INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Two-Dimensional Zero-Phase FIR Filter Design with Nonuniform Frequency Sampling

Valentin Ninov

A Thesis

in

The Department

of

Electrical Engineering

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Applied Science at
Concordia University
Montreal, Quebec, Canada

April 1999

ABSTRACT

Two-Dimensional Zero-Phase FIR Filter Design with Nonuniform Frequency Sampling

Valentin Ninov

The present thesis is concerned with design techniques for two-dimensional zero-phase finite impulse response digital filters with nonuniform frequency samples. Using the freedom and flexibility of the nonuniform frequency sampling, several techniques for taking samples in the frequency plane have been proposed. The design problem is treated as a bivariate interpolation problem with unevenly spaced data. The main idea is to select (find) sampling locations in the (ω_1, ω_2) frequency plane and corresponding sample values $H_d(\omega_{1k}, \omega_{2k})$ of the desired filter frequency response such that the approximation error in the designed filter is reduced significantly and the shape performance is high.

Three main types of two-dimensional zero-phase FIR filters that are most frequently used in practice have been considered: rectangular, circular, and 2-D halfband (diamond and fan) FIR filters. At least two sampling techniques for each filter type have been proposed and examined. Common features of all of the proposed sampling approaches are that the design is performed entirely with real number arithmetic, no computationally expensive iterative procedures are used, and the samples are taken on curves that match the contours of the desired filter frequency response magnitude,.

Although the filters designed with the proposed techniques are not optimal (in strict sense), the methods are conceptually simple and produce filters with high degree of shape regularity and approximation error comparable and sometimes even smaller than the "conventional" 2-D FIR filter design methods.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my thesis supervisor, Dr. Venkat Ramachandran for his valuable guidance and encouragement throughout the course of this thesis, for his extremely careful, critical, and thorough review of my work. I would also like to thank Dr. William Lynch for his consideration and useful suggestions. I thank Dr. E. Doedel for the useful discussion. I am extremely grateful to my dear wife Nelly for her enormous support, encouragement, and understanding.

Table of Contents

List of Figures

List of Tables

List of Symbols and Important Abbreviations

All vectors and matrices are denoted with boldface letters.

ω'	1-D radian frequency
ω_1 , ω_2	2-D horizontal and vertical radian frequencies
$\boldsymbol{\omega}$	frequency vector, $\boldsymbol{\omega} = (\omega_1 , \omega_2)^T$
$h(n_1, n_2) \equiv h[n_1, n_2]$	impulse response of a two-dimensional digital filter, usually the filter to be designed
$h_d(n_1, n_2)$	impulse response of a desired (ideal) 2-D digital filter
$H(\omega_1, \omega_2) \equiv H(e^{j\omega_1}, e^{j\omega_2})$	frequency response of the filter to be designed
$H_d(\omega_1 , \omega_2)$	desired (ideal) frequency response
NDFT	nonuniform discrete Fourier transform
$\hat{X}(z_k)$	samples of the NDFT
$\hat{\mathbf{X}}$	vector containing the NDFT points
$\mathbf{D}$	NDFT matrix
$\mathbf{V}_1$, $\mathbf{V}_2$	Vandermonde matrices
N	order of interpolation polynomial
$N_1 \times N_2$	2-D FIR filter impulse response region of support size
$N \times N$	2-D FIR filter size in the case $N_1 = N_2$
M_1, M_2	numbers related to the filter size: $N_1 = 2M_1+1$, $N_2 = 2M_2+1$
M	for a filter of size $N \times N$, $M = (N-1)/2$
L	number of independent filter points
N_{1D}	length of a 1-D FIR filter
$a(n_1, n_2)$	a 2-D sequence obtained from a fourfold $h(n_1, n_2)$
$b(n_1, n_2)$	a 2-D sequence obtained from an eightfold $h(n_1, n_2)$
$b(i)$	a 1-D sequence obtained from $b(n_1, n_2)$
$\phi_k(\omega_1, \omega_2)$	2-D interpolation basis functions

P	number of sampling contours in the passband
S	number of sampling contours in the stopband
$\Phi(\omega_k)$	characteristic vector associated with a particular point ω_k
$T_n(x)$	n-th order Chebyshev polynomial
$\mathbf{C}$	interpolation matrix for circular shaped FIR filters
$\hat{\mathbf{H}}$	vector containing the frequency sample values
R_π	region in the (ω_1, ω_2) plane defined as $R_\pi : \left\{\sqrt{\omega_1^2 + \omega_2^2} \le \pi\right\} \cap \{0 \le \omega_1 \le \pi,\ 0 \le \omega_2 \le \omega_1\}$
R_C	region in the (ω_1, ω_2) plane defined as $R_C : \left\{\sqrt{\omega_1^2 + \omega_2^2} > \pi\right\} \cap \{0 \le \omega_1 \le \pi,\ 0 \le \omega_2 \le \omega_1\}$
$\mathscr{R}_p$	passband region
$\mathscr{R}_s$	stopband region
ω_p	2-D passband frequency edge
ω_s	2-D stopband frequency edge
ω'_p	1-D passband frequency edge
ω'_s	1-D stopband frequency edge
δ_{1D}	ripple of 1-D FIR digital filter, usually optimal filter
Q	number of 1-D extremal frequencies
Exp	Sampling technique on contours with exponential distribution
Sc	sampling technique by scaling the 1-D frequency axis, Sec. 3.3.3
HC	samples on hyperbolic contours in R_C, Sec. 3.3.4
1-D Opt	sampling technique using the extremal frequencies of 1-D optimal filter design
Win	Rotated 1-D Kaiser window method
FT	Frequency transformation using the McClellan's transformation sequence
LS	Least squares solution

Foreword (Motivation for the Study)

In the last two decades there has been a great deal of interest in designing two-dimensional (2-D) digital filters and their applications to 2-D digital signal processing. This interest has been boosted by the advances of very large scale integrated (VLSI) circuits which have allowed real-time operations of 2-D digital filters. 2-D digital filters find applications in such versatile areas as digital processing of aerial and satellite photographs, enhancement of X-rays, computer tomography, digital television, radio astronomy, processing of geophysical data, and radar, to name just a few.

The two classes of digital filters depending on the nature of their impulse response are the infinite-extent impulse response (IIR) filters and finite-extent impulse response (FIR) digital filters. The IIR filters are more economical, i.e., they can meet particular design specifications with a significantly smaller number of filter coefficients than do the FIR filters. However, the 2-D FIR filters are widely used in the field of digital signal processing and often preferred to the 2-D IIR filters because of the attractive properties of the former:

(i) inherent stability, the impulse response is finite and therefore, it is always absolutely summable, or alternatively, all poles of the system function are at the origin of the complex (z_1, z_2) space. The stability is never an issue in design or implementation;

(ii) the ability to attain a linear or zero phase response, constant group delay, respectively; the zero phase is often a requirement, especially in the image processing, and it cannot be satisfied with a single IIR filter;

(iii) relative ease of design, arbitrary frequency response can be closely approximated with sufficiently large size of impulse response region of support;

(iv) efficient realization through high-speed convolution using FFT. If an FFT realization of FIR filters is used the advantage of IIR filters is not so big.

(v) if a zero phase FIR filter is required the design problem is simplified: the frequency response is purely real and the number of independent coefficients is reduced;

(vi) FIR filters have good quantization properties;

(vii) efficient implementation of zero phase FIR filters: using the existing symmetries the number of arithmetic operations per output point can be reduced at least by 50 percent.

The problem of designing a digital filter is basically a problem of finding the impulse response (or transfer function) coefficients that meet the design specifications. Existing standard methods for designing 2-D FIR filters are the windowing, the frequency-sampling method, the frequency transformation method, and optimal (minimax) methods [10].

The 2-D window method is based on the same concept as the 1-D counterpart. The desired filter frequency response $H_d(\omega_1, \omega_2)$ is known and the corresponding impulse response $h_d(n_1, n_2)$ is found by inverse Fourier transforming $H_d(\omega_1, \omega_2)$. Then this desired or ideal frequency response $h_d(n_1, n_2)$ which is of infinite extent is truncated to a finite-extent sequence. In order to moderate the Gibbs phenomenon, this truncation performed with a window function $w(n_1, n_2)$, that is,

$$h(n_1, n_2) = h_d(n_1, n_2)w(n_1, n_2).$$

If $h_d(n_1, n_2)$ and $w(n_1, n_2)$ are both symmetric with respect to the origin, $h_d(n_1, n_2)$ will be also and the designed filter has a zero phase frequency response $H(\omega_1, \omega_2)$. This frequency response is a smoothed version of the ideal frequency response. One-dimensional windows are often used as a basis for generating 2-D windows [10, 11, 25]. There are two methods by which this is usually done. The first method is to obtain a 2-D window $w(n_1, n_2)$ as an outer product of two 1-D windows, $w(n_1, n_2) = w_1(n)\, w_1(n)$. The second method, proposed by Huang [25], is to obtain a 2-D window $w(n_1, n_2)$ by sampling a circularly rotated 1-D continuous window function, $w(n_1, n_2) =$

$w_c\left(\sqrt{n_1^2 + n_2^2}\right)$. Among the most popular 1-D windows that can be used to obtain a 2-D window are the rectangular, the Hamming and the Kaiser windows. the window method is quite general and it is not optimal. There is no control over the frequency domain specifications and sometimes it is necessary to design several filters to meet the design specifications. Although, the window method is simple conceptually and computationally and it is often used.

The frequency transformation method was originally proposed by McClellan [24] and further developed by other authors [11]. In this method, a 2-D zero-phase FIR filter is designed from a 1-D zero-phase FIR filter using the frequency transformation

$$H(\omega_1, \omega_2) = H(\omega)\Big|_{\cos\omega = F(\omega_1, \omega_2)}$$

where $F(\omega_1, \omega_2)$ is the Fourier transform of a finite-extent zero-phase sequence $t(n_1, n_2)$. Some first-order transformations for the design of circularly symmetric, diamond, and fan shaped filters are shown in Chapters 3 and 4, respectively. There are two specific design approaches. In the firs method, $t(n_1, n_2)$ is chosen among the results given in the open literature. In the second approach, $t(n_1, n_2)$ is designed for the specific need. The frequency transformation method appears more complex conceptually than the other standard methods for 2-D FIR filter. Nevertheless, this method has short design time and its performance appear to be better than the window method and frequency sampling (uniform) method [10]. In general., the frequency transformation method does not produce optimal filters. In some restricted set of cases, however, this method can produce optimal filters in the Chebyshev sense [10].

The optimal 2-D FIR design involve optimization of the designed filter coefficients such that some function of the error between the resulting filter frequency response and the desired filter frequency response is minimized. Usually, this is the Chebyshev (L_∞) norm

$$E_\infty = \max_{(\omega_1, \omega_2)} \left| H(\omega_1, \omega_2) - H_d(\omega_1, \omega_2) \right|$$

that is minimized. The error can also be weighted. In 2-D the Haar condition is not in general satisfied, see Sec. 1.2, ant the alternation theorem does not apply to the minimization of the approximation error. Some iterative algorithms of the Remez exchange type have been developed [11, 21, 22]. These algorithms are very expensive computationally, take long time, and do not always converge to a correct solution. Some terminology and theorems of bivariate best approximation are considered in Sec. 1.2. In contrast to the 1-D case, a practical procedure for reliably designing a 2-D optimal FIR filter remains as an area of research.

The uniform frequency sampling technique is not optimal in any sense and does not give control over the frequency domain parameters. Yet it is widely used because of its conceptual and computational simplicity. The traditional frequency sampling design methods are based on the discrete Fourier transform (DFT), the fast Fourier transform (FFT) in particular, and they are, therefore, applicable to uniform frequency samples. Traditionally, in the 1-D frequency sampling approach, the desired frequency response is sampled at N equally spaced frequencies, where N is the filter length. An N-point inverse DFT is used to compute the filter coefficients. The 2-D frequency sampling method involve sampling a desired frequency response at the vertices of a uniform Cartesian grid. Main disadvantage of the DFT and uniform sampling in general, is the limited (and uniform) spectral resolution. For a given DFT length N (or of size $N \times N$ in 2-D), the spectral resolution is fixed by the number N and is $2\pi/N$ (horizontally an vertically in 2-D). The mathematical framework of the uniform frequency sampling method is presented in Chapter 2.

Some relatively recent studies [1, 2, 6-9] showed the great potential of the nonuniform frequency sampling for designing of 2-D FIR filters. The nonuniform frequency sampling gives flexibility in choosing the sampling points and, therefore, controlling the spectral resolution. The samples can be chosen so that the resulting interpolated frequency response is very close to the optimal response, i. e., a response with maximum error in the domain of approximation near to its minimum. The FIR filter design is generally an

approximation problem. The transfer function of a 1-D filter is a 1-D polynomial of a finite order and therefore can be reconstructed from a finite number of uniform or nonuniform frequency samples. However, some mathematical results in 1-D do not hold in two or higher dimensions. For example, the fundamental theorem of algebra of factorizability of univariate polynomials does not hold in two or more dimensions. As it has been mentioned above, iterative design techniques leading to 2-D optimal FIR filters have been developed. The final stage of such an algorithm is a 2-D polynomial interpolation with a set of nonuniformly spaced frequency samples. A set of L+1 linear equations[†] with L+1 unknowns is solved in order to find the L independent filter points. Therefore, if this set of frequency samples was known at the beginning, the optimal design problem would be solved in one step, without complex iterative multiple exchange algorithms of the Remez type. Hence, one of the ideas and motivations of the present study is to determine, using simple non-iterative procedures, a frequency sample point set in the domain of approximation which will be a kind of approximation to the point set producing an optimal (minimax) filter. In this way, a good approximation of the desired frequency response would be obtained and the designed filter would be nearly optimal in the Chebyshev sense. The present thesis does not pretend to give the exact algorithm but it is believed that it makes one step forward in the right direction. The 2-D nonuniform discrete Fourier transform (NDFT) and the bivariate polynomial interpolation are the mathematical framework of the 2-D zero-phase FIR filter design with nonuniform frequency samples. An overview of this theory is presented in Chapter 1.

[†] The notation L is used later for the number of frequency samples used in the proposed nonuniform sampling techniques for 2-D FIR filter design. The meaning is the same.

Scope of the Thesis

The objective of the thesis is to investigate possible sampling techniques and to propose the satisfactory of them for 2-D zero-phase FIR digital filter design with nonuniform frequency sampling. These sampling techniques should be conceptually simple and not requiring computationally expensive iterative procedures. Also, it is expected these techniques to produce 2-D FIR filters with low approximation error, comparable to the error produced by other methods as the window method, the uniform frequency sampling method, and the frequency transformation method for the same filter size. Additionally, the proposed techniques should produce filters with regular shapes, e.g., circular, square, fan, etc.

Chapter 1 begins with a brief review of the 1-D and 2-D nonuniform discrete Fourier transform [1, 2] since the inverse NDFT is the basis for the nonuniform frequency sampling design from signal processing point of view. Some particular cases in which the inverse NDFT has a unique solution are considered. Next, an overview is given of some of the most important results in the bivariate interpolation theory. This theory is the framework of the nonuniform frequency sampling FIR filter design from polynomial interpolation point of view.

The next three chapters, 2, 3, and 4, are concerned with nonuniform sampling design techniques of 2-D zero-phase FIR filters of different shapes. Several sampling techniques have been proposed, which are illustrated with many design examples.

In Chapter 2, the design of zero-phase rectangularly shaped FIR filters with nonuniform sampling is considered. Having in consideration the existing symmetries in the frequency response and consequently, in the impulse response of these filters, the necessary number of frequency samples is taken only in the first quadrant of the (ω_1, ω_2) plane. The filters

with rectangular shapes are the easiest to design among the considered shapes. The two particular cases in which the inverse NDFT is nonsingular can be used: 1) frequency samples taken arbitrary on vertical lines and 2) samples taken at the vertices of a nonuniform rectangular grid. In both cases the frequency samples lie on straight horizontal and/or vertical lines which naturally describe the filter shape contours. Several techniques for determining the line coordinates are proposed. They include line coordinates obtained from the extremal frequencies of a 1-D optimal FIR filter, from samples of exponential functions, and from harmonic series. By adjustment of a single parameter, the exponential and harmonic series techniques allow the approximation error to be controlled to a certain degree. The methods are computationally efficient since instead of one large 2-D problem several small 1-D problems are solved.

In Chapter 3, the design of zero-phase circularly shaped FIR filters with nonuniform samples is considered. In this case, in order to obtain regular circular shape, the frequency samples are taken on circular contours centered at the origin of the (ω_1, ω_2) plane and extending from (0, 0) till $\omega_1^2 + \omega_2^2 \le \pi^2$. The sampling on straight lines cannot be applied and therefore, the problem cannot be decomposed to several small 1-D problems. In compensation, the number of independent filter coefficients is smaller because of the eightfold symmetry. The approximation domain is chosen to be the first octant of the frequency plane. The most important issues here are the distribution of circular contours in the region $\omega_1^2 + \omega_2^2 \le \pi^2$ and the contour shape in $\{ \omega_1^2 + \omega_2^2 > \pi^2 , | \omega_1 |, |\omega_2 | \le \pi^2 \}$. The first problem is solved by taking samples on circular contours having radii equal to the extremal frequencies with the 1-D optimal filter design method [18]. This is a right choice since the frequency response behavior of the designed filter is (almost) equiripple till $\omega_1^2 + \omega_2^2 \le \pi^2$ provided no samples were taken in the region $\omega_1^2 + \omega_2^2 > \pi^2$. Determining the shape of the contours in the corner near the point (π, π) proved to be the most problematic issue. Filters with good performance were obtained in the case of hyperbolic contours in this region. Several other techniques have been tried with less success. Some measures should be taken to avoid singularities of the inverse problem. Fortunately, the interpolation problem runs into singularities very seldom. Though the system of linear equations is often

ill-conditioned, the proposed techniques allow circularly shaped FIR filters with relatively low deviations and high shape regularity to be designed.

In Chapter 4, the design of 2-D half-band FIR filters, namely 90° fan and diamond shaped filters, with nonuniform samples is investigated. Two sampling techniques are proposed. They are conceptually similar to the techniques used in the previous chapters. In both cases frequency samples are taken on straight lines parallel to the contours of the desired filter frequency response. One of the techniques calculates the location of these lines using extremal points of 1-D optimal filter design. The other technique uses exponential distribution of parallel lines in the domain of approximation. The designed filters have low approximation error and very regular shape. The second method produces superior results. At the end, the design of 2-D zero-phase FIR filters with different shapes is considered and some of the capabilities of the nonuniform frequency sampling approach are demonstrated with several design examples. The sampling techniques proposed in the previous two chapters are also employed.

In Chapter 5, conclusions about the 2-D zero-phase FIR filter design with nonuniform frequency sampling, in general, and for the proposed sampling techniques, in particular, are made. Some possible directions and recommendations for further research are given.

The Appendix contains listings of the Matlab code for the most important algorithms proposed and used in the present study.

Chapter 1
Theoretical Preliminaries

The 2-D FIR filter design problem with nonuniform frequency sampling can be approached using two major frameworks: the nonuniform discrete Fourier transform (NDFT) and bivariate interpolation theory. These theoretical frameworks are reviewed in the two sections of this chapter. In Section 1.1, a review of the one- and two-dimensional nonuniform discrete Fourier transform is presented. The basic idea of the NDFT and existing methods for computing the inverse NDFT are presented in the first two subsections. In Subsection 1.1.3, the two-dimensional NDFT is reviewed. At the end of Section 1.1, two special cases of 2-D NDFT are discussed in which the inverse NDFT is guaranteed to exist. In Section 1.2, an overview of some important results in bivariate polynomial interpolation theory is presented. This section begins by outlining the main differences between univariate and bivariate polynomial interpolation. Next, some existing results in bivariate polynomial interpolation are presented. These results are based on restrictive cases in which the interpolation problem has a unique solution. Some conditions under which the 2-D interpolation problem cannot be uniquely solved are also considered. At the end, some terminology and results characterizing the best approximation by generalized polynomials in 2-D are reproduced.

The inverse nonuniform Fourier transform, as defined in [1, 2], is the basic framework for nonuniform frequency sampling. The uniform frequency sampling methods are based on the inverse conventional discrete Fourier transform (DFT). Given an aperiodic

sequence $x(n)$ of length N with Fourier transform $X\left(e^{j\omega}\right)$, its DFT is evaluation of the z-transform $X(z)$ of $x(n)$ on the unit circle at N equally spaced points [13].

$$\tilde{X}(k)=X(z)\Big|_{z=e^{j(2\pi/N)k}}=X\left(e^{j(2\pi/N)k}\right) \tag{1.0.1}$$

where $\tilde{X}(k)$ is the discrete Fourier series (DFS) representation of the periodic extension $\tilde{x}(n)$ of $x(n)$,

$$\tilde{x}(n)=\sum_{m=-\infty}^{\infty}x(n+mN) \tag{1.0.2}$$

The DFT $X(k)$ of $x(n)$ is the sequence of length N obtained by taking just one period of $\tilde{X}(k)$:

$$X(k)=\begin{cases}\tilde{X}(k), & 0\le k\le N-1\\ 0, & \text{otherwise}\end{cases} \tag{1.0.3}$$

The discrete Fourier transform is algebraically a matrix-vector product. Calling $\mathbf{x} = [x(0), x(1), ..., x(N-2)\ x(N-1)]^T$ the vector of the input values, $\mathbf{X} = [X(0), X(1), ..., X(N-1)]^T$ the vector of transform values, and $W_N = e^{-j2\pi/N}$ is the primitive N-th root of unity, the DFT can be written as

$$\begin{bmatrix}X(0)\\X(1)\\\vdots\\X(N-1)\end{bmatrix}=\begin{bmatrix}1 & 1 & 1 & \cdots & 1\\ 1 & W_N & W_N^2 & \cdots & W_N^{N-1}\\ 1 & W_N^2 & W_N^4 & \cdots & W_N^{2(N-1)}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & W_N^{N-1} & W_N^{2(N-1)} & \cdots & W_N^{(N-1)(N-1)}\end{bmatrix}\begin{bmatrix}x(0)\\x(1)\\\vdots\\x(N-1)\end{bmatrix} \tag{1.0.4}$$

The two-dimensional DFT is a straightforward extension [10, 11] of the 1-D case. In 2-D, the discrete transform pair is given by

$$X(k_1,k_2)=\begin{cases}\sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1}x(n_1,n_2)W_{N_1}^{k_1n_1}W_{N_2}^{k_2n_2}, & \begin{matrix}0\le k_1\le N_1-1\\ 0\le k_2\le N_2-1\end{matrix}\\ 0, & \text{otherwise}\end{cases} \tag{1.0.5a}$$

$$x(n_1,n_2)=\begin{cases}\dfrac{1}{N_1N_2}\displaystyle\sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1}X(k_1,k_2)W_{N_1}^{-k_1n_1}W_{N_2}^{-k_2n_2}, & \begin{cases}0\le n_1\le N_1-1\\0\le n_2\le N_2-1\end{cases}\\ 0, & \text{otherwise}\end{cases} \tag{1.0.5b}$$

1.1 The Nonuniform Discrete Fourier Transform, Review.

1.1.1 Definition

The nonuniform discrete Fourier transform [1, 2] is a generalization of the DFT and the chirp z-transform (CZT) [12]. The nonuniform discrete Fourier transform of a sequence $x(n)$ of length N is defined as [1, 2]

$$\hat{X}(z_k)=\sum_{n=0}^{N-1}x(n)z_k^{-n}, \qquad k=0,1,\ldots,N-1 \tag{1.1.1}$$

where $z_0, z_1, \ldots, z_{N-1}$ are distinct points located arbitrarily on the z-plane. When these points are located on the unit circle and spaced at equal angles the NDFT reduces to the conventional DFT. Eq. (1.1.1) can be expressed in matrix form as

$$\hat{\mathbf{X}}=\mathbf{V}\mathbf{x} \tag{1.1.2}$$

where

$$\hat{\mathbf{X}}=\begin{bmatrix}\hat{X}(z_0)\\ \hat{X}(z_1)\\ \vdots\\ \hat{X}(z_{N-1})\end{bmatrix}, \tag{1.1.3a}$$

$$\mathbf{x}=\begin{bmatrix}x(0)\\ x(1)\\ \vdots\\ x(N-1)\end{bmatrix} \tag{1.1.3b}$$

and

$$\mathbf{V}=\begin{bmatrix}1 & z_0^{-1} & z_0^{-2} & \cdots & z_0^{-N+1}\\ 1 & z_1^{-1} & z_1^{-2} & \cdots & z_1^{-N+1}\\ 1 & z_2^{-1} & z_2^{-2} & \cdots & z_2^{-N+1}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & z_{N-1}^{-1} & z_{N-1}^{-1} & \cdots & z_{N-1}^{-N+1}\end{bmatrix}. \tag{1.1.4}$$

When the NDFT is evaluated on the unit circle, the above Vandermonde matrix becomes

$$\mathbf{V}=\begin{bmatrix} 1 & e^{-j\omega_0} & e^{-j2\omega_0} & \cdots & e^{-j(N-1)\omega_0} \\ 1 & e^{-j\omega_1} & e^{-j2\omega_1} & \cdots & e^{-j(N-1)\omega_1} \\ 1 & e^{-j\omega_2} & e^{-j2\omega_2} & \cdots & e^{-j(N-1)\omega_2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j\omega_{N-1}} & e^{-j2\omega_{N-1}} & \cdots & e^{-j(N-1)\omega_{N-1}} \end{bmatrix} \tag{1.1.5}$$

In the design of digital filters from nonuniform frequency samples we are primarily interested in the existence and calculation of the inverse NDFT. If the N sampling points are distinct then the Vandermonde matrix $\mathbf{V}$ is non-singular. This can be seen easily from the factored form of the determinant of $\mathbf{V}$ [3, 29] :

$$\det(\mathbf{V}) = \prod_{i\neq j,\ i>j} \left(z_i^{-1} - z_j^{-1}\right) = \prod_{j=0}^{N-2}\left[\prod_{i=j+1}^{N-1}\left(z_i^{-1} - z_j^{-1}\right)\right] \tag{1.1.6}$$

Therefore, if $z_i \neq z_j$ for $i \neq j$, then $\det(\mathbf{V}) \neq 0$ and $\mathbf{V}^{-1}$ exists. The inverse NDFT then exists and is given by

$$\mathbf{x} = \mathbf{V}^{-1}\mathbf{X} \tag{1.1.7}$$

When the N sampling points are located at equal angles on the unit circle in the z-plane then the matrix $\mathbf{V}$ reduces to the 2-D DFT matrix (cf. Eq. (1.0.4))

$$\mathbf{V} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & W_N & W_N^2 & \cdots & W_N^{N-1} \\ 1 & W_N^2 & W_N^4 & \cdots & W_N^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & W_N^{N-1} & W_N^{2(N-1)} & \cdots & W_N^{(N-1)(N-1)} \end{bmatrix} \tag{1.1.8}$$

where $W_N = e^{-j2\pi/N}$.

1.1.2 Computing the inverse NDFT.

The problem of computing the inverse NDFT can be viewed as polynomial interpolation problem. The theorem of polynomial interpolation states that [3]: Given N distinct (real or complex) points $z_0, z_1, \ldots, z_{N-1}$ and N (real or complex) values $f_0, f_1, \ldots, f_{N-1}$, there exist a unique polynomial of order $N-1$

$$p(z) = a_0 + a_1 z + a_2 z^2 + \ldots + a_{N-1} z^{N-1}$$

for which

$$p(z_k) = f_k, \qquad k = 0, 1, \ldots, N-1 \tag{1.1.9}$$

This interpolation problem can be solved using different methods.

(a) **Direct Method**

The inverse NDFT **x** is found by directly solving the linear system given by Eq. (1.1.2) using Gaussian elimination. The complexity is on the order of $O(N^3)$ arithmetic operations.

(b) **Lagrange Interpolation**

$X(z)$ is expressed as a Lagrange polynomial of order $N-1$,

$$X(z) = \sum_{k=0}^{N-1} \frac{L_k(z)}{L_k(z_k)} \hat{X}(k) \tag{1.1.10}$$

where $L_0(z), L_1(z), \ldots, L_{N-1}(z)$ are the fundamental polynomials, defined as

$$L_k(z) = \prod_{i \neq k} (1 - z_i z^{-1}), \qquad k = 0,1,\ldots,N-1. \tag{1.1.11}$$

(c) **Newton Interpolation**

The z-transform of $x(n)$, $X(z)$ is obtained as a Newton interpolating polynomial from the given sample values $\hat{X}(z_k)$ and the sample locations z_k and is expressed as

$$X(z)=c_0+c_1(z-z_0)+c_2(z-z_0)(z-z_1)+\ldots+c_{N-1}\prod_{k=0}^{N-2}(z-z_k), \tag{1.1.12}$$

Each coefficient c_m can be represented as a divided difference of the m-th order of the sample values $X(0), X(1), \ldots, X(m)$ with respect to the sampling points $z_0, z_1, \ldots, z_m$. Each divided difference c_m is a linear combination of $X(k)$ and z_k, $k = 0, 1, 2, \ldots, m$.

$$\begin{aligned} c_0 &= \hat{X}(0) \\ c_1 &= \frac{\hat{X}(1)-\hat{X}(0)}{z_1-z_0} = \frac{\hat{X}(1)-c_0}{z_1-z_0} \\ c_2 &= \frac{\dfrac{\hat{X}(2)-\hat{X}(1)}{z_2-z_1} - \dfrac{\hat{X}(1)-\hat{X}(0)}{z_1-z_0}}{z_2-z_0} = \frac{\hat{X}(2)-\hat{X}(1)-c_1(z_2-z_1)}{(z_2-z_0)(z_2-z_1)} \\ &\vdots \end{aligned} \tag{1.1.13}$$

The Newton interpolation has a permanence property. This means that if an additional point is included, the coefficients c_m need not to be recomputed as in the Lagrange representation. One more term in (1.1.12) is added instead.

1.1. 3 Two-dimensional NDFT

The 2-D nonuniform discrete Fourier transform of a sequence $x(n_1, n_2)$ of size $N_1 \times N_2$ is defined as [1, 2]

$$\hat{X}(z_{1k}, z_{2k}) = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} x(n_1, n_2)\, z_{1k}^{-n_1} z_{2k}^{-n_2}, \qquad k = 0, 1, \ldots, N_1N_2-1 \tag{1.1.14}$$

In other words, the 2-D NDFT corresponds to sampling the z-transform of the 2-D sequence $x(n_1, n_2)$. Without loss of generality, here $x(n_1, n_2)$ is assumed a first-quadrant sequence. The sampling points (z_{1k}, z_{2k}) are distinct 4-D points in the (z_1, z_2) space. They

can be chosen arbitrary but in a way that the inverse NDFT exists. For the purpose of 2-D FIR filter design these points are usually taken on the unit surface ($|z_1| = 1$, $|z_2| = 1$). Eq. (1.1.14) can be expressed in a matrix form as

$$\hat{\mathbf{X}} = \mathbf{D}\mathbf{x} \tag{1.1.15}$$

where

$$\hat{\mathbf{X}} = \begin{bmatrix} \hat{X}(z_{10}, z_{20}) \\ \hat{X}(z_{11}, z_{21}) \\ \vdots \\ \hat{X}\left(z_{1(N_1-1)}, z_{2(N_2-1)}\right) \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} x(0,0) \\ x(0,1) \\ \vdots \\ x(N_1-1, N_2-1) \end{bmatrix} \tag{1.1.16}$$

and

$$\mathbf{D} = \begin{bmatrix} 1 & z_{20}^{-1} & z_{20}^{-2} & \cdots & z_{10}^{-N_1+1} z_{20}^{-N_2+1} \\ 1 & z_{21}^{-1} & z_{21}^{-2} & \cdots & z_{11}^{-N_1+1} z_{21}^{-N_2+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{2(N_2-1)}^{-1} & z_{2(N_2-1)}^{-2} & \cdots & z_{1(N_1-1)}^{-N_1+1} z_{2(N_2-1)}^{-N_2+1} \end{bmatrix} \tag{1.1.17}$$

Clearly, in order (1.1.14) to be a transform, the number of the sampling points in the (z_1, z_2) space must be $N = N_1 \times N_2$. The size of the 2-D NDFT matrix is $N \times N$.

In general, the determinant of the 2-D NDFT matrix cannot be factored. Distinct sampling points (z_{1k}, z_{2k}) do not guarantee that the matrix **D** is not singular [2]. There are, however, some special cases in which the determinant can be factored.

From a 2-D FIR filter design point of view, finding the inverse NDFT can be alternatively stated as a 2-D polynomial interpolation problem. The filter impulse response coefficients appear as coefficients of a bivariate polynomial (the transfer function)

$$H(z_1, z_2) = \sum_{n_1,\ n_2 \in R_h} \sum h(n_1, n_2)\, z_1^{-n_1} z_2^{-n_2} \tag{1.1.18}$$

where R_h is the region of support of $h(n_1, n_2)$. In [9] a number of sufficient conditions have been derived under which the two-dimensional polynomial interpolation problem

has a unique or nonunique solution. However, no set of necessary and sufficient conditions has been found. Some of the theorems stating the above-mentioned conditions are presented in the next section. From filter design point of view, the possible singularity of the determinant of **D** is not a serious problem because if that is the case, a different set of sampling points can be chosen. Actually, the practice shows that when the sampling points are distinct, the 2-D NDFT matrix **D** is very seldom singular.

1.1.4 Special cases of two-dimensional NDFT

If we constrain the locations of the sampling points (z_{1k}, z_{2k}) in a special way, we can obtain a 2-D NDFT matrix **D** whose determinant is factorizable. Thus, it is guaranteed that this matrix is nonsingular and, therefore, the inverse NDFT exists.

A) Sampling points located on parallel vertical (or horizontal) straight lines in (ω_1, ω_2) plane.

This means the sampling points are taken on the unit surface ($|z_1|=1$, $|z_2|=1$) in the (z_1, z_2) space. Arbitrary samples taken on parallel vertical lines is the strategy employed in [6] and [7].

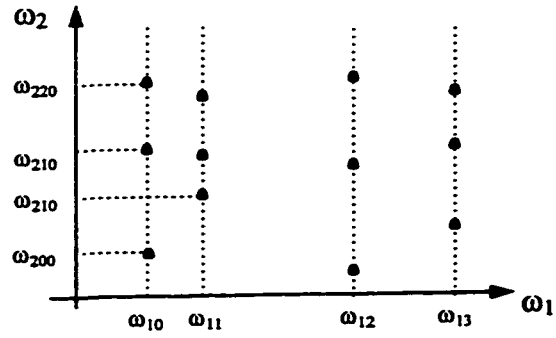


Figure 1.1 Sampling points located on parallel lines to ω_2 axis with $N_1 = 4$ and $N_2 = 3$.

For an $N_1 \times N_2$-point sequence, the number of samples is $N_1 \times N_2$. These samples are taken on N_1 vertical lines with coordinates $\omega_{10}, \omega_{11}, \ldots, \omega_{1k}, \ldots, \omega_{1,N1-1}$. For each selected value ω_{1i}, $(i = 0, 1, \ldots, N_1\text{-}1)$, N_2 values of ω_2 are arbitrarily selected and denoted as $\omega_{2j(i)}$, $j = 0, 1, \ldots, N_2$, Fig. 1.1. The equivalent points in (z_1, z_2) space lie on the unit surface:

$$z_{1i} = e^{j\omega_{1i}}, \quad z_{2m(k)} = e^{j\omega_{2j(i)}}, \qquad i = 0,\ldots, N_1 - 1, \qquad j = 0,\ldots, N_2 - 1 \tag{1.1.19}$$

Kronecker left product (tensor product) of two matrices **A** $(m \times n)$ and **B** $(p \times q)$ is defined as the $(mp \times nq)$ matrix

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} \mathbf{A}b_{11} & \mathbf{A}b_{12} & \cdots & \mathbf{A}b_{1q} \\ \mathbf{A}b_{21} & \mathbf{A}b_{22} & \cdots & \mathbf{A}b_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}b_{p1} & \mathbf{A}b_{p2} & \cdots & \mathbf{A}b_{pq} \end{bmatrix} \tag{1.1.20}$$

One of the Kronecker product properties states that [15] if **A** is an $n \times n$ matrix and **B** is $q \times q$, then

$$\det(\mathbf{A} \otimes \mathbf{B}) = (\det \mathbf{A})^q (\det \mathbf{B})^n \tag{1.1.21}$$

In the case of samples on parallel lines, Eq. (1.1.15) can be expressed in the form

$$\mathbf{X} = \mathbf{D}\mathbf{x} = \left(\{\mathbf{V}_2\} \otimes \mathbf{V}_1\right)\mathbf{x} \tag{1.1.22}$$

In other words, the 2-D NDFT matrix **D** is decomposed as a generalized Kronecker product of a set $\{\mathbf{V}_2\}$ of N_1 Vandermonde matrices and another Vandermonde matrix $\mathbf{V}_1$.

$$\mathbf{D} = \{\mathbf{V}_2\} \otimes \mathbf{V}_1 = \begin{Bmatrix} \mathbf{V}_{20} \\ \mathbf{V}_{21} \\ \vdots \\ \mathbf{V}_{2,N_1-1} \end{Bmatrix} \otimes \mathbf{V}_1 = \begin{bmatrix} \mathbf{V}_{20} \otimes \mathbf{v}_0 \\ \mathbf{V}_{21} \otimes \mathbf{v}_1 \\ \vdots \\ \mathbf{V}_{2,N_1-1} \otimes \mathbf{v}_{N_1-1} \end{bmatrix} \tag{1.1.23}$$

where $\mathbf{v}_i$, $i = 0, \ldots, N_1\text{-}1$, denotes the i-th row vector of matrix $\mathbf{V}_1$.

$$\mathbf{V}_1 = \begin{bmatrix} 1 & z_{10}^{-1} & z_{10}^{-2} & \cdots & z_{10}^{-(N_1-1)} \\ 1 & z_{11}^{-1} & z_{11}^{-2} & \cdots & z_{11}^{-(N_1-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{1,N_1-1}^{-1} & z_{1,N_1-1}^{-2} & \cdots & z_{1,N_1-1}^{-(N_1-1)} \end{bmatrix} \tag{1.1.24}$$

$$\mathbf{V}_{2i} = \begin{bmatrix} 1 & z_{20i}^{-1} & z_{20i}^{-2} & \cdots & z_{20i}^{-(N_2-1)} \\ 1 & z_{21i}^{-1} & z_{21i}^{-2} & \cdots & z_{21i}^{-(N_2-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{2,N_2-1,i}^{-1} & z_{2,N_2-1,i}^{-2} & \cdots & z_{1,N_2-1,i}^{-(N_2-1)} \end{bmatrix}, \qquad i = 0, 1, \ldots, N_1 - 1 \tag{1.1.25}$$

According to property (1.1.21), the determinant of **D** can be factored in the following way

$$\det(\mathbf{D}) = \left[\det(\mathbf{V}_1)\right]^{N_2} \prod_{i=1}^{N_1-1} \det(\mathbf{V}_{2i}) = \prod_{i \neq j,\, i>j} \left(z_{1i}^{-1} - z_{1j}^{-1}\right)^{N_2} \prod_{i=1}^{N_1-1} \prod_{m \neq n,\, m>n} \left(z_{2mi}^{-1} - z_{2ni}^{-1}\right) \tag{1.1.26}$$

From (1.1.26) it is clear, that if the vertical lines with coordinates z_{1i}, $i = 0, \ldots, N_1\text{-}1$, are all distinct, i. e., $z_{1i} \neq z_{1j}$, $i \neq j$, and if no two samples on the same line coincide, i. e., $z_{1mi} \neq z_{1ni}$, $m \neq n$, then the 2-D NDFT matrix D is nonsingular and the inverse transform exists. Hence, the 2-D interpolation problem can be solved uniquely. The inverse 2-D NDFT in this case can be obtained as follows. First, Eq. (1.1.14) is written as [2], [6]

$$\hat{X}\left(z_{1i}, z_{2ij}\right) = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} x(n_1, n_2)\, z_{1i}^{-n_1} z_{2ij}^{-n_2} = \sum_{n_2=0}^{N_2-1} g(z_{1i}, n_2) z_{2ij}^{-n_2} \tag{1.1.27}$$

$$i = 0, 1, \ldots, N_1 - 1 \quad , \qquad j = 0, 1, \ldots, N_2 - 1$$

where

$$g(z_{1i}, n_2) = \sum_{n_1=0}^{N_1-1} x(n_1, n_2)\, z_{1i}^{-n_1} , \qquad n_2 = 0, 1, \ldots, N_2 - 1. \tag{1.1.28}$$

For constant i (for each i-th vertical line), we have a 1-D interpolation problem with N_2 points. Clearly, we have a system of N_2 linear equations, one equation per sample point on the i-th line. For fixed value of i, Eq. (1.1.27) can be written in matrix notation as

$$\hat{\mathbf{X}}_i = \mathbf{V}_{2i}\,\mathbf{g}_i \tag{1.1.29}$$

where $\hat{\mathbf{X}}_i$ and $\mathbf{g}_i$ are $N_2 \times 1$ column vectors, and $\mathbf{V}_{2i}$ is the i-th $N_2 \times N_2$ Vandermonde matrix as given in (1.1.25). Therefore, the vectors $\mathbf{g}_i$ are directly computed from

$$\mathbf{g}_i = \mathbf{V}_{2i}^{-1}\hat{\mathbf{X}}_i \tag{1.1.30}$$

and this is repeated for each value of i, $i = 1, \ldots, N_1$. In other words, $\mathbf{g}_i$ is the 1-D inverse NDFT of $\hat{\mathbf{X}}_i$. Once the coefficients $g(z_{1i}, n_2)$ have been calculated, the sequence $x(n_1, n_2)$ is calculated from (1.1.28) which in matrix form is written as

$$\mathbf{q}_{n2} = \mathbf{V}_1\mathbf{x}_{n2} \tag{1.1.31}$$

where $\mathbf{q}_{n2}$ is the $(N_1 \times 1)$ n_2-th row of $g(z_{1i}, n_2)$, $\mathbf{x}_{n2}$ is the n_2-th row of $x(n_1, n_2)$, and $\mathbf{V}_1$ is the $N_1 \times N_1$ Vandermonde matrix as given in (1.1.24). Hence,

$$\mathbf{x}_{n2} = \mathbf{V}_1^{-1}\,\mathbf{q}_{n2} \tag{1.32}$$

and this is repeated for each value of n_2, $n_2 = 0, 1, \ldots, N_2\text{-}1$, in order to obtain the whole sequence $x(n_1, n_2)$.

Therefore, the inverse NDFT (INDFT) in the case of nonuniformly spaced samples taken on nonuniformly spaced vertical (or horizontal) lines is computed with a total of N_1 1-D N_2-point INDFT for the column operations and N_2 1-D N_1-point INDFT for the row operations. The process is the same as the FFT method by row-column decomposition. Solving the linear systems (1.1.30) N_1 times and (1.1.32) N_2 times using Gaussian elimination, the number of operations involved is $O(N_1N_2^3 + N_2N_1^3)$.

B) Nonuniformly spaced rectangular grid.

Imposing an additional constraint that the samples in z_2 direction have to be at the same locations on each vertical line, a nonuniform grid in the (z_1, z_2) space is produced. When

taken on the unit surface ($|z_1|=1$, $|z_2|=1$), the sample locations are vertices of a rectangular grid in (ω_1, ω_2) plane, Fig. 1.2.

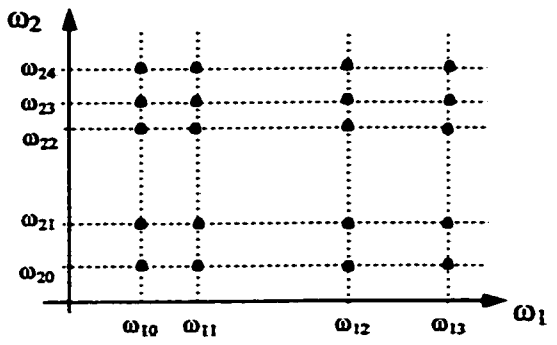


Figure 1.2 2-D NDFT with samples at the vertices of a nonuniformly spaced rectangular grid for $N_1 = 4$ and $N_2 = 5$.

For a 2-D sequence $x(n_1, n_2)$ of size $N_1 \times N_2$, the samples are taken at (z_{1k}, z_{2l}), $k = 0, 1, \ldots, N_1\text{-}1$, $l = 0, 1, \ldots, N_2\text{-}1$. Now, Eq. (1.1.14) can be expressed in a matrix form as

$$\hat{\mathbf{X}} = \mathbf{V}_1 \mathbf{X} \mathbf{V}_2^T \tag{1.1.33}$$

where

$$\hat{\mathbf{X}} = \begin{bmatrix} \hat{X}(0,0) & \hat{X}(0,1) & \cdots & \hat{X}(0,N_2-1) \\ \hat{X}(1,0) & \hat{X}(1,1) & \cdots & \hat{X}(1,N_2-1) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{X}(N_1-1,0) & \hat{X}(N_1-1,1) & \cdots & \hat{X}(N_1-1,N_2-1) \end{bmatrix} \tag{1.1.34}$$

Here,

$$\hat{X}(k, l) = X(z_1, z_2)\big|_{z_1=z_{1k},\ z_2=z_{2l}}, \quad \begin{matrix} k = 1, 2, .., N_1 - 1, \\ l = 1, 2, .., N_2 - 1. \end{matrix} \tag{1.1.34a}$$

and

$$\mathbf{X} = \begin{bmatrix} x(0,0) & x(0,1) & \cdots & x(0,N_2-1) \\ x(1,0) & x(1,1) & \cdots & x(1,N_2-1) \\ \vdots & \vdots & \ddots & \vdots \\ x(N_1-1,0) & x(N_1-1,1) & \cdots & x(N_1-1,N_2-1) \end{bmatrix} \tag{1.1.35}$$

As before, $\mathbf{V}_1$ and $\mathbf{V}_2$ are Vandermonde matrices of sizes $N_1 \times N_1$ and $N_2 \times N_2$, respectively.

$$\mathbf{V}_i = \begin{bmatrix} 1 & z_{i0}^{-1} & z_{i0}^{-2} & \cdots & z_{i0}^{-(N_i-1)} \\ 1 & z_{i1}^{-1} & z_{i1}^{-2} & \cdots & z_{i1}^{-(N_i-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{i,N_i-1}^{-1} & z_{i,N_i-1}^{-2} & \cdots & z_{i,N_i-1}^{-(N_i-1)} \end{bmatrix}, \qquad i = 1, 2 \tag{1.1.36}$$

The equivalent NDFT matrix is a direct matrix product of $\mathbf{V}_1$ and $\mathbf{V}_2$

$$\mathbf{D} = \mathbf{V}_1 \otimes \mathbf{V}_2 \tag{1.1.37}$$

and according to property (1.1.21)

$$\det(\mathbf{D}) = \left[\det(\mathbf{V}_1)\right]^{N_2}\left[\det(\mathbf{V}_{2i})\right]^{N_1} = \prod_{i \neq j,\ i>j} \left(z_{1i}^{-1} - z_{1j}^{-1}\right)^{N_2} \prod_{m \neq n,\ m>n} \left(z_{2m}^{-1} - z_{2n}^{-1}\right)^{N_1} \tag{1.1.38}$$

Clearly, the matrix is nonsingular provided that the sampling point coordinates are distinct, i.e., when $z_{1i} \neq z_{1i}$ for $i \neq j$, and $z_{2m} \neq z_{2n}$ for $m \neq n$. Hence, the inverse NDFT exist and is unique or, in other words, the 2-D interpolation problem has unique solution. The computing of the INDFT in this case involves the solution of two separate systems of linear equations of sizes N_1 and N_2, respectively. Using Gaussian elimination, the number of operations is on the order of $N_1^3 + N_2^3$. In the general case the operations involved in the solution of INDFT are $O(N_1^3\, N_2^3)$.

1.2 Overview of Bivariate Polynomial Interpolation Theory

As we mentioned in Section 1.1, the inverse NDFT problem is equivalent to a 2-D polynomial interpolation problem. The transfer function of a 2-D FIR digital filter is a 2-D polynomial of finite order. The purpose of this section is to select some appropriate tools for filter design among the vast amount of theoretical results of approximation theory. Unlike the 1-D case, the bivariate polynomial interpolation is a nontrivial task. The question is in which cases the 2-D polynomial can be reconstructed from a finite number of nonuniform frequency samples. In which cases this reconstruction is unique? A number of theoretical results on univariate and especially on bivariate interpolation theory will be reviewed. Also, another purpose of this section is to introduce and equate terminology.

The univariate approximation by polynomial interpolation rests mainly on two theorems. The first one is the classical interpolation theorem and it states that [3, 5]

Theorem 1.1 *Given $n+1$ distinct (real or complex) points $z_0, z_1, \ldots, z_n$ and $n+1$ (real or complex) values $f_0, f_1, \ldots, f_n$. There exists a unique polynomial $p_n(z) \in \mathcal{P}_n$ for which*

$$p(z_i) = f_i \qquad i = 0, 1, \ldots, n. \tag{1.2.1}$$

$\mathcal{P}_n$ denotes the class of polynomials (a linear space) of degree $\leq n$. The proof is based on the Vandermonde's determinant factored form. Since $p_n(z)$ is a polynomial of degree $\leq n$, it may be expressed as $p(z) = \sum_{k=0}^{n} c_k z^k$. The interpolation conditions (1.21) written out in matrix form now become

$$\begin{bmatrix} 1 & z_0 & z_0^2 & \cdots & z_0^n \\ 1 & z_1 & z_1^2 & \cdots & z_1^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_n & z_0^2 & \cdots & z_n^n \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} f_0 \\ f_1 \\ \vdots \\ f_n \end{bmatrix} \tag{1.2.2}$$

where the c's are the unknowns. This system has a unique solution because the coefficient matrix is nonsingular. It can be shown [3] that the determinant of this matrix (Vandermonde's determinant) can be factored in the form

$$V = \prod_{0 \le i < j \le n} (z_i - z_j) \tag{1.2.3}$$

From this formula it is clear that $V \neq 0$ if and only if the points z_i are distinct.

The second fundamental theorem in the 1-D polynomial approximation is the Weierstrass approximation theorem of 1885, [3].

Theorem 1.2. *Let $f(x)$ be a continuos function defined on $[a, b]$. For given an $\varepsilon > 0$, it is possible to find a polynomial $p(x)$ of sufficiently high degree for which*

$$|f(x) - p(x)| \le \varepsilon, \qquad a \le x \le b \tag{1.2.4}$$

The Weierstrass' theorem asserts the possibility of uniform approximation by polynomials to continuous (not necessarily analytic) functions over a closed interval.

The ordinary polynomials are simply linear combinations of the monomials $1, x, x^2, \ldots, x^n$. A continuous function $f(x)$ on $[a, b]$ can be approximated also by a linear combination of other fixed functions $\phi_0, \phi_1, \ldots, \phi_n$ on the same interval $[a, b]$. Their linear combinations $\sum_{i=0}^{n} c_i \phi_i$ are termed generalized polynomials. The problem of best approximation by generalized polynomials requires the Haar condition [3]:

Given a system of $n+1$ functions $\{\phi_0, \ldots, \phi_n\}$ continuous on some fixed metric space X. This system is said to satisfy the *Haar condition* if every set of $n+1$ vectors of the form

$$\Phi(x) = \begin{bmatrix} \phi_0(x) & \phi_1(x) & \cdots & \phi_n(x) \end{bmatrix}^T \tag{1.2.5}$$

is independent for any choice of distinctly different x. Expressed otherwise , the determinant

$$\Delta(x_0, x_2 \ldots, x_n) = \begin{vmatrix} \phi_0(x_0) & \cdots & \phi_0(x_n) \\ \cdots & \cdots & \cdots \\ \phi_n(x_0) & \cdots & \phi_n(x_n) \end{vmatrix} \tag{1.2.5a}$$

is zero only when two of the points x_i coincide in the fixed compact metric space X. The non-zero Vandermonde determinant (1.2.3) implies that $\{1, x, x^2, \ldots, x^n\}$ satisfies the Haar condition on any interval and for any n. A system of functions satisfying the Haar condition is called a Chebyshev system. The Chebyshev polynomials of first and second kind form a Chebyshev system, $x \in [-1, 1]$. Another example of a Chebyshev system is $\{\varphi_k = e^{\alpha_k x}\}_{k=0}^{n}$, for distinct α_k and $x \in (-\infty, +\infty)$.

Now it will be pertinent to reproduce the central theorem of univariate best approximation, the alternation theorem [3, 13, 16].

Theorem 1.3. *Let* $\{\phi_0, \ldots, \phi_N\}$ *be a system of* $N + 1$ *functions of* $C[a, b]$ *satisfying the Haar condition, and let* X *be any closed subset of* $[a, b]$. $P(x) = \sum_{i=0}^{N} c_i \phi_i(x)$ *denotes a certain generalized polynomial on* X. *Also,* $f(x)$ *is a continuos function on* X *to be approximated and* $w(x)$ *is a positive and continuous on* X. *The weighted error function is given by*

$$E(x) = w(x)[f(x) - P(x)] \tag{1.2.6}$$

and the error norm is the weighted maximum error

$$\|E\| = \max_{x \in X} |E(x)| = \delta \tag{1.2.7}$$

A necessary and sufficient condition that $P(x)$ *is the unique polynomial that minimizes* $\|E\|$ *is that the error function* $E(x)$ *exhibits on* X *at least* $N+2$ *alternations, thus:*

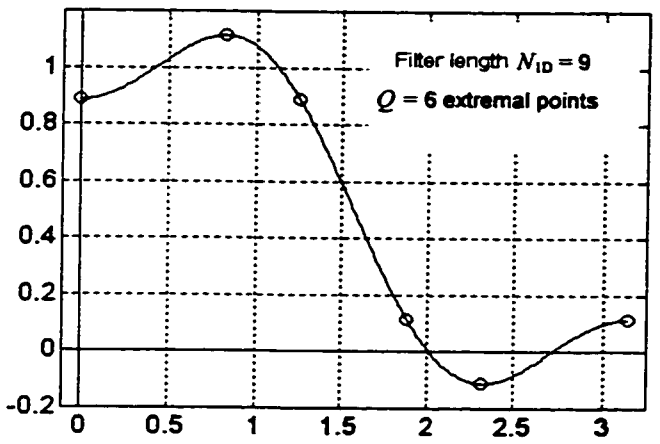
$E(x_i) = -E(x_{i+1}) = \pm\delta$, *for* $i = 0, 1, 2, \ldots, N+1$, *and* $x_i \in X$.

The alternation theorem means that there is a *unique* best approximation for a given set of points (frequencies), filter length N_{1D}, and weight function $w(x)$. It also states that the best Chebyshev approximation must necessarily have an equiripple error function. In the

optimum approximation of 1-D FIR filters Chebyshev polynomials are used as basis functions with $x = \cos\omega$:

$$\phi_k(x) = T_k(\cos\omega) = \cos(k\omega), \qquad k = 0, 1, ..., N. \tag{1.2.8}$$

The approximation polynomial $P(x)$ is a sum of these weighted cosines. The number of the extremal points is at least $N+2$ and is related to the filter length N_{1D}. A subset of $N+1$ of these points are sufficient to reconstruct the N-th order polynomial $P(x)$. These $N+1$ extremal points are in fact frequency samples trough which the filter amplitude response $A(\omega)$ is constructed [18].



$$P(\cos\omega) = \sum_{k=0}^{4} a_k \cos(\omega)^k$$

Min. number of extremal points Q:

$Q = N + 2$, or

$$Q = \begin{cases} (N_{1D}+3)/2, & N_{1D} = \text{odd} \\ (N_{1D}+2)/2, & N_{1D} = \text{even} \end{cases}$$

Figure 1.3 Frequency response amplitude of an optimal FIR filter designed using the algorithm in [18]. Five of the six alternation frequency samples (two of them are at the band edges) have been used for the interpolation.

The amplitude response $A(\omega) = P(\cos\omega)$ of the length 9 FIR minimax filter, shown in Fig. 1.3, is obtained with a linear combination of five cosine basis functions. The amplitude function is the analytic (real-valued) version of the magnitude response $M(\omega)$, $A(\omega) = \pm M(\omega)$.

Unfortunately, there is an essential difficulty in the extension of the Chebyshev approximation theory to functions of more than one variable. As it has been shown [4], there are no universal Chebyshev sets of functions of two and more variables that can be used for interpolation in 2-D or higher dimensions. In other words, the Haar condition is

not in general satisfied. There are, however, some special metric spaces X, on which a Chebyshev system of continuous functions can be defined. For example, this is true when X is homeomorphic to a subset of a circle [4]. Such special cases are of little interest in practice. The absence of Chebyshev sets means that, in general, the polynomials are not specified uniquely by samples at arbitrary locations. Furthermore, it may happen that a solution to the interpolation equations does not exist. The nonuniqueness is not too much of an issue, because one of the many possible solutions is enough (provided all the solutions are optimal in Chebyshev sense). Finite point sets can be considered as approximations to infinite point regions. In such cases we have a set of functions $\{\phi_i(x)\}$ that may form a Chebyshev system or may not [4].

Prior discussing optimal solutions in two and higher dimensions, we will consider some basic results from the bivariate polynomial interpolation theory. The emphasis will be on interpolation using only function sample values. The methods employing function derivatives at the sample points will not be considered because they are quite involved and, therefore, of little interest for the practice of digital filter design.

The interpolation in 2-D can be done with bivariate polynomials that are either in Π_n, the space of polynomials with total degree less or equal to n, or in $\Pi_{(n,m)}$, the space of polynomials $P(z_1, z_2)$ with maximum degree n in z_1 and m in z_2. The total degree of a 2-D polynomial $P(z_1, z_2)$ is defined to be the degree of the 1-D polynomial $P(z_1, z_1)$.

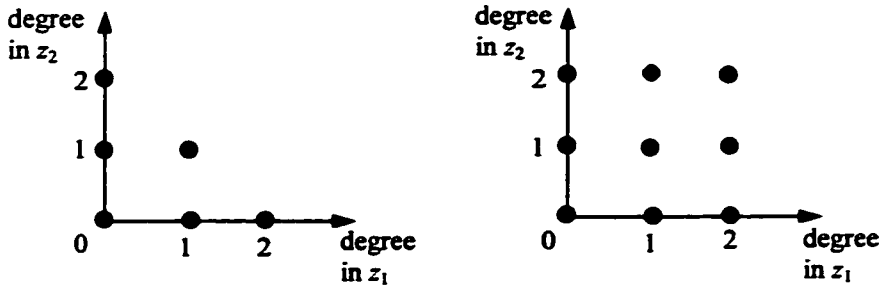


Figure 1.4 Degree of polynomials (a) in Π_2 and (b) in $\Pi_{(2,2)}$.

One important result in 2-D interpolation, that will be of interest in our problem of 2-D FIR filter design, was derived by Gasca and Maeztu in 1982, [23]. Consider a set of straight lines r_i in $\mathbf{R}^2$, each of which is associated with a polynomial of first degree in x and y, also denoted by r_i. With each line r_i a set of straight lines r_{ij} is considered in such a way that the intersections determined by r_i and r_{ij} are points at which the interpolation data (sample values) are given, denoted by u_{ij}. The lines r_i and/or r_{ij} may appear with multiplicity greater than one, leading to derivative values as interpolation data. A formulation with derivatives results in a Hermite interpolation problem. When no two lines coincide we have the particular case of a Lagrange interpolation problem and this can be stated as follows [23]:

Theorem 1.4 *If the intersection u_{ij} of r_i and r_{ij} with $i+j \neq 0$, does not lie on any of the lines*

$$r_0, \ldots, r_{i-1}, r_{i0}, \ldots, r_{ij-1} \quad \text{if } i > 0,\ j > 0,$$
$$r_0, \ldots, r_{i-1} \quad \text{if } i > 0,\ j = 0,$$
$$r_{00}, \ldots, r_{0j-1} \quad \text{if } i = 0,\ j > 0,$$

then we have defined a Lagrange interpolation problem with a unique solution.

The proof can be found in [23]. Fig. 1.5 shows an example of geometric distributions of the sampling points of Theorem 1.4.

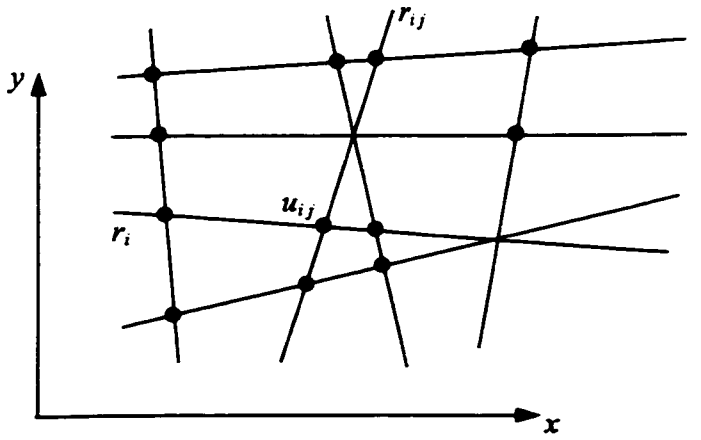


Figure 1.5 Example of geometric distribution of the sampling points of Theorem 1.4, Ref. [23].

A special case of the above theorem has been proposed earlier by Chung an Yao [30]. Their results deal with special sets (lattices) of nodes in $\mathbf{R}^m$ for which the Lagrange interpolation problem has a unique solution. An example of a so-called natural lattice is presented in Fig. 1.6.

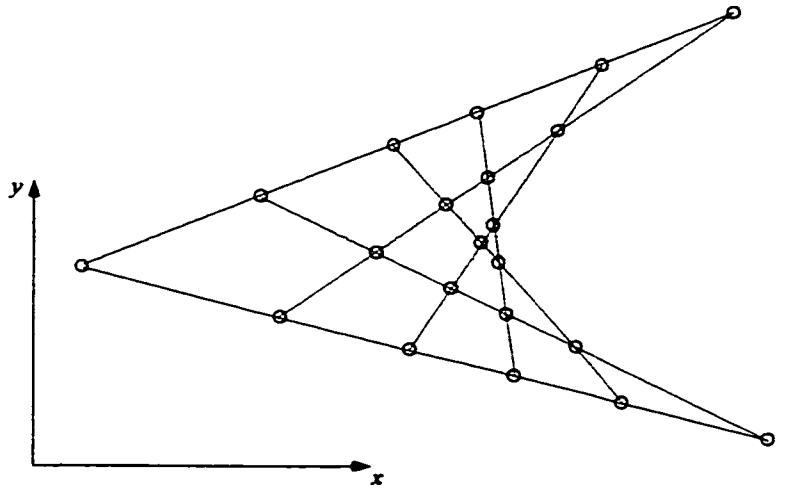


Figure 1.6 A two-dimensional example of natural lattice of sampling points of fifth order [30]..

The condition for uniqueness is that corresponding to each sampling point $x_i \in \mathbf{R}^2$, there exist k distinct straight lines (or hyperplanes in $\mathbf{R}^m$) such that (i) x_i does not lie on any of these lines (hyperplanes), and (ii) all the other sample points lie on at least one of these lines (hyperplanes). In the example above $k = 5$. Stated plainly, the above results require no sample points taken on the intersection of three or more straight lines (in $\mathbf{R}^2$) in order to guarantee unique interpolation polynomial.

A special case of the above results is the sampling on a nonuniform rectangular grid. In Sec. 1.1 it has been shown that a unique solution of the interpolation problem exists in this case.

Some results concerning sufficient conditions under which the 2-D polynomial interpolation problem has a unique or nonunique solution has been derived in [9]. These results are on interpolation in the polynomial space $\Pi_{(n,\ m)}$. The sampling points are chosen on curves with nonfactorizable equations. However, the results concerning the

unique solution are for very specific arrangement of the irreducible curves and the samples on each curve. Some conditions under which the interpolation polynomial coefficients cannot be uniquely determined are given. For completeness, these conditions will be repeated here as

Theorem 1.5, [9] *Let the bivariate interpolation polynomial be of the form*

$$P(x,y)=\sum_{i=0}^{N_x}\sum_{j=0}^{N_y}a(i,j)x^i y^j \tag{1.2.9}$$

and let the sampling points lie on p irreducible distinct curves $r_1, r_2, \ldots, r_p$. *Each of these curves* r_i *has maximum degree in x given by* $M_x^{(i)}$, *and in y given by* $M_y^{(i)}$. *If the following inequalities are simultaneously satisfied, i. e.,*

$$N_x > \sum_{i=1}^{p} M_x^{(i)} \qquad N_y > \sum_{i=1}^{p} M_y^{(i)}\,, \tag{1.2.10}$$

then the polynomial coefficients cannot be uniquely determined.

Theorem 1.6, [9] *Let the interpolation polynomial be of the form* (1.2.9) *and the* $(N_x+1)(N_y+1)$ *sampling points are taken again on irreducible curves. If there is an irreducible curve of the form*

$$x^{M_x} = ky^{M_y} \qquad or \qquad x^{M_x}y^{M_y} = k \tag{1.2.11}$$

which contains more than $M_yN_x+M_xN_y+1$ *sampling points, then the coefficients of the polynomial (1.2.9) cannot be uniquely determined.*

Proofs of the above two theorems can be found in [9]. Theorem 1.5 simply states that if the sum of the degrees of the curves on which the samples are taken is small compared to the degree of the desired polynomial, then the interpolation problem becomes singular. Theorem 1.6 states that if too many points are taken on any curve with equation of the form (1.2.11), then overspecification is done and the interpolation problem runs into singularity.

In [31] the Lagrange interpolation is extended to 2-D functions (signals) using polar coordinates. A 2-D function $f(\rho, \theta)$ is reconstructed from a set of nonuniform samples. The function is presented in polar coordinates ρ and θ, and it has a finite circular region of support.

As it was pointed out in the foreword, one of the ideas in the present thesis is to determine locations and values of the frequency samples in the (ω_1, ω_2) plane, such that the resulting continuous frequency response is as much as possible close to optimal in Chebyshev sense. That is why it would be pertinent to reproduce some terminology and results concerning the characterization of the Chebyshev approximation in 2-D and m-D.

The domain of approximation is a compact subset K of the two-dimensional frequency plane (ω_1, ω_2). For an ideal filter with piecewise constant frequency bands, K would be chosen to be the union of the passband regions R_p and stopband regions R_s, omitting the transition bands. The error criterion is the minimization of the maximum error between the desired frequency response (the function to be approximated) and the resulting filter frequency response (the 2-D interpolation polynomial). As in the 1-D case, the maximum error magnitude $|E(\omega)|$ given by

$$\|E\| = \max_{\omega \in K} |E(\omega)| = \max_{\omega \in K} \left| W(\omega)\left(H_D(\omega) - H(\omega)\right) \right| \tag{1.2.12}$$

Here $\omega = (\omega_1, \omega_2)$ is the frequency vector. The error norm $\|E\| = \delta$ is sometimes referred to as the L_∞ error criterion. The frequency response of a 2-D zero-phase FIR can be expressed as (cf. Sec. 3.1)

$$H(\omega) = H(\omega_1, \omega_2) = \sum_{(n_1, n_2) \in R_a} \sum a(n_1, n_2) \cos(\omega_1 n_1 + \omega_2 n_2) = \sum_{i=0}^{N} a_i \phi_i(\omega) \tag{1.2.13}$$

Clearly, the bivariate polynomial $H(\omega)$ can be constructed using $N+1$ real-valued basis functions $\phi_i(\omega)$,

$$\phi_i(\omega) = \phi_i(\omega_1, \omega_2) = \cos(\omega_1 n_1 + \omega_2 n_2) \tag{1.2.14}$$

where i depends on n_1 and n_2. The optimal approximation problem can be stated as: Given R_p R_s, passband tolerance $\delta_p = k\delta_s$, and stopband tolerance δ_s, determine a_i so that δ_s is minimized. Stated otherwise, *determine the set of coefficients a_i that minimizes* $\max_{\omega \in K} |E(\omega)|$ *over the compact subset K.*

In 2-D, however, the $N+1$ vectors of basis functions

$$\Phi(\omega_i) = \left[\phi_0(\omega_i) \quad \phi_1(\omega_i) \quad \cdots \quad \phi_N(\omega_i)\right]^T \qquad i = 0, 1, ..., N,$$

(cf. Eq. 1.2.5) are not always linearly independent for any $N+1$ points ω_i in K. No set of nontrivial bivariate functions satisfies the Haar condition [4, 10]. The absence of Chebyshev sets means that the alternation theorem does not hold in m-D, $m \geq 2$. There are, however not so powerful, some theorems characterizing the best approximation in 2-D. First of all, the concept of convexity should be considered.

A set of points (vectors) is said to be convex if with each two of its points it contains also the line segment containing them. In $\mathbf{R}^2$, the line segment joining points $\mathbf{v}_1$ and $\mathbf{v}_2$ consists of all points of the form $\alpha\mathbf{v}_1 + (1 - \alpha)\mathbf{v}_2$ for $\alpha \in [0, 1]$.

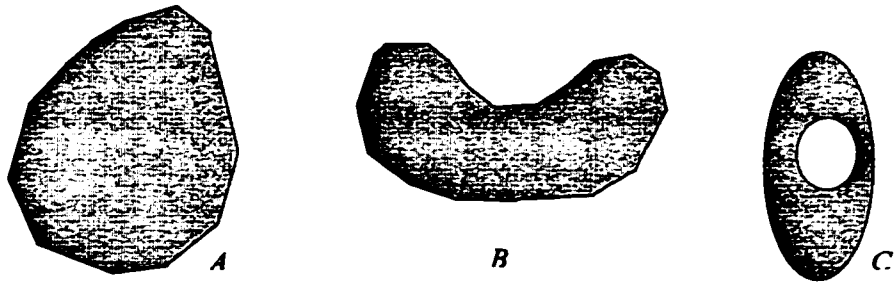


Figure 1.7 Convex and nonconvex sets: set *A* is convex, *B* and *C* are not.

In an arbitrary m-D linear space (in our case $m = N+1$), if a set of vectors $\mathbf{v}_i$ is finite, then a vector $\mathbf{v}$ is said to lie in the convex hull of $\mathbf{v}_i$ if and only if there exist constants α_i subject to three simultaneous conditions:

$$\text{(a)}\quad \mathbf{v}=\sum_{i=1}^{m}\alpha_i\mathbf{v}_i \qquad \text{(b)}\quad \alpha_i \geq 0 \quad \text{for all } i \qquad \text{(c)}\quad \sum_{i=1}^{m}\alpha_i = 1 \tag{1.2.15}$$

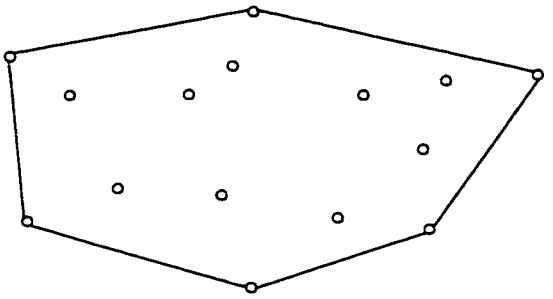


Figure 1.8 Convex hull of a finite set of discrete points.

The convex hull of a set of points lying in a 2-D plane may be "found" by driving nails in at each point and wrapping a string around the configuration, as it is illustrated in Fig. 1.8.

An extremal point is a point $\omega_i = (\omega_1, \omega_2)$ in K where the error functions attains its norm, given by (1.2.12). The extremal point set is the set of all such extremal points. The characteristic vector associated with a particular point ω_i is the vector $\Phi(\omega_i)$, see above.

A **critical point set** is defined as a set of extremal points of minimum size such that the zero vector lies in the convex hull of the of the signed characteristic vectors $\sigma(\omega_i)\Phi(\omega_i)$ where

$$\sigma(\omega_i)=\begin{cases}-1, & \text{if } H(\omega_i)-H_D(\omega_i)<0\\ 1, & \text{if } H(\omega_i)-H_D(\omega_i)>0\end{cases} \tag{1.2.16}$$

By minimum size it is meant that if any point were removed from the set, the zero vector (the origin of the m-D space) would no longer lie within the convex hull of the reduced set.

Theorem 1.7 *A critical point set contains* $p \leq N + 2$ *critical points.* ($N + 1$ *is the number of basis functions* $\phi_i(\omega)$), [3, 4, 21].

This theorem limits the dimensionality of the problem. The originally stated approximation problem can be replaced by a search for a critical point set of $p \leq N + 2$ members. In the 1-D case, the alternation theorem asserts that the unique solution necessitates at least N+2 alternations. The choice of p in the second Remez exchange algorithm [18] is $p \leq N + 2$. However, in the 2-D case all what is known is that $p \leq N + 2$ and typically $p = N + 2$ is chosen for the practical iterative algorithms [21, 22]. At each iteration $N + 2$ points (frequency samples) are necessary for solving a set of $N + 2$ linear equations in order to determine the error deviation δ and the candidate $N + 1$ coefficients a_i. These equations possess a unique solution. The case when $p < N + 2$ is called the degeneracy case.

The next theorem is a characterization theorem for the optimal approximation [3, 4, 21].

Theorem 1.8 *$H(\omega_1, \omega_2)$ is a best approximation to $H_D(\omega_1, \omega_2)$ in the Chebyshev sense if and only if the zero vector lies in the convex hull of a critical point set associated with the extremal points of the error function.*

Proofs of theorems 1.7 and 1.8 can be found, for example, in Rice [4] and Cheney [3]. The best Chebyshev approximations are not uniquely determined. Additional criteria are considered which separate one of the best approximations as the "best of the best" which is called also the strict approximation [4]. The strict approximation is unique. Several iterative techniques for the design of optimal 2-D linear phase FIR digital filters have been developed [19, 20, 21, 22]. The methods based on linear programming used by Hu and Rabiner[20] and improved by Fiasconaro [19] are usually very slow. Much faster methods using single-exchange and multiple-exchange ascent algorithms have been proposed by Kamp and Thiran [22], Hersey and Mersereau, and Harris [21]. The mathematical details of these algorithms, which are out of the scope of this thesis, can be found in [22, 4, 3, 21]. The ascent algorithms include a search for the local maxima of the error function magnitude. This is much more involved in 2-D than in the 1-D case. The 2-D function $H(\omega_1, \omega_2)$ can have ridges or nearly flat edges along which the error function may not vary much. That is why it must be carefully searched in many directions. In contrast to the 1-D case, in the 2-D case not all the critical points are replaced in each

iteration, and this increases the number of iterations. The 2-D iterative algorithms of the Remez exchange type developed so far are very expensive computationally, and have not been demonstrated to reliably converge to a correct solution. Developing a computationally efficient algorithm to design 2-D optimal filters remains an area for research [10].

Some important elements of the above-mentioned techniques deserve more attention as far as the filter design by nonuniform frequency samples is concerned. First, in order to determine the best approximation digitally, the continuos domain of approximation K is represented by a finite set of discrete points K'. These are samples of K lying on a Cartesian grid. It has been found that a grid density sufficient to sample the highest frequency basis function ten to twenty times per period in each spatial frequency will adequately represent the continuum [21]. Also, points of the discrete set K' are located along the edges of the transition regions. On the second place, although the Haar condition is not satisfied in the continuos domain K, as a practical matter it is true that N+1 characteristic vectors randomly chosen from the discrete domain K' are nearly always linearly independent. If for a given choice of the N+1 frequency points $\omega_i = (\omega_{1i}, \omega_{2i})$ the characteristic vectors are not linearly independent, then one or more values of ω_i are perturbed slightly to remove the degeneracy.

At the end, a few words should be said about the possible arrangements of the positive and negative sampling points in a critical point set. The positive and negative points of the set of extremal points of an approximation $H(\omega)$ to $H_D(\omega)$ are, respectively

$$\mathcal{P} = \{\omega \mid \omega \in K, \quad H(\omega) - H_D(\omega) = \max |H(\omega) - H_D(\omega)| \},$$
$$\mathcal{N} = \{\omega \mid \omega \in K, \quad H(\omega) - H_D(\omega) = -\max |H(\omega) - H_D(\omega)| \}. \qquad (1.2.17)$$

In the one-dimensional case, the nature of the critical point set is simply described. The Chebyshev approximation is characterized by at least $N + 2$ extremal (alternating) points of the error function. This set contains a subset of N + 1 points in the order

● ○ ● ○ ● ○ ● ○

where ○, ● denote positive and negative points, respectively. There is no simple geometric identification of a critical point set in two or higher dimensions. Some possible arrangement of positive and negative points in a critical point set for approximation by low degree polynomials and rational functions are illustrated in Fig. 1.9, [4].

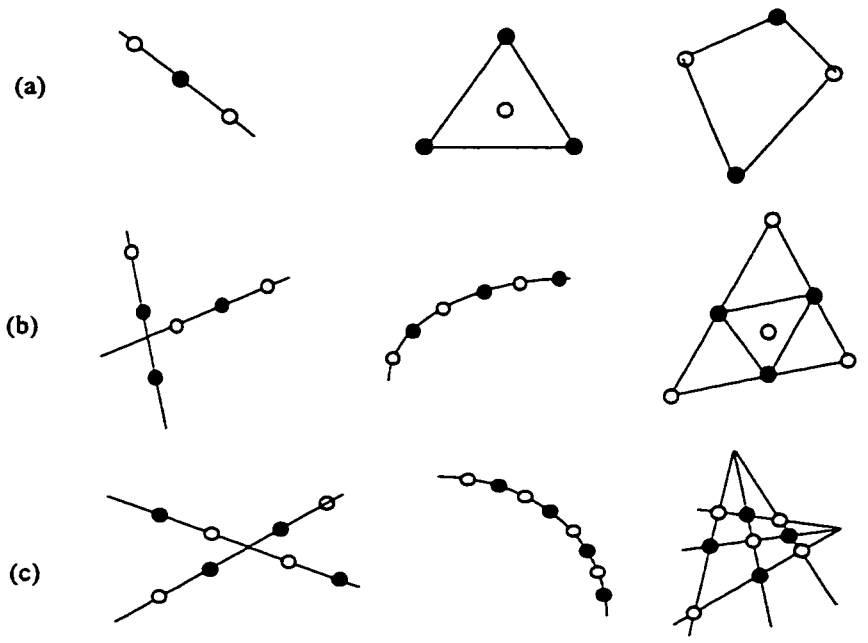


Figure 1.9 Some critical point sets for approximation in 2-D by polynomials or by rational functions [4]

For approximation in 2-D and higher dimensions, approximating functions which are tensor products can be used. The critical point sets in a product space must be products in a certain sense of critical point sets in the spaces entering into the product. Some examples are illustrated in Fig.1.10.

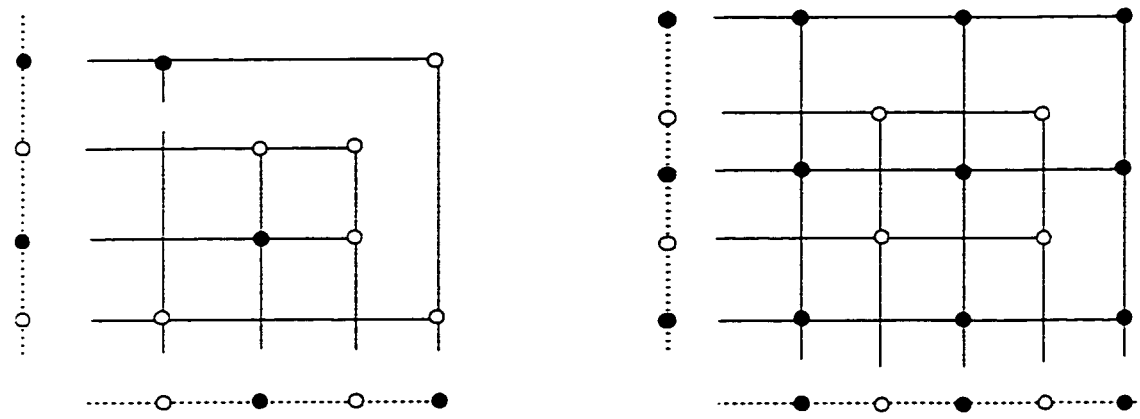


Figure 1.10 Some 2-D critical point sets formed by tensor products. The points on the dotted lines only indicate the construction and are not part of the sets.

Chapter 2

Rectangular Shape FIR Filter Design

In this chapter we will consider nonuniform sampling techniques for designing 2-D zero-phase FIR filters with rectangular shape of their frequency response. In Section 2.1, existing methods for designing 2-D FIR filters with uniform as well as with nonuniform frequency sampling are briefly discussed. Existing symmetries in the frequency response and impulse response, tolerance scheme specifications, and general solution to the 2-D interpolation problem for rectangular FIR filters are discussed in Section 2.2. In Section 2.3, the arbitrary sampling technique is investigated. In Section 2.4, a sampling technique based on exponentially distributed parallel lines is proposed and demonstrated. Several sampling techniques based on sampling at the vertices of a nonuniformly spaced rectangular grid are proposed in Section 2.5. These techniques employ grid lines obtained using harmonic series, exponential functions, 1-D extremal frequencies, and Chebyshev polynomials.

2.1 Existing Methods for 2-D Frequency Sampling FIR Filter Design .

The uniform frequency sampling technique for designing 2-D FIR filters is the most straightforward extension of the corresponding 1-D case. It has been shown by Hu and Rabiner [20] in 1972 how the powerful FFT can be used in the 2-D frequency sampling design method. The method consists of taking samples at the vertices of a Cartesian grid

in the 2-D frequency plane (ω_1, ω_2). The frequency samples $H'(k_1, k_2)$ are simply the coefficients of the 2-D DFT of the shifted desired filter $h_D(n_1, n_2)$, see Eq. (2.1.6). The shift is due to the fact that the discrete Fourier transform is defined only for first-quadrant support sequences, that cannot be zero phase. To make that more clear, let's begin with the transfer function of the desired zero phase filter.

$$H_D(z_1,z_2) = \sum_{n_1=-\infty}^{+\infty} \sum_{n2=-\infty}^{+\infty} h_D(n_1,n_2) z_1^{-n_1} z_2^{-n_2} \tag{2.1.1}$$

This filter is stable and its Fourier transform exists iff

$$\sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} |h_D(n_1,n_2)| < \infty \tag{2.1.2}$$

The desired frequency response is (2.1.1) evaluated on the unit surface ($z_1=e^{j\omega 1}$, $z_2=e^{j\omega 2}$):

$$H_D(\omega_1,\omega_2) = \sum_{n_1=-\infty}^{+\infty} \sum_{n2=-\infty}^{+\infty} h_D(n_1,n_2) e_1^{-j\omega_1 n_1} e_2^{-j\omega_2 n_2} \tag{2.1.3}$$

Now, if we want to approximate this desired (ideal) filter by an FIR filter with $N_1 \times N_2$-point impulse response $h(n_1, n_2)$, where $N_1 = 2\ M_1+1$, and $N_2 = 2\ M_2+1$, the impulse response of the ideal filter has to be shifted by M_1 points in n_1 direction, and by M_2 points in n_2 direction.

$$h'_D(n_1,n_2) = h_D(n_1 - M_1, n_2 - M_2) \tag{2.1.4}$$

Using the shifting property of the Fourier transform, the frequency response of the shifted version is given by

$$H'_D(\omega_1,\omega_2) = e_1^{-j\omega_1 M_1} e_2^{-j\omega_2 M_2} H_D(\omega_1,\omega_2) \tag{2.1.5}$$

The frequency samples can be expressed as the coefficients of an $N_1 \times N_2$-point DFT of the

shifted ideal impulse response (2.1.3).

$$H'(k_1,k_2) = H'_D(\omega_1,\omega_2)\Big|_{\omega_1=\frac{2\pi}{N_1}k_1,\,\omega_1=\frac{2\pi}{N_1}k_1} = H_D(\omega_1,\omega_2)e_1^{-j\omega_1 M_1}e_2^{-j\omega_2 M_2}\Big|_{\omega_1=\frac{2\pi}{N_1}k_1,\,\omega_1=\frac{2\pi}{N_1}k_1},$$

$$0 \le k_1 \le N_1 - 1,\ 0 \le k_2 \le N_2 - 1$$

(2.1.6)

The corresponding sequence $h'(n_1, n_2)$ is obtained by the application of the inverse discrete Fourier transform to (2.1.6), which gives

$$h'(n_1,n_2) = \frac{1}{N_1N_2}\sum_{k_1=0}^{N_1-1}\sum_{k_2=0}^{N_2-1} H'(k_1,k_2)\exp\left(j\frac{2\pi}{N_1}k_1n_1\right)\exp\left(j\frac{2\pi}{N_2}k_2n_2\right) \tag{2.1.7}$$

At the end, the designed filter is obtained by shifting (2.1.7) back.

$$h(n_1,n_2) = h'(n_1 + M_1, n_2 + M_2) \tag{2.1.8}$$

The frequency response of the designed filter is given by

$$H(\omega_1,\omega_2) = \sum_{n_1=-M_1}^{M_1}\sum_{n2=-M_2}^{M_2} h(n_1,n_2)\,e_1^{-j\omega_1 n_1}e_2^{-j\omega_2 n_2} \tag{2.1.9}$$

From the frequency samples, a direct interpolation formula for the frequency response can be obtained by plugging the shifted version (2.1.8) of (2.1.7) into (2.1.9), giving

$$H(\omega_1,\omega_2) = \sum_{n_1=-M_1}^{M_1}\sum_{n_2=-M_2}^{M_2}\left[\frac{1}{N_1N_2}\sum_{k_1=0}^{N_1-1}\sum_{k_2=0}^{N_2-1} H'(k_1,k_2)\exp\left(j\frac{2\pi}{N_1}k_1(n_1+M_1)\right)\exp\left(j\frac{2\pi}{N_2}k_2(n_2+M_2)\right)\right] .\exp(-j\omega_1 n_1)\exp(-j\omega_2 n_2) \tag{2.1.10}$$

Interchanging the order of summation, and summing over the n_1 and n_2, produces

$$H(\omega_1,\omega_2)=\frac{e^{-j\omega_1 M_1}e^{-j\omega_1 M_2}\left(1-e^{-j\omega_1 N_2}\right)\left(1-e^{-j\omega_1 N_2}\right)}{N_1 N_2} \cdot \sum_{k_1=0}^{N_1-1}\sum_{k_2=0}^{N_2-1}\frac{H'(k_1,k_2)}{\left(1-e^{j(2\pi/N_1)k_1}e^{-j\omega_1}\right)\left(1-e^{j(2\pi/N_2)k_2}e^{-j\omega_2}\right)} \tag{2.1.11}$$

Equation (2.1.11) is the basis of the 2-D FIR filter design with *uniform* frequency samples. In the design of piecewise constant filters, the ideal frequency response changes sharply from one to zero or vice versa. These sharp transitions cause large deviations in both passbands and stopbands. This can be considerably improved by introducing transition bands and samples in these bands. The values of these samples can be chosen so as to minimize the maximum approximation error δ_s in the stopband(s) and $k\delta_p$ in the passband(s). this is not a trivial problem and requires linear programming methods, [19, 20], which are computationally intensive. The problem can be considerably simplified, and will work well [10], if the sample values in the transition band(s) are obtained by linear interpolation.

The uniform frequency sampling gives the least error control in both passbands and stopbands. This is due to the inherent disadvantage of the method: the lack of flexibility in choosing the frequency sample locations. For a fixed filter size the frequency samples appear always at the same locations, regardless of the filter shape, e.g., square, circular, fan, etc. As a consequence, the filters deviate from the desired passband shape, especially when the support size of the impulse response is small.

The idea of 2-D FIR filter design by nonuniform frequency sampling is not new. Here some of the most interesting approaches will be considered. The details of these techniques will be given in the next section during the comparison with the proposed methods. Rozwod, Therrien, and Lim [6] proposed a method for nonuniform frequency sampling design of 2-D FIR filters in which the locations of the frequency samples are constrained to be on parallel vertical or horizontal lines in the (ω_1, ω_2) plane. This case has been considered in Sec. 1.1. As it has been shown there, this sampling technique reduces the large 2-D system of linear equations to several smaller 1-D systems. Along with the reduction of computational complexity, the method guarantees existence and

uniqueness of a solution provided that the sampling points on the same line are distinct. However, no efficient algorithm for the sample locations on each line and the line locations has been offered. Angelidis [7] improved this method. Using the same constraints on the sampling locations, he improved the computational efficiency by introducing a Newton polynomial representation of the filter's transfer function. This representation guarantees accurate solutions even in cases of high-order filters or when the interpolation matrix is ill-conditioned. This technique is a special case of the corollary given in [9], see next. Two design examples are given, one circular lowpass filter, and one circular bandpass filter.

Zakhor and Alvstad [9] apply a number of theoretical results to the problem of nonuniform frequency sampling design of 2-D FIR filters. Some conditions are given, under which the interpolation problem might become singular. Specifically, these conditions concern the sum of degrees of the curves on which the sampling points are chosen and the number of the points on each curve, Sec. 1.2, Theorem 1.5 and Theorem 1.6. Also, a corollary is given providing an exact description of the distributions of the frequency sampling points required for unique specification of the filter coefficients. The corollary specifies the number of the sampling lines and their slopes, the distributions of the sampling lines in the frequency plane, and the distributions of the samples on each line. A recursive algorithm is proposed for computing the 2-D polynomial coefficients. This recursive algorithm can only be applied to lines of identical slope in the (ω_1, ω_2) plane. The authors design a circularly symmetric FIR filters using this approach and via a linear least squares (LLS) approach which involves the solution of an overdetermined system of linear equations. Actually, the recursive approach is an LLS fit. In the first case the samples are taken on straight lines in the 2-D frequency plane with slope -1, while in the LLS case the lines have slopes $\pm$ 1. Common feature of the designs shown in [9] is that samples are taken at the intersections of the sampling lines and the pass- and stopband edges. Also, a few samples are taken in the transition regions with values chosen linearly from 0 to 1. The shape of the designed filters deviates from circular due to the fact that the samples are on straight lines and cannot match exactly the circular shape.

Another feature is the huge number of sampling points, for example 534 frequency samples for only 36 independent filter coefficients, filter size 15×15 points. The resulting approximation error can be found in the comparison table in Chapter 3.

Mitra *et al.* [1], and Bagchi and Mitra [2] have generalized the definition of the DFT and introduced the NDFT (see Sec.1.1). They showed how this framework of nonuniform frequency sampling can be used in the design of 2-D FIR filters. Their examples include the design of nonseparable filters with different shapes: square, circular, diamond, and fan. It has been shown how the number of the frequency samples can be decreased using the symmetry properties of these filters' coefficients. Using the freedom of the nonuniform sampling, the frequency samples are placed on contour lines that match the desired passband shape, e.g., for the design of a square-shaped filter, the samples are taken on a set of square contours in the (ω_1, ω_2) plane. The result is a closer match to the desired shape. The sample values of the desired filter are obtained by approximating a cross-section of the 2-D frequency response by 1-D analytic functions derived from Chebyshev polynomials. The design is completed by solving a system of L linear equations for the L independent filter coefficients. Though some guidelines have been given for the choice of the number of samples and their locations, there is nothing said about the minimum distance between the sampling contours and the maximum sample density.

Angelidis and Diamessis [14] proposed a method for designing 1-D FIR filters from nonuniform frequency sampling. The method is based on an interpolation polynomial of Newton type and works well with complex values of the frequency samples. in other words, nonlinear phase of the desired filter can be specified and supplied to the algorithm through the sample values. The polynomial coefficients are calculated recursively and the method is numerically well-conditioned. Based on this Newton type polynomial, Angelidis [7] extended the method to 2-D. The frequency samples are taken in the first quadrant of the frequency plane at the vertices of a nonuniformly spaced rectangular grid. In this way, the 2-D problem of the filter coefficients calculation is reduced to 1-D

formulas, like the case considered in Sec. 1.1., case B. Here, the number of arithmetic operations is reduced because two 1-D triangular systems are solved. Since the polynomial is of Newton type, the coefficients are permanent which means that in cases where an additional row of frequency samples is taken, the filter coefficients are obtained by calculating only the new coefficients and updating the old ones. As in [14], the frequency samples can be complex numbers. i.e., the phase of the desired filter can be specified.

Angelidis [8] further improved this method and applied it the case when the frequency samples are taken arbitrary along vertical or horizontal lines in the (ω_1, ω_2) plane, as in [6]. The 2-D interpolation problem is divided into several 1-D problems. As it was described in Sec. 1.1, case A, in such an sample arrangement, the solution is guaranteed to exist an it is unique. Again, the method is recursive, fast, and guarantees accurate solutions even in cases of high-order filter design when the interpolation matrix is ill-conditioned. As before, in case of new samples, the old interpolation coefficients are updated and only the new coefficients are calculated.

2.2 Rectangular shape FIR filters: symmetry constraints

The 2-D finite impulse response (FIR) digital filters have impulse response $h(n_1, n_2)$ which is of a finite extent. Therefore, $h(n_1, n_2)$ is always absolutely summable and FIR filters are always stable. Among the other advantages over the infinite-extent impulse response (IIR) filters can be pointed out the ability of the FIR filters to attain purely real frequency responses. Such filters are termed zero phase filters. The 2-D FIR can be efficiently realized, for example, trough a high-speed convolution using the FFT.

Similar to 1-D digital filters, 2-D digital filters are generally specified in the frequency

domain. The frequency response of a 2-D digital filter is periodic with period 2π in both spatial frequencies ω_1 and ω_2, i.e.,

$$H(\omega_1, \omega_2) = H(\omega_1 + 2\pi, \omega_2) = H(\omega_1, \omega_2 + 2\pi) \qquad \text{for all } (\omega_1, \omega_2) \qquad (2.2.1)$$

Therefore, $H(\omega_1, \omega_2)$ would be completely specified if known in the region $-\pi \le \omega_1 \le \pi$, $-\pi \le \omega_2 \le \pi$.

In many applications, for example image processing, a zero phase characteristic is needed. A zero phase filter has tendency to preserve the shape of the signal component in the passband region of the filter. The frequency response of zero-phase 2-D FIR filters is a real valued function, i. e.,

$$H(\omega_1, \omega_2) = H^*(\omega_1, \omega_2) . \qquad (2.2.2)$$

If in some frequency regions $H(\omega_1, \omega_2)$ becomes negative, then a phase shift of - π radians occurs. Typically, the frequency response can become negative in regions corresponding to the stopbands, and a phase of π rad. in the stopbands has little significance. Provided that the impulse response $h(n_1, n_2)$ is real, the constraint (2.2.2) is equivalent in the space domain to a symmetric impulse response with respect to the origin of (n_1, n_2) plane:

$$h(n_1, n_2) = h(-n_1, -n_2) \qquad (2.2.3)$$

In this chapter, the design of 2-D rectangularly shaped zero phase FIR filters from nonuniform samples in the frequency plane (ω_1, ω_2) is considered. Frequency response specifications for the four basic types of 2-D filters with rectangular shape are presented in Fig. 2.1. Since the filters are zero phase, only the magnitude specifications are given, using a tolerance scheme, Eq.(2.2.4). The darkly shaded regions are the passband regions, denoted as R_p, and the unshaded regions, R_s, correspond to the stopbands.

$$1 - \delta_p \le |H(\omega_1, \omega_2)| \le 1 + \delta_p, \qquad (\omega_1, \omega_2) \in R_p$$
$$|H(\omega_1, \omega_2)| \le \delta_s, \qquad (\omega_1, \omega_2) \in R_s \tag{2.2.4}$$

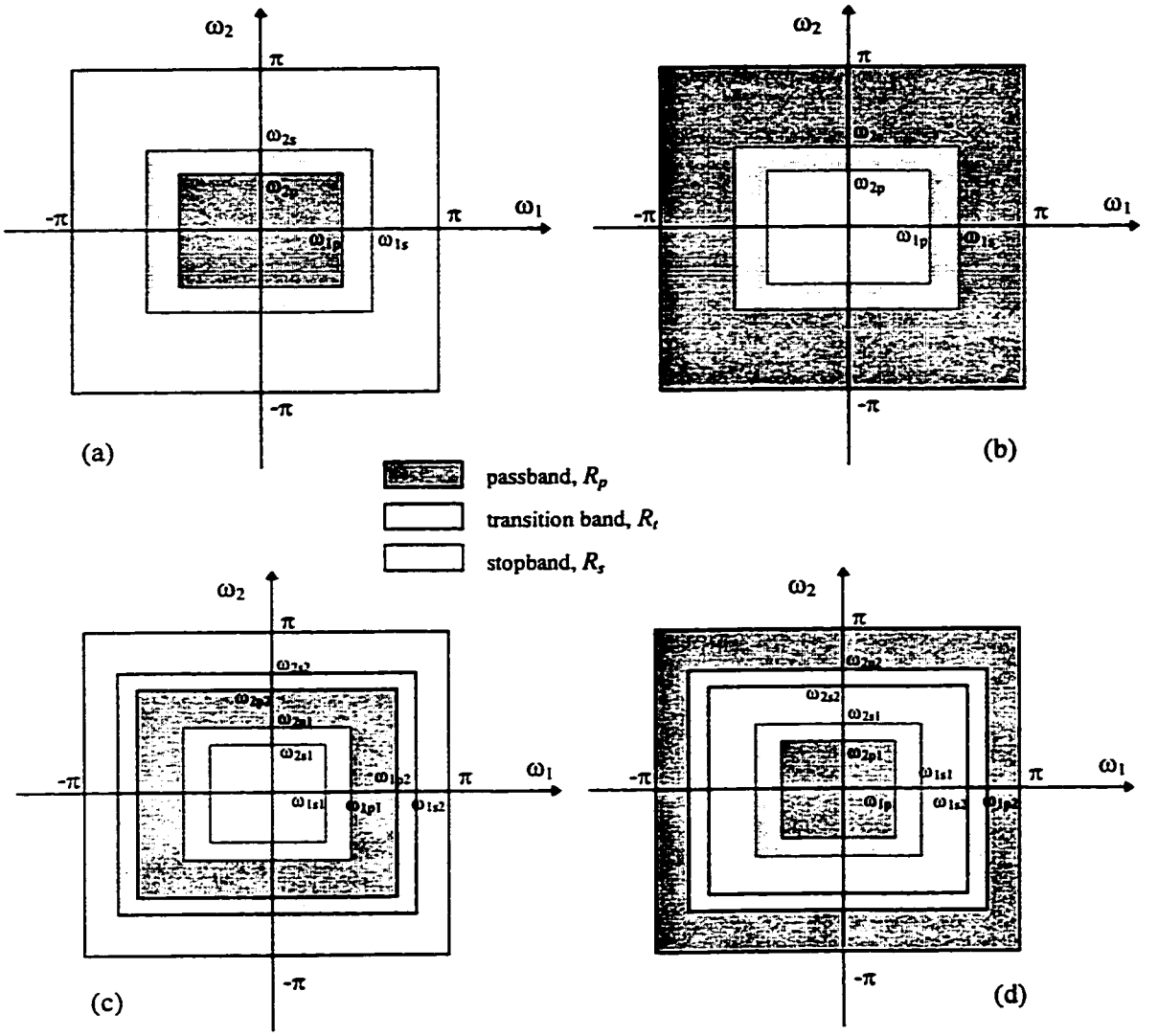


Figure 2.1 Frequency response specifications of rectangularly shaped ideal FIR filters. (a) Lowpass filter; (b) highpass filter; (c) bandpass filter; (d) bandstop filter.

From the figure above it is clear that the rectangularly shaped filters have fourfold (quadrant) symmetric frequency responses.

$$H(\omega_1, \omega_2) = H(-\omega_1, \omega_2) = H(\omega_1, -\omega_2) \tag{2.2.5}$$

In the space domain this constraint is equivalent to a fourfold symmetry of the impulse response $h(n_1, n_2)$ given by

$$h(n_1, n_2) = h(-n_1, n_2) = h(n_1, -n_2) \tag{2.2.6}$$

The symmetry constraints reduce the number of the independent parameters to be estimated during the design. Also, the number of arithmetic operations in the implementation is reduced. This reduction is more than 3 times compared to an arbitrary FIR filter and is greater for larger filter sizes. For instance, this factor is 3.52 for a filter of size 15×15, and 3.75 for a 31×31-point filter. The independent points are usually chosen in the first quadrant of the (n_1, n_2) plane. This is illustrated in Fig. 2.2 for a 7×7-point filter.

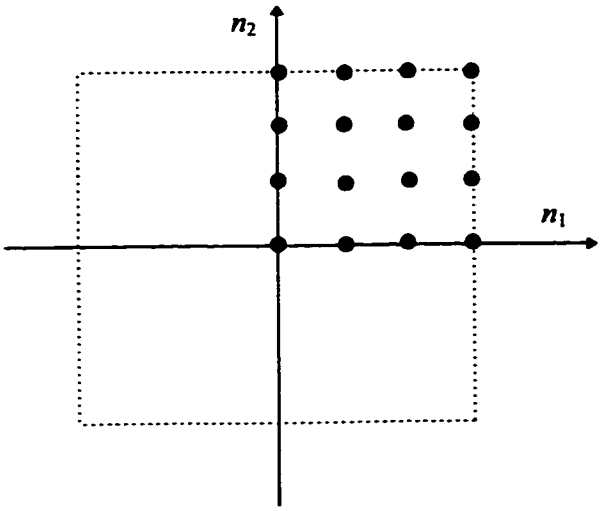


Figure 2.2 Independent points of an FIR filter with a fourfold symmetry of size 7×7.

Let's consider a 2-D FIR rectangularly shaped filter with $N_1 \times N_2$-point impulse response region of support. If N_1 and N_2 both are odd integers and $h(n_1, n_2)$ is centered at the origin of (n_1, n_2) plane, then the frequency response can be expressed as

$$H(\omega_1, \omega_2) = \sum_{n_1=-M_1}^{M_1} \sum_{n_2=-M_2}^{M_2} h(n_1, n_2) \exp(-j\omega_1 n_1) \exp(-j\omega_2 n_2) \tag{2.2.7}$$

where $M_1 = \dfrac{N_1 - 1}{2}$, $M_2 = \dfrac{N_2 - 1}{2}$.

To ensure a quadrant symmetric, zero-phase frequency response, the impulse response $h(n_1, n_2)$ must have the following fourfold symmetry:

$$h(n_1,n_2) = h(-n_1,n_2) = h(n_1,-n_2) \tag{2.2.8}$$

Applying these symmetry conditions to Eq. (2.2.7), the following expression for $H(\omega_1, \omega_2)$ is obtained

$$\begin{aligned} H(\omega_1,\omega_2) &= h(0,0)+\sum_{n_1=1}^{M_1} 2h(n_1,0)\cos(\omega_1 n_1)+\sum_{n_2=1}^{M_2} 2h(0,n_2)\cos(\omega_2 n_2)+ \\ &+\sum_{n_1=1}^{M_1}\sum_{n_2=1}^{M_2} 4h(n_1,n_2)\cos(\omega_1 n_1)\cos(\omega_2 n_2) \\ &=\sum_{n_1=0}^{M_1}\sum_{n_2=0}^{M_2} a(n_1,n_2)\cos(\omega_1 n_1)\cos(\omega_2 n_2) \end{aligned} \tag{2.2.9}$$

where $a(0, 0) = h(0, 0)$,

$a(n_1, 0) = 2h(n_1, 0)$,

$a(0, n_2) = 2h(0, n_2)$, and

$a(n_1, n_2) = 4h(n_1, n_2)$ for $n_1, n_2 \neq 0$. (2.2.10)

Therefore, the number of independent coefficients is

$$L = 1+\frac{N_1-1}{2}+\frac{N_2-1}{2}+\frac{(N_1-1)(N_2-1)}{4} = \frac{(N_1+1)(N_2+1)}{4} = (M_1+1)(M_2+1) \tag{2.2.11.a}$$

In the case where $N_1 = N_2 = N = 2M + 1$ this number is

$$L = \frac{(N+1)^2}{4} = (M+1)^2 \tag{2.2.11.b}$$

Hence, L frequency samples $H(\omega_{1k}, \omega_{2k})$ will be sufficient to solve for the $N_1 \times N_2$ filter coefficients $h(n_1, n_2)$. These samples are taken in the first quadrant of (ω_1, ω_2) plane: the desired frequency response characteristic is to be sampled at L points in the region $\{ (\omega_1, \omega_2): 0 \leq \omega_1 \leq \pi; 0 \leq \omega_2 \leq \pi \}$. Let's denote the discrete set of interpolation points

$K=\left\{\left(\omega_{1k},\omega_{2k}\right)\right\}_{k=1}^{L}$. The design problem becomes a problem of solving the following system of linear equations for the coefficients $a(n_1, n_2)$, $0 \le n_1 \le M_1$, $0 \le n_2 \le M_2$:

$$\begin{bmatrix} 1 & \cos\omega_{11} & \cos\omega_{11}\cos\omega_{21} & \cdots & \cos M_1\omega_{11}\cos M_2\omega_{21} \\ 1 & \cos\omega_{12} & \cos\omega_{12}\cos\omega_{22} & \cdots & \cos M_1\omega_{12}\cos M_2\omega_{22} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \cos\omega_{1L} & \cos\omega_{1L}\cos\omega_{2L} & \cdots & \cos M_1\omega_{1L}\cos M_2\omega_{2L} \end{bmatrix} \begin{bmatrix} a(0,0) \\ a(1,0) \\ \vdots \\ a(M_1, M_2) \end{bmatrix} = \begin{bmatrix} H(\omega_{11},\omega_{21}) \\ H(\omega_{12},\omega_{22}) \\ \\ H(\omega_{1L}\omega_{2L}) \end{bmatrix}$$

(2.2.12)

If the coefficients $a(n_1, n_2)$ are to be determined via a linear least-squares fit, then the number of the frequency samples will be denoted N_s , and usually $N_s > L$.

2.3 Arbitrary sampling.

Arbitrary sampling means allowing frequency samples to occur anywhere in the (ω_1, ω_2) plane. This approach involves the solution of L linear equations, where L is the number of independent filter coefficients, Eq. (2.2.11), and is computationally intensive for filters of high order.

The arbitrary sampling technique provides great flexibility in choosing the sampling locations but it suffers from many theoretical and practical problems. As it has been pointed out in Sec. 1.2, in 1-D case a set of $N+1$ arbitrary samples guarantees that an N order interpolating polynomial can be placed through them. In 2-D, however, there are no sets of $N+1$ universal functions which can be used for interpolation at any $N+1$ distinct points. In other words the polynomials in 2-D do not form a Chebyshev system, and therefore, bivariate polynomials are not specified in general from samples at arbitrary locations. Solutions to the interpolation equations may not exist or be unique. This is illustrated with the following examples. Since the filters are fourfold symmetric, the number of the independent coefficients, and, therefore, the number of the necessary frequency samples is L, given by (2.2.11).

Example 2.1: A square-shape lowpass FIR filter with impulse response region of support size 7×7 points designed by arbitrary sampling. Filter specifications:

$$|H_d(\omega_1, \omega_2)| = 1 \quad \text{for} \quad |\omega_1| \le 0.2\pi \text{ and } |\omega_2| \le 0.2\pi$$
$$|H_d(\omega_1, \omega_2)| = 0 \quad \text{for} \quad 0.4\pi \le |\omega_1| \le \pi \text{ or } 0.4\pi \le |\omega_2| \le \pi$$

The samples are taken arbitrarily in the first quadrant of the (ω_1, ω_2) plane. Since a 7×7-point square filter is to be designed, only 16 coefficients are independent out of 49. Therefore, 16 frequency samples will be sufficient to solve the problem. These samples are taken at locations shown in Fig. 2.3 (a). The magnitude response perspective and contour plots are shown in Fig 2.3 (b), and (c), respectively. Isocontours at levels 1, 0.8, 0.4, 0.2, and 0.02 are shown.

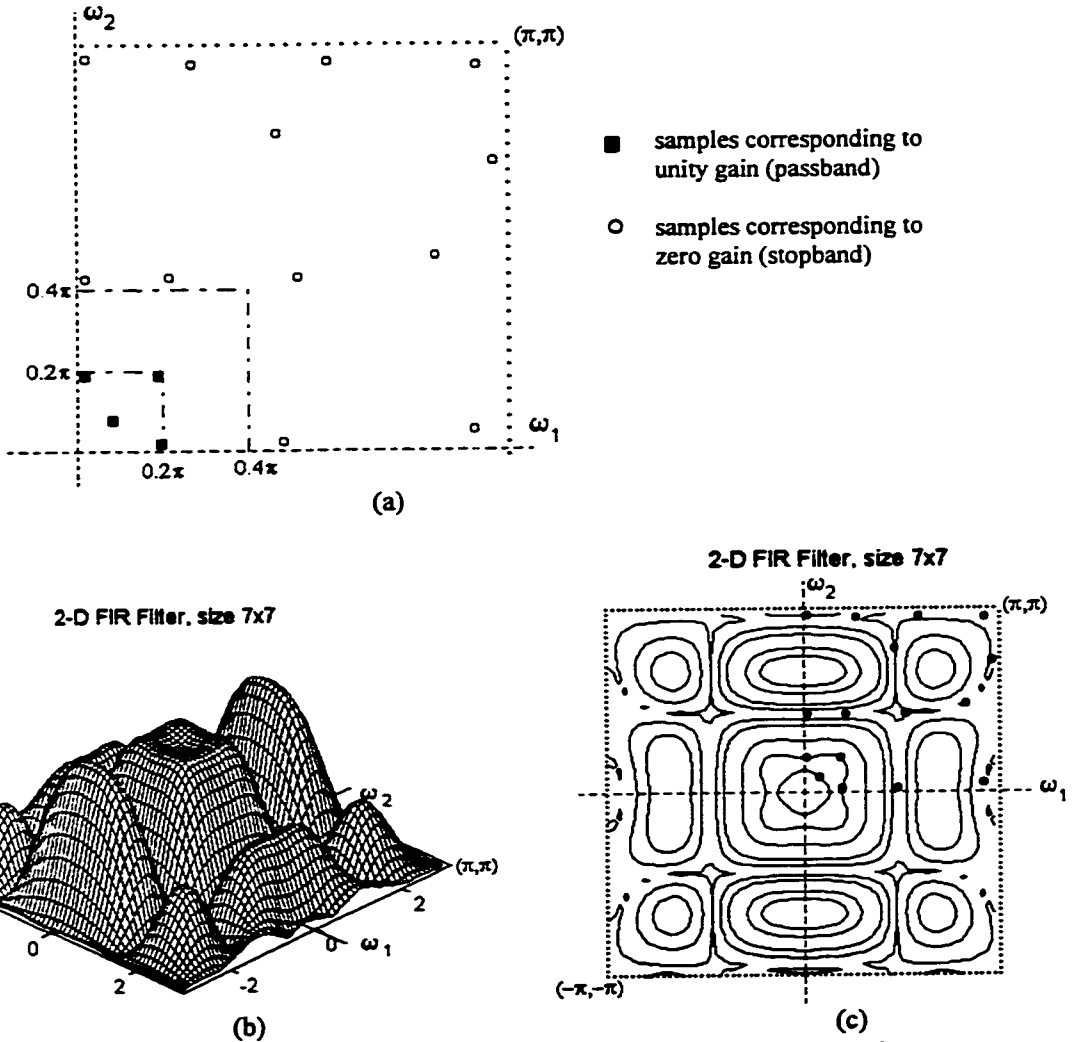


Figure 2.3 Example 2.1: A 7×7-point FIR filter designed using 16 arbitrary taken samples. (a) Sample locations; (b) Perspective plot of the frequency response magnitude; (c) Contour plot and sample locations (denoted by the dots).

Clearly, the error in the stopband is extremely large and this design is unacceptable. In the next example this error is decreased by increasing the size of the filter.

Example 2.2: The desired filter frequency response is specified using the same tolerance scheme as in Example 2.1. Now the filter is of size 9×9. The number of independent samples is L =25. Now the samples are taken more uniformly both in the passband and stopband region, i. e., their density is approximately constant in these regions. This improves the characteristic in the stopband but the error is still large: δ_p = 0.6301, δ_s = 0.1704. The sample locations and the resulting frequency response are shown in Fig. 2.4 (a), (b), and (c).

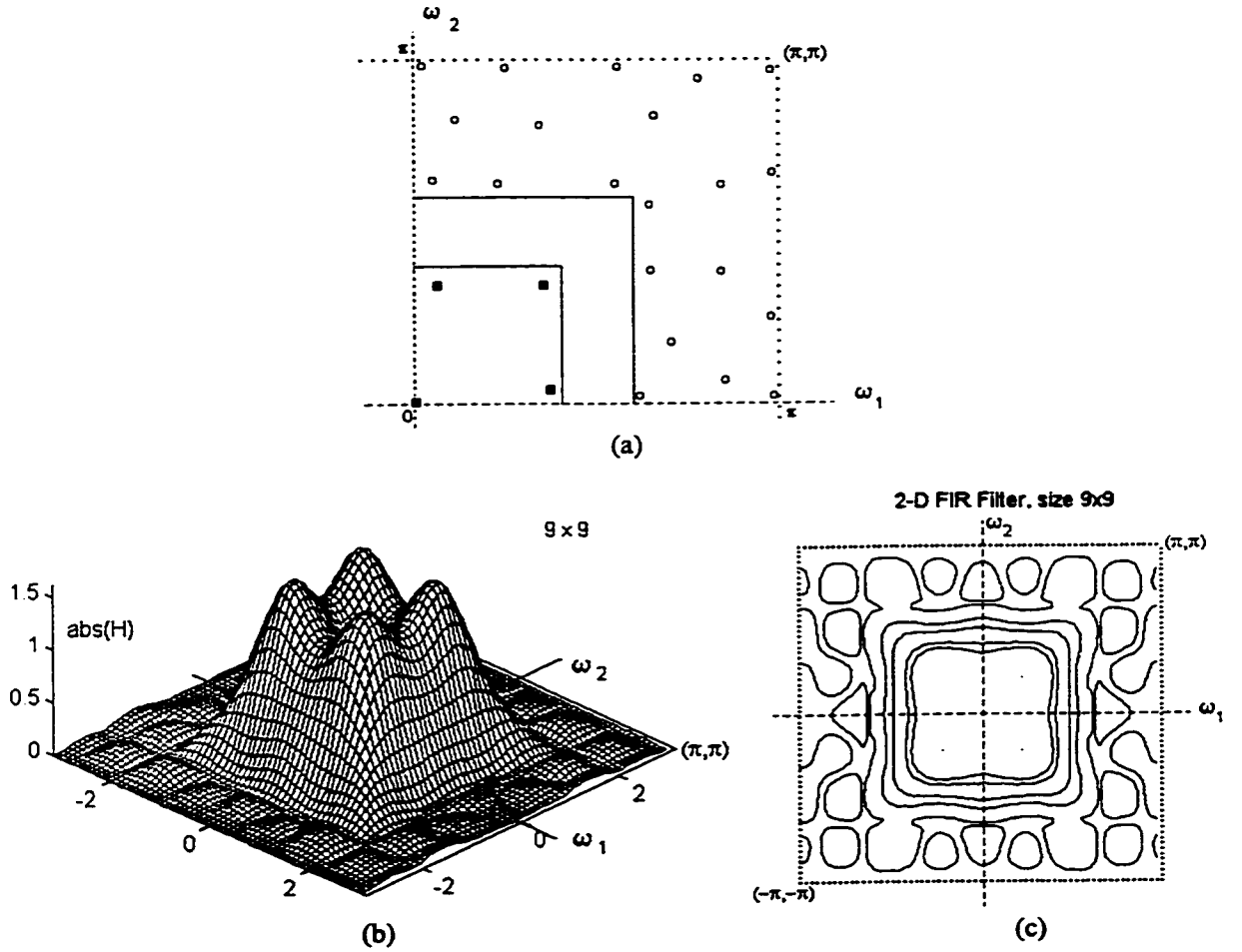
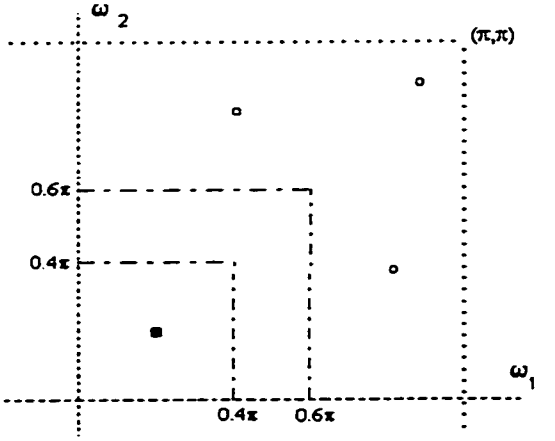


Figure 2.4 Example 2.2: A 9×9-point FIR filter designed using 25 arbitrary taken samples. (a) Sample locations; (b) Perspective plot of the frequency response magnitude; (c) Contour plot at levels 1, 0.8, 0.4, 0.2, and 0.02 and sample locations (the dots).

Example 2.3: This example shows the great potential of the nonuniform frequency sampling. The same tolerance scheme as in Example 2.1 is used. Now the filter is of size 3×3 and only $L = 4$ samples are necessary. The sample locations and the resulting frequency response are shown in Fig. 2.5. The deviations are $\delta_p = 0.5075$ and $\delta_s = 0.3510$ in the pass- and stopband, respectively. The error is not relatively so large if compared to the "standard" uniform sampling approach or to other 2-D FIR filters of the same size. For example, the following 3×3 filter [10]

$$h = \begin{bmatrix} & 1/6 & \\ 1/6 & 1/3 & 1/6 \\ & 1/6 & \end{bmatrix}$$

has deviations $\delta_p = 0.4731$ and $\delta_s = 0.5544$ (assumed that $\omega_{1p} = \omega_{2p} = 0.4\pi$ and $\omega_{1s} = \omega_{2s} = 0.6\pi$). The filter designed with the uniform sampling approach shows deviations $\delta_p = 0.7224$, $\delta_s = 0.3325$. The results are summarized in Table 2.1.



(a)

Figure 2.5 Example 2.3: (a) Sample locations for a 3×3-point FIR filter. The resulting perspective and contour plots are shown in Fig. 2.5, (b) and (c) on the next page.

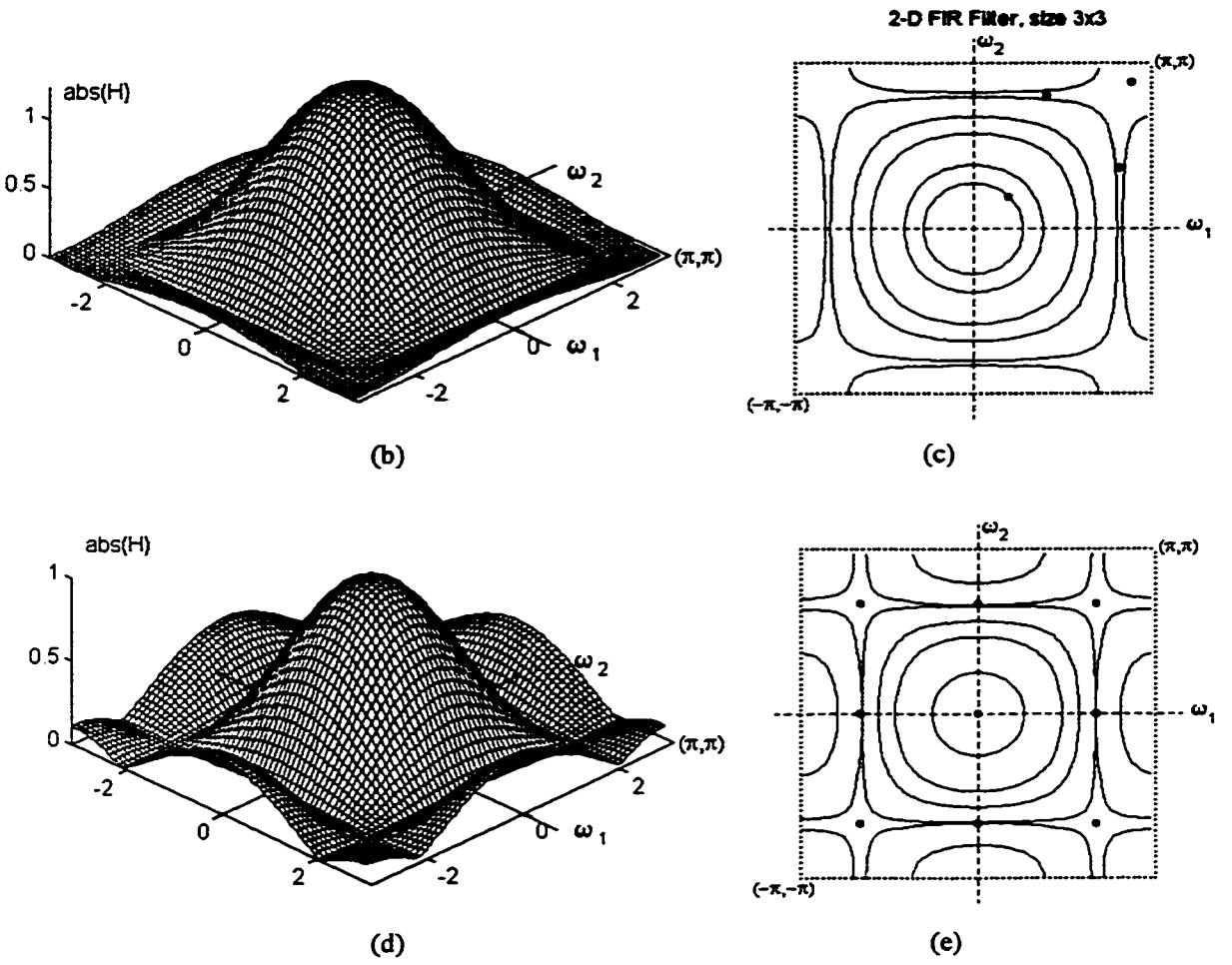


Figure 2.5 (continued) (b) and (c): Example 2.3: A 3×3-point FIR filter designed using 4 nonuniformly taken samples shown in Fig. 2.5 (a)..
(d) and (e): A 3×3-point FIR filter designed using 9 uniformly taken samples.

Besides the theoretical disadvantages as possibly degenerate or ill-conditioned matrices, it became clear from the experiments that the arbitrary sampling approach suffers from practical difficulties too. The prescribed deviations in the passband(s) and stopband(s) cannot be guaranteed, even worse, the design error can be inadmissibly large, as in Example 2.1. Therefore, some restrictions to the frequency sample locations should be imposed in order to avoid the above mentioned disadvantages.

2.4 Rectangular Shape FIR filter design by nonuniform sampling on parallel lines.

This is the special case (A) of the NDFT, considered in Sec. 1.1. It has been shown there (see Eqs. (1.22)-(1.26)) that the 2-D NDFT matrix is nonsingular provided the vertical lines are distinct and the samples on each vertical line are distinct. Instead of solving one large system of linear equations, this approach solves several smaller 1-D systems. Each line determines one 1-D FIR filter or a slice of the 2-D filter. For a FIR filter with impulse response of size $N_1 \times N_2$ points, where $N_1 = 2M_1+1$ and $N_2 = 2M_2+1$, there are M_1+1 vertical lines with coordinates ω_{1k} . On each line are taken M_2+1 samples (cf. Fig. 1.2). For simplicity and avoiding repetition, the frequency samples are taken only in the first quadrant of the (ω_1,ω_2) plane, $0 \le \omega_1, \omega_2 \le \pi$. The expression (2.2.9) for a zero phase fourfold symmetric 2-D FIR filter evaluated at the sampling locations $(\omega_{1k}, \omega_{2km})$ can be expressed as

$$\begin{aligned} H(\omega_{1k},\omega_{2km}) &= \sum_{n_1=-M_1}^{M_1} \sum_{n_2=-M_2}^{M_2} h(n_1,n_2)\, e^{-j\omega_{1k}n_1} e^{-j\omega_{2km}n_2} \\ &= \sum_{n_1=0}^{M_1} \sum_{n_2=0}^{M_2} a(n_1,n_2)\cos(n_1\omega_{1k})\cos(n_2\omega_{2km}) \\ &= \sum_{n_2=0}^{M_2} g(\omega_{1k},n_2)\cos(n_2\omega_{2km}) \end{aligned} \tag{2.4.1}$$

for $k = 0, 1, \ldots, M_1$, $m = 0, 1, \ldots, M_2$,

where

$$g(\omega_{1k},n_2) = \sum_{n_1=0}^{M_1} a(n_1,n_2)\cos(n_1\omega_{1k}) , \qquad n_2 = 0, 1, \ldots, M_2. \tag{2.4.2}$$

For constant k (for each k-th vertical line), we have a 1-D interpolation problem with

M_2+1 points or a system of M_2+1 linear equations. For fixed value of k, Eq. (2.4.1) can be written in matrix notation as

$$\mathbf{H}_k = \mathbf{V}_{2k}\mathbf{g}_k \tag{2.4.3}$$

where $\mathbf{H}_k$ is an $(M_2+1) \times 1$ column vector containing the sample values on the k-th line. The vector $\mathbf{g}_k$ is $(M_2+1) \times 1$ column vector and $\mathbf{V}_{2k}$ is the k-th $(M_2+1)\times(M_2+1)$ matrix of cosine terms

$$\mathbf{V}_{2k} = \begin{bmatrix} 1 & \cos\omega_{2k0} & \cdots & \cos M_2\omega_{2k0} \\ 1 & \cos\omega_{2k1} & \cdots & \cos M_2\omega_{2k1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \cos\omega_{2kM_2} & \cdots & \cos M_2\omega_{2kM_2} \end{bmatrix}, \qquad k = 0, 1, \ldots, M_1 \tag{2.4.4}$$

Therefore, the vectors $\mathbf{g}_k$ are directly computed from

$$\mathbf{g}_k = \mathbf{V}_{2k}^{-1}\mathbf{H}_k \tag{2.4.5}$$

and this is repeated for each value of k, $k = 1, \ldots, N_1$. The sequence $h(n_1, n_2)$ is calculated from (2.4.2) which in matrix form is written as

$$\mathbf{q}_{n2} = \mathbf{V}_1\mathbf{a}_{n2} \tag{2.4.6}$$

where $\mathbf{q}_{n2}$ is the $(M_1+1)\times 1$ n_2-th row of $g(\omega_{1k}, n_2)$, $\mathbf{a}_{n2}$ is the n_2-th row of $a(n_1, n_2)$, and $\mathbf{V}_1$ is a $(M_1+1) \times (M_1+1)$ matrix

$$\mathbf{V}_1 = \begin{bmatrix} 1 & \cos\omega_{10} & \cdots & \cos M_1\omega_{10} \\ 1 & \cos\omega_{11} & \cdots & \cos M_1\omega_{11} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \cos\omega_{1M_1} & \cdots & \cos M_1\omega_{1M_1} \end{bmatrix} \tag{2.4.7}$$

Hence,

$$\mathbf{a}_{n2} = \mathbf{V}_1^{-1}\,\mathbf{q}_{n2} \tag{2.4.8}$$

and this is repeated for each value of n_2 , $n_2 = 0, 1, \ldots, M_2$, in order to obtain the whole sequence $a(n_1 , n_2)$. Then, the impulse response $h(n_1 , n_2)$ is obtained from the sequence $a(n_1 , n_2)$ using the relations (2.2.10). The matrices $\mathbf{V}_1$ and $\mathbf{V}_{2k}$ are nonsingular provided the parallel lines and the samples are selected at non-repeated locations. The nonsingularity is guaranteed by the fact that the set of functions $\{\cos k\omega\}_{k=0}^{M}$ satisfy the Haar condition [3, 4], i.e., they form a Chebyshev set, see Sec. 1.2.

The basis for the algorithm is presented by Rozwod, Therrien and Lim in [6] and further developed by Angelidis in [7]. However, the coordinates of the parallel lines and the sample locations on each line is an issue that needs further investigation. An arbitrary choice of line coordinates and samples does not produce the desired results in terms of deviations and band edges. It may even produce very unwanted results. This is illustrated by the following example.

Example 2.4: Filter specifications:

$$|H_d(\omega_1, \omega_2)| = 1 \quad \text{for} \quad |\omega_1| \le 0.35\pi \ \text{and} \ |\omega_2| \le 0.35\pi$$

$$|H_d(\omega_1, \omega_2)| = 0 \quad \text{for} \quad 0.65\pi \le |\omega_1| \le \pi \ \text{or} \ 0.65\pi \le |\omega_2| \le \pi$$

A. Filter size: 9×9 points.

The sample locations are shown in Fig. 2.6 (a). The samples are taken arbitrary on 5 vertical lines, 5 samples on each line. No points are taken in the transition region. The magnitude of the filter is shown in Fig. 2.6 (b).

B. Filter size: 11×11 points.

A total of 36 samples are taken on 6 lines, 6 samples on each, Fig. 2.6 (c). The resulting magnitude is shown on Fig. 2.6 (d).

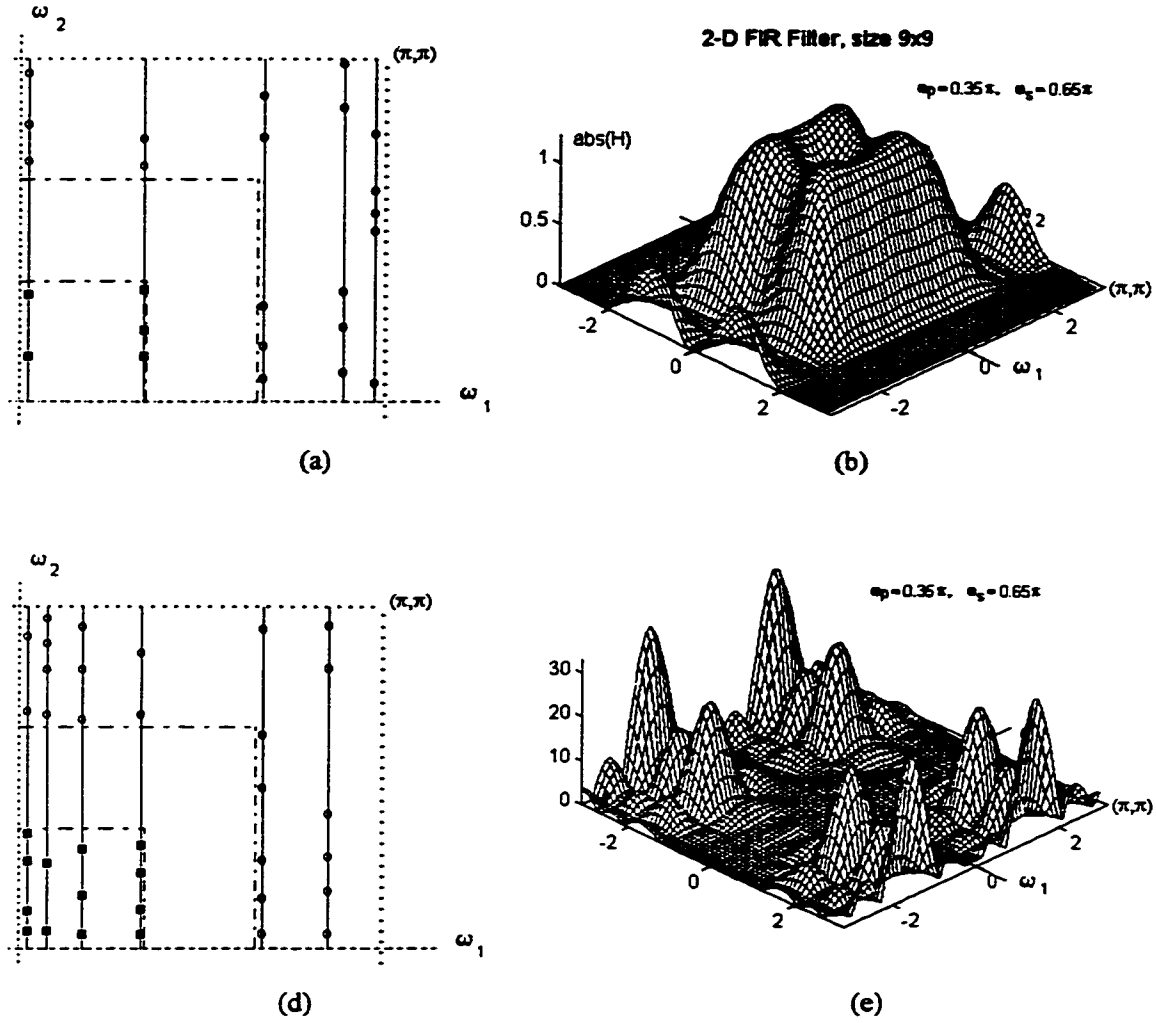


Figure 2. 6 Sample point locations and perspective plots of the frequency response for Example 2.4.
(a) and (b) case A, 9×9-point impulse response;
(d) and (e) case B, 11×11-point impulse response.

The above example clearly shows that for good design results some appropriate locations of the vertical (or horizontal) lines and the samples on each line should be found. Here, the method will be developed a little bit further. It will be shown that it is possible to find such locations of the line and sample coordinates that the resulting filters will be with good performance.

In the proposed in this section sampling method, the coordinates of the vertical lines ω_{1k}, $k = 0, 1, \ldots, M_1$, are calculated using exponential functions. For the passband the

exponential function used is

$$f_p(x)=1-e^{-\alpha x}, \qquad x \in [0, 1], \tag{2.4.9}$$

and for the stopband the function used is

$$f_s(x)=e^{\alpha x}-1, \qquad x \in [0, 1], \tag{2.4.10}$$

In both cases α is a positive constant. The functions f_s and f_p are sampled uniformly in the interval [0, 1] and then the samples are linearly mapped to the ω_1 axis, Fig. 2.7.

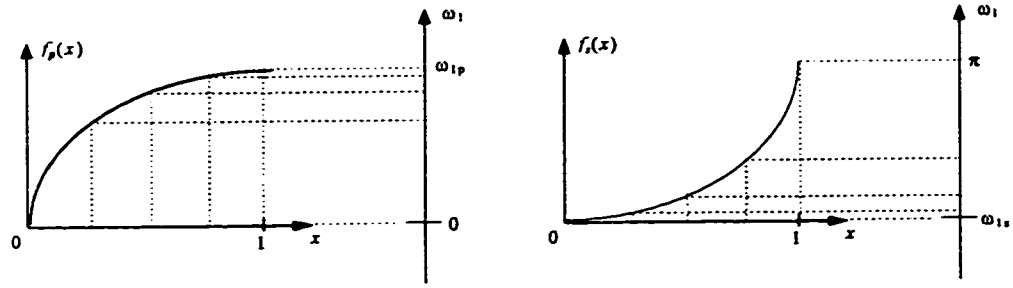


Figure 2.7 Nonlinear mapping of the uniform nodes in the interval [0, 1] to the passband (a) and the stopband (b) along the ω_1 axis.

The function $f_p(x)$ is sampled uniformly at P_1 points in the interval [0, 1]. The sample values are mapped to the interval [0, ω_{1p}]. ω_{1p} is the passband edge in ω_1 direction. The final effect is a nonlinear (exponential) mapping of the uniform nodes.

$$\omega_{1i} = \frac{\omega_{1p}}{f_{pm}} f_p(x_i), \tag{2.4.11}$$

where $x_i = i/(P_1-1)$ for $i=0, 1, ..., P_1-1$, and $f_{pm}=f_p(1)$

Similarly, the function $f_s(x)$ is sampled uniformly at S_1 points and the sample locations in

the stopband along ω_1 axis are obtained according to the following transformation

$$\omega_{1(i+P_1)} = \frac{\pi - \omega_{1s}}{f_{sm}} f_s(x_i) + \omega_{1s}, \tag{2.4.12}$$

where $x_i = i/(S_1 - 1)$ for $i = 0, 1, ..., S_1 - 1$, and $f_{sm} = f_s(1)$.

The numbers P_1 and S_1 are chosen in proportion to the passband and stopband width, respectively, and such that $P_1 + S_1 = M_1 = (N_1-1)/2$. The filter impulse response size $N_1 \times N_2$ can be roughly estimated in terms of the desired deviations δ_p , δ_s , and transition band width by [11]

$$N_i = \frac{-20\log_{10}\sqrt{\delta_p \delta_s} - 8}{2.10(\omega_{is} - \omega_{ip})}, \qquad i = 1, 2 \tag{2.4.13}$$

The sample point locations along each of the vertical lines are obtained using the same exponential mappings (2.4.11) and (2.4.12), this time in ω_2 direction. The number of the samples on each line in the passband region is P_2 and in the stopband region this number is S_2, such that $P_2 + S_2 = M_2 = (N_2-1)/2$. This mapping is used only for $0 \le \omega_1 \le \omega_{1p}$. For the stopband region $\omega_{1s} \le \omega_1 \le \pi$, M_2 samples are taken uniformly on each line for $0 \le \omega_2 \le \pi$. Eqs. (2.4.11) and (2.4.12) now become

$$\omega_{2ij} = \frac{\omega_{2p}}{f_{pm}} f_p(x_j), \qquad \text{for } i = 0, 1, ..., P_1 - 1 \tag{2.4.14}$$

where $x_j = j/(P_2-1)$ for $j = 0, 1, ..., P_2-1$, and $f_{pm} = f_p(1)$.

$$\omega_{2i(j+P_2)} = \frac{\pi - \omega_{2s}}{f_{sm}} f_s(x_j) + \omega_{2s}, \qquad \text{for } i = 0, 1, ..., P_1 - 1 \tag{2.4.15}$$

where $x_j = j/(S_2-1)$ for $j = 0, 1, ..., S_2-1$, and $f_{sm} = f_s(1)$.

For the region $\omega_{1s} \le \omega_1 \le \pi$,

$$\omega_{2ij} = j\frac{\pi}{2M_2}, \quad \text{for} \quad j = 1, 3, 5, ..., 2M_2-1, \quad i = P_1, \ldots, M_1+1 \tag{2.4.16}$$

The value of the frequency samples in the passband region is set to 1, and in the stopband region it is 0.

The following example demonstrates the good quality of rectangular-shape FIR filters designed with frequency samples determined in the above-described fashion.

Example 2.5 Square shape 2-D FIR filter with specifications

$|H_d(\omega_1, \omega_2)| = 1$ for $|\omega_1| \le 0.35\pi$ and $|\omega_2| \le 0.35\pi$
$|H_d(\omega_1, \omega_2)| = 0$ for $0.65\pi \le |\omega_1| \le \pi$ or $0.65\pi \le |\omega_2| \le \pi$

Impulse response size: (A) 11×11; (B) 21×21, (C) 81×81

The parameter α = 1.25. The sample locations and the resulting frequency responses are shown in Fig. 2.8, Fig. 2.9, and Fig. 2.10 for case A, B, and C, respectively

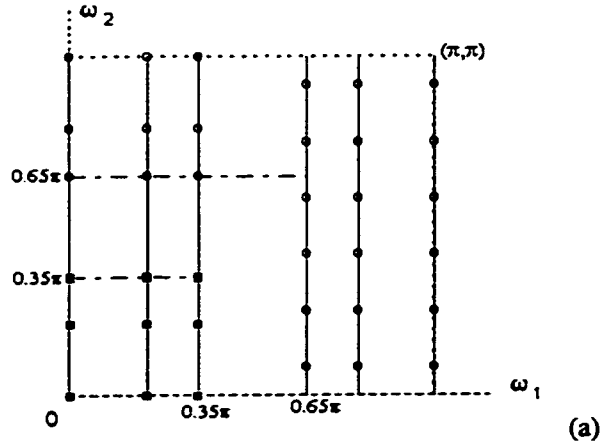
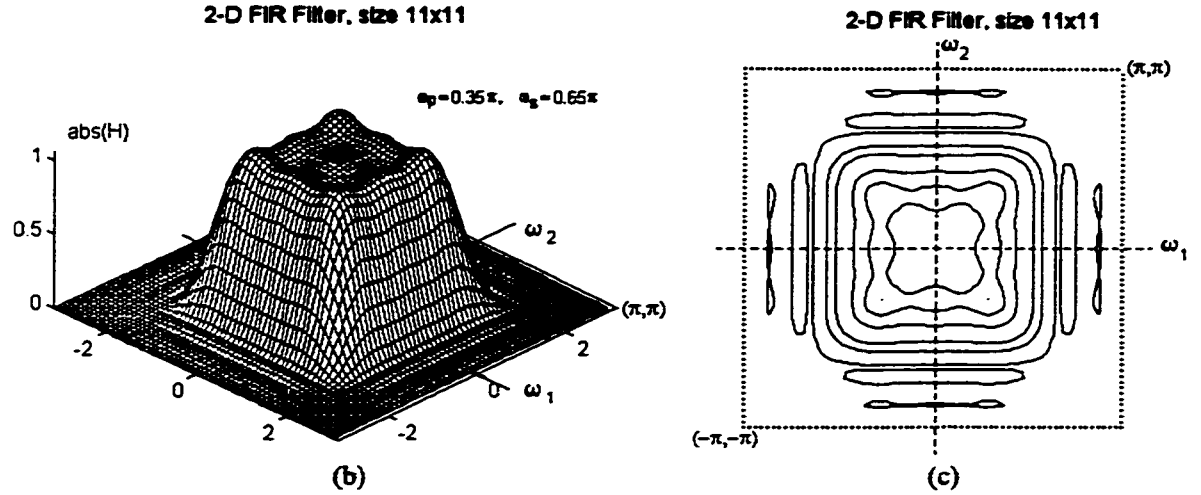
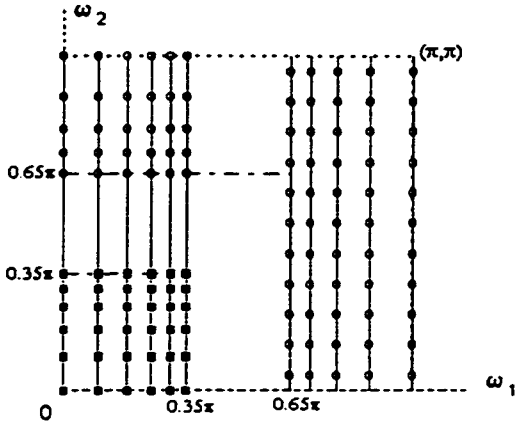


Figure 2.8 Example 2.5 (A): An 11×11-point zero phase square shape FIR filter designed using the proposed method for the frequency sample locations. (a) Sample locations.

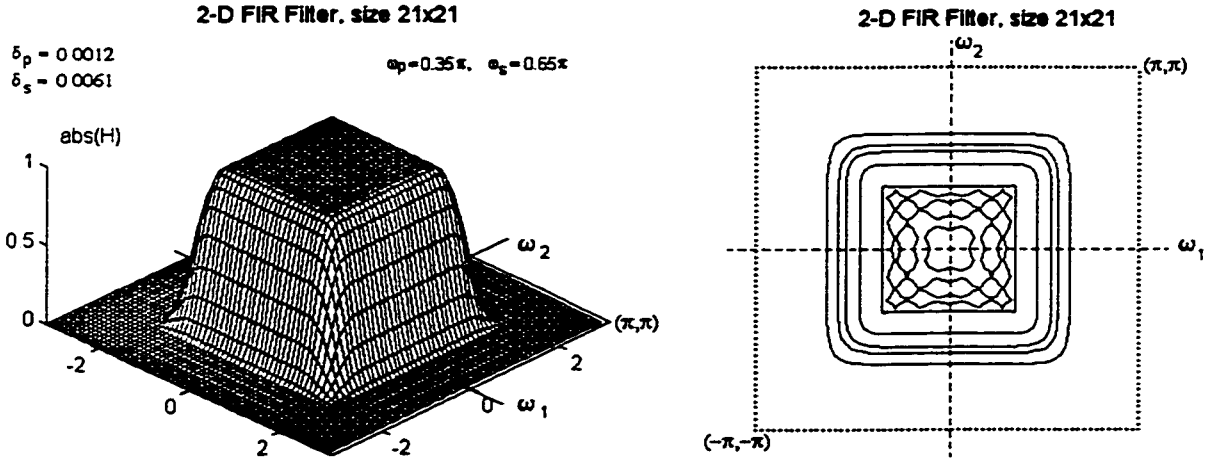


(b) (c)

Figure 2.8 (continued) Example 2.5(A): (b) Frequency response perspective plot; (c) frequency response contour plot. The maximum error is $\delta_p = 0.0682$ in the passband and $\delta_s = 0.0346$ in the stopband.

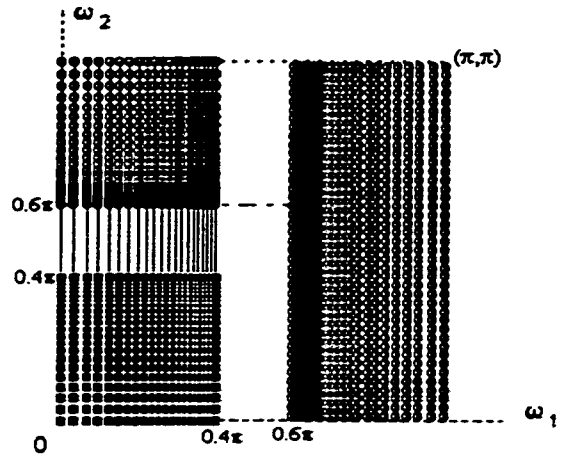


(a)

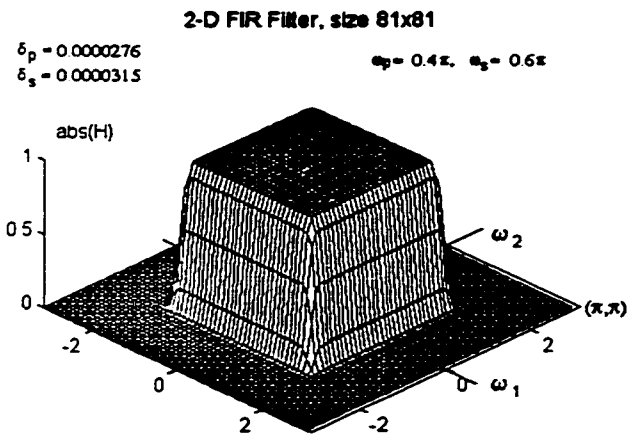


(b) (c)

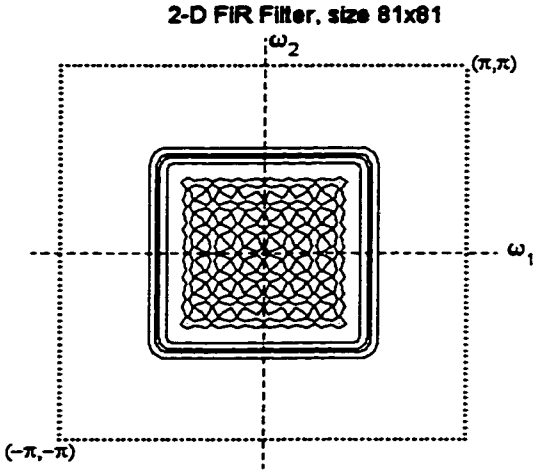
Figure 2.9 Example 2.5 (B): a 21×21-point zero phase square shape FIR filter designed using the proposed method for the frequency sample locations. (a) Sample locations; (b) Frequency response perspective plot; (c) frequency response contour plot. The approximation error is $\delta_p = 0.0012$ in the passband and $\delta_s = 0.0061$ in the stopband.



(a)



(b)



(c)

Figure 2.10 Example 2.5 (C): an 81×81-point zero phase square shape FIR filter designed using the proposed method for the frequency sample locations. (a) Sample locations; (b) Frequency response perspective plot; (c) frequency response contour plot. The deviations are $\delta_p = 2.76\times10^{-5}$ in the passband and $\delta_s = 3.15\times10^{-5}$ in the stopband.

In the first case, $N_1 = N_2 = 11$, $M_1 = M_2 = 6$, $P_1 = P_2 = 3$, and $S_1 = S_2 = 3$. The samples are taken on six vertical lines. The coordinates of the first three lines are calculated according (2.4.9). The next three lines have coordinates given by (2.4.10). For $0 \le \omega_1 \le 0.35\pi$ the samples are taken on each line according to (2.4.12) and (2.4.13). For $0.65\pi \le \omega_1 \le \pi$ the locations are calculated using (2.4.14). A total of 36 samples is taken in the

first quadrant of the (ω_1, ω_2) plane, shown in Fig. 2.8 (a). The third example shows that the method produces accurate results even for the design of high-order filters which maintain very good shape and low peak approximation error. A total of 1681 samples are taken on 41 vertical lines. In these examples the parameter α in Eqs. (2.4.9) and (2.4.10) was set to 1.25, a value that gives good results. Determining the optimal value of α is a possible direction for further research.

The results are compared with the uniform sapling approach and summarized in Table 2.1

Filter Size	Passb. edges ω_{1p}, ω_{2p}	Stopb. edges ω_{1s}, ω_{2s}	Number of samples	Passband dev., δ_p	Stopband dev., δ_s	Algorithm (Sampling)
3×3	0.4π	0.6π	9	0.7224	0.3325	Uniform
			4	0.5075	0.3510	Arbitrary
9×9	0.35π	0.65π	81	0.1849	0.2235	Uniform
			25	0.0314	0.2464	PLE
11×11	0.35π	0.65π	121	0.2027	0.1496	Uniform
			36	0.0682	0.0346	PLE
21×21	0.35π	0.65π	441	0.1034	0.1115	Uniform
			121	0.0012	0.0061	PLE
81×81	0.35π	0.65π	6561	0.0636	0.0422	Uniform
			1681	0.0000276	0.0000315	PLE

Table 2.1 Performance of the proposed sampling technique on parallel lines..
PLE stands for "sampling on parallel lines using exponential coordinates".

The above described algorithm has been implemented using Matlab. The m file `tsam2.m` is used to generate $(M_1+1)\times(M_2+1)$ frequency samples. Input arguments are the passband and stopband frequency edges (normalized by π) along ω_1 and ω_2 axes, and the numbers M_1+1 and M_1+1. The function returns a vector `w1` containing the sample coordinates along ω_1 axis, a matrix `w12`, each column containing the sample coordinates along ω_2 for each vertical line, and a matrix `Hk` containing the corresponding sample values. The Matlab file `pnint2d.m` calculates and returns the coefficients of the designed 2-D rectangular-shape zero-phase FIR filter. Input arguments are `w1`, `w12`, and `Hk`, returned by `tsam2.m`. The number of arithmetic operations for an $N_1\times N_2$ filter is on the order of $(M_1+1)(M_2+1)^3+(M_2+1)(M_1+1)^3$ where $M_i=(N_i-1)/2$, $i = 1, 2$. The number

of operations needed to calculate the sample locations and values is small compared to the one required for the linear system solution. The program listings are given in Appendix A.

2.5 Samples taken on the vertices of a nonuniformly spaced rectangular grid.

This is the second special case of the NDFT where the inverse NDFT is guaranteed to exist and it is unique. It has been shown in subsection 1.1.4 that in this case the 2-D NDFT matrix is a Kronecker product of two Vandermonde matrices, see Eq. (1.1.37). The corresponding determinant can be expressed as product of terms of the form $(z_{im}^{-1} - z_{in}^{-1})$, $i = 1, 2$, Eq. (1.1.38). Therefore, the nonsingularity is guaranteed by the choice of distinct sampling points. For the purpose of FIR filter design, the sampling points are taken at the vertices of a nonuniform rectangular grid in the (ω_1, ω_2) plane. The ω_1 coordinates of the grid lines running parallel to ω_2 axis can be chosen arbitrary, as long as they are distinct. The same holds for the ω_2 coordinates of the lines parallel to ω_1 axis, cf. Fig. 1.3. In this way the matrices of the systems of linear equations will not be singular. The expression (2.2.9) for the frequency response of an zero phase fourfold symmetric 2-D FIR filter evaluated at the sampling locations $(\omega_{1k}, \omega_{2m})$ now takes the form

$$H(\omega_{1k}, \omega_{2m}) = \sum_{n_1=0}^{M_1} \sum_{n_2=0}^{M_2} a(n_1, n_2)\cos(n_1\omega_{1k})\cos(n_2\omega_{2m}) \tag{2.5.1}$$

for $k = 0, 1, \ldots, M_1$, $m = 0, 1, \ldots, M_2$,

Using matrix notation, this expression can be written as

$$\hat{\mathbf{H}} = \mathbf{V}_1\, \mathbf{a}\, \mathbf{V}_2^T \tag{2.5.2}$$

where $\hat{\mathbf{H}}$ is the $(M_1+1)\times(M_2+1)$ matrix containing the frequency samples, $\mathbf{a}$ is the matrix form of the $(M_1+1)\times(M_2+1)$ sequence $a(n_1, n_2)$ related to the impulse response $h(n_1, n_2)$ by (2.2.10). $\mathbf{V}_1$ and $\mathbf{V}_2$ are matrices of size $(M_1+1)(M_1+1)$ and $(M_2+1)(M_2+1)$, respectively.

$$\mathbf{V}_i = \begin{bmatrix} 1 & \cos\omega_{i0} & \cdots & \cos M_i\omega_{i0} \\ 1 & \cos\omega_{i1} & \cdots & \cos M_i\omega_{i1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \cos\omega_{iM_i} & \cdots & \cos M_i\omega_{iM_i} \end{bmatrix}, \qquad i = 1, 2. \tag{2.5.3}$$

Since a set of 1-D functions $\{\cos k\omega\}_{k=0}^{M}$ forms a Chebyshev system, the matrices $\mathbf{V}_1$ and $\mathbf{V}_2$ are nonsingular and $a(n_1, n_2)$ can be determined directly from

$$\mathbf{a} = \mathbf{V}_1^{-1}\hat{\mathbf{H}}\left(\mathbf{V}_2^T\right)^{-1} \tag{2.5.4}$$

The impulse response of the designed filter $h(n_1, n_2)$ is then determined using (2.2.10). The computational complexity is significantly reduced and it is $O(M_1^3+M_2^3)$. Several techniques for choosing the grid line coordinates will be considered which produce good performance filters.

2.5.1. Grid line coordinates using harmonic series.

In this case the coordinates of the nonuniform rectangular grid are calculated as partial sums of harmonic series.

$$\omega_{ik} = b_{1i}\left[\sum_{m=1}^{k}\frac{1}{m^{\beta}} - 1\right], \qquad k = 1, \ldots, P_i, \qquad i = 1, 2 \tag{2.5.5}$$

for the passband and

$$\omega_{i\,k} = b_{2i} + b_{3i}\left[1 - \sum_{m=1}^{k} \frac{1}{m^{\beta}}\right], \qquad k = 1, \ldots, S_i, \qquad i = 1, 2 \tag{2.5.6}$$

for the stopband, $i = 1, 2$. The total number of samples is $(P_1+S_1)(P_2+S_2) = (N_1+1)(N_2+1)/4$ for an $N_1 \times N_2$ -point filter. The constants b_1, b_2, and b_3 are chosen so as to map the above functions to the passband $[0, \omega_{ip}]$ and stopband $[\omega_{is}, \pi]$ in ω_1 and ω_2 directions, $i = 1, 2$.

$$b_{1i} = \frac{\omega_{ip}}{\sum_{m=1}^{P_i} m^{-\beta} - 1} \tag{2.5.7}$$

$$b_{2i} = \pi \tag{2.5.8}$$

$$b_{3i} = \frac{(\omega_{is} - \pi)}{1 - \sum_{m=1}^{S_i} m^{-\beta}} \tag{2.5.9}$$

for $i = 1, 2$.

The frequencies ω_{1k} and ω_{2k} determine the locations of the vertical and horizontal lines of a nonuniformly spaced rectangular grid. The frequency samples are then taken at the vertices of this grid. Their values are set to 1 in the passband, and to 0 in the stopband. The impulse response is computed according to (2.5.4) and (2.2.10). The Matlab code for obtaining the frequency samples using harmonic series, `sharm3.m`, along with the code for computing the filter impulse response, `grd2d.m`, are presented in Appendix A.

The design results obtained using this nonuniform sampling technique are very similar to the results obtained in Sec. 2.4 using exponential functions and sampling on vertical lines. The two techniques cannot be directly compared because the results were obtained with different values of the parameters α and β and these values were not the optimal. In most of the design examples $\beta = 1.0$. However, for higher order filters, e.g., 21×21-point

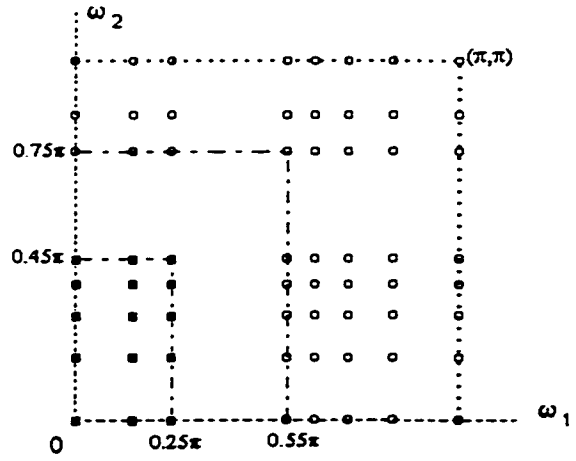
and larger, this value should be decreased to 0.5 --- 0.1. The results are summarized in Table 2.2.

Example 2.6 Rectangular shape 2-D FIR filter designed from frequency samples taken at the vertices of a nonuniform grid obtained with harmonic series.

Filter frequency edge specifications:

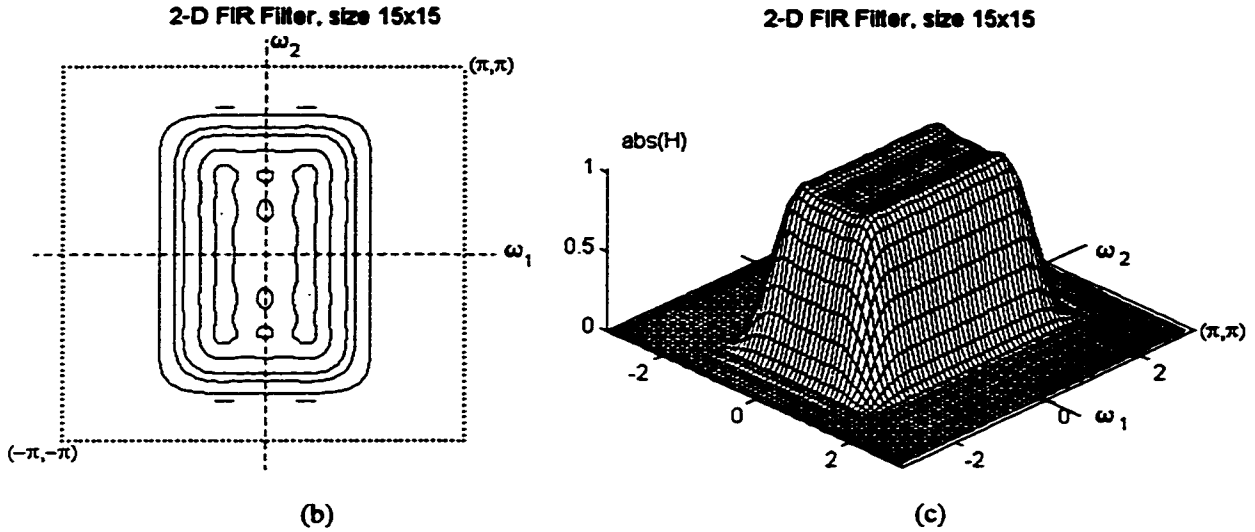
$|H_d(\omega_1, \omega_2)| = 1$ for $|\omega_1| \le 0.25\pi$ and $|\omega_2| \le 0.45\pi$
$|H_d(\omega_1, \omega_2)| = 0$ for $0.55\pi \le |\omega_1| \le \pi$ or $0.75\pi \le |\omega_2| \le \pi$

The numbers P_1=3, S_1=5, P_2=5, S_2=3, resulting in M_1=8 and M_2=8 grid lines in ω_1 direction and ω_2 direction, respectively. The total number of samples is 64, corresponding to a 15×15-point impulse response. The samples taken in the first quadrant of the frequency plane with $\beta = 1.0$ are shown in Fig. 2.11(a). The resulting frequency response is shown in Fig. 2.11 (b) and (c). The maximum deviation in the passband is $\delta_p = 0.0249$ and in the stopband $\delta_s = 0.0205$.



(a)

Figure 2.11 A 15×15-point square shape zero-phase FIR filter designed from frequency samples with locations obtained using harmonic series. (a) Sampling locations;
(continued on the next page).



(b) (c)

Figure 2.11 (continued) A 15×15-point 2-D square shape zero phase FIR filter designed from frequency samples with locations obtained using harmonic series, $\beta = 1.0$.
(b) frequency response contour plot; (c) frequency response perspective plot.

While for low-order filters good results were obtained with $\beta = 1.0$, for filters of higher order filters (impulse response 21×21 points and more) this value gives large error and β should be decreased. Values between 0.5 and 0.1 work well, the smaller values to be used for filters of larger size and narrower transition bands. Evidently, the small values of β decrease the "degree of nonuniformity" and the design results become similar to these obtained with uniform sampling.

2.5.2 Grid line coordinates with exponential distribution.

This nonuniform sampling technique is very similar to that considered in Sec. 2.4. The coordinates of the vertical grid lines ω_{1k}, $k = 0, 1, \ldots, M_1$, are obtained in the same way as it was done in Sec. 2.4, using formulas (2.4.9) - (2.4.11). The following exponential functions are used:

$$f_p(x) = 1 - e^{-\alpha x}, \qquad x \in [0, 1], \tag{2.5.10}$$

for the passband and

$$f_s(x)=e^{\alpha x}-1, \qquad x \in [0, 1], \tag{2.5.11}$$

for the stopband. The coordinates are obtained using

$$\omega_{1i} = \frac{\omega_{1p}}{f_{pm}} f_p(x_i), \tag{2.5.12}$$

where $x_i = i/(P_1\text{-}1)$ for $i = 0, 1, ..., P_1\text{-}1$, and $f_{pm} = f_p(1)$

Similarly, the function $f_s(x)$ is sampled at S_1 points and the sample locations in the stopband along ω_1 axis are obtained according to the following transformation

$$\omega_{1(i+P_1)} = \frac{\pi - \omega_{1s}}{f_{sm}} f_s(x_i) + \omega_{1s}, \tag{2.5.13}$$

where $x_j = i/(S_1\text{-}1)$ for $i = 0, 1, ..., S_1\text{-}1$, and $f_{sm} = f_s(1)$.

The same formulas are used to obtain the grid line locations in ω_2 direction. The difference with the case considered in Sec. 2.4 is that the frequency samples in the region $(\omega_{1s}, \pi)\times(0, \pi)$ are not taken uniformly. Instead, these locations are calculated using the same expressions (2.5.10)-(2.5.13), except that ω_2 is used in place of ω_1. This sampling method produces filters with good performance for α between 1.0 and 1.5. The results are comparable with the previous two cases.

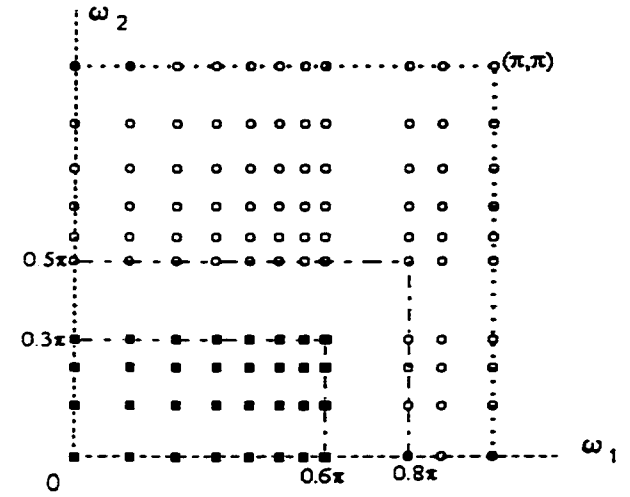
Example 2.7 Rectangular shape 2-D FIR filter designed from frequency samples taken at the vertices of a nonuniform grid obtained with exponential functions.

Filter frequency edge specifications:

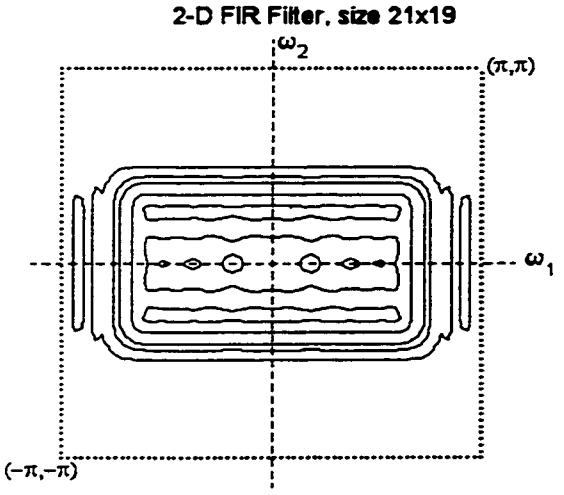
$|H_d(\omega_1, \omega_2)| = 1$ for $|\omega_1| \le 0.6\pi$ and $|\omega_2| \le 0.3\pi$
$|H_d(\omega_1, \omega_2)| = 0$ for $0.8\pi \le |\omega_1| \le \pi$ or $0.5\pi \le |\omega_2| \le \pi$

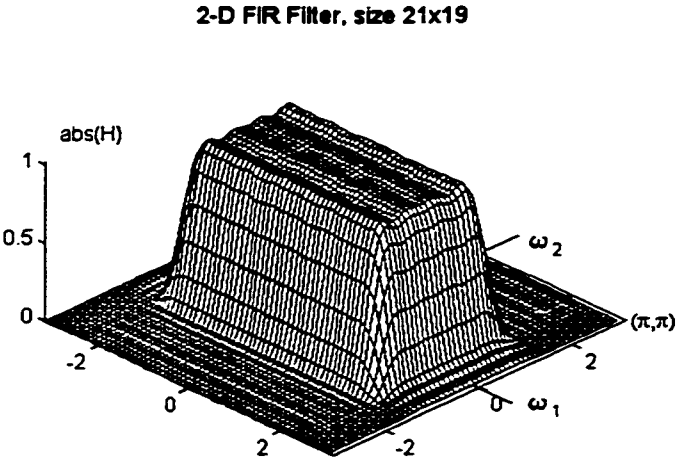
The number of lines in the pass- and stopbands is P_1=8, S_1=3, P_2=4, S_2=6, respectively, resulting in M_1=11 and M_2=10 grid lines in ω_1 direction and ω_2 direction, respectively. The corresponding filter is of size 21×19 points. The sample locations are shown in Fig.

2.12 (a), and the resulting frequency response is shown in Fig. 2.12 (b) and (c).



(a)





(b) (c)

Figure 2.12 A 21×19-point 2-D square shape zero phase FIR filter designed from frequency samples with locations obtained using exponential functions, $\alpha = 1.25$. (a) Sampling locations; (b) frequency response contour plot; (c) frequency response perspective plot.

The value of α in the above example is 1.25. The resulting filter impulse response is not square but the filter frequency response is still with zero phase. The maximum deviations are $\delta_p = 0.0283$ for the passband, and $\delta_s = 0.0398$ for the stopband. Although no samples

are taken in the region $\{\omega_{1s} < \omega_1 \le \pi\ ,\ \omega_{2p} < \omega_2 < \omega_{2s}\} \cup \{\omega_{1p} < \omega_1 < \omega_{1s},\ \omega_{2p} < \omega_2 \le \pi\}$ the filter behavior is good in this region. There is no need to take samples in the transition band either. This is typical for lowpass and highpass filters designed using the sampling technique with exponential functions. However, for bandpass and bandstop filters, as well as for some other sampling approaches, it is necessary to take samples in the above-mentioned frequency regions, as it will be shown in the next subsection.

The performance of some of the designed filters from frequency samples obtained by exponential functions is presented in Table 2.2. For all designs $\alpha = 1.25$. The performance is almost the same as that obtained with sampling on vertical lines. However, in the case of frequency samples on a rectangular grid the computational complexity is substantially reduced and the algorithm is much faster. The listing of Matlab code for obtaining the frequency samples, `nexp2g.m`, is presented in Appendix A. The filter impulse response is obtained using the same m file as in the previous case, `grd2d.m`.

These sampling approaches, harmonic series and exponential distribution, can be used for designing filters other than lowpass, as highpass, bandpass, and bandstop. The next example shows a bandpass, zero phase, square shape FIR filter.

Example 2.8 A bandpass square shape, zero phase FIR filter with edge specifications:

$|H_d(\omega_1, \omega_2)| = 1$ for $\{0.4\pi \le |\omega_1| \le 0.6\pi,\ 0.4\pi \le |\omega_2| \le 0.6\pi\}$
$|H_d(\omega_1, \omega_2)| = 0$ for $\{|\omega_1| \le 0.2\pi,\ |\omega_2| \le 0.2\pi\} \cup \{0.8\pi \le |\omega_1| \le \pi\} \cup \{0.8\pi \le |\omega_2| \le \pi\}$

Filter impulse response size: 35×35 points.

The number of independent filter coefficients is 324 out of a total of 1225. That is why the samples are taken at the vertices of a grid with M_1=18 lines along the ω_1 axis, and M_2=18 lines along the ω_2 axis, Fig. 2.13 (a). The number of lines in the passband is P_1=6 and P_2=6, and in the stopbands we have S_{11}=6, S_{12}=6 and S_{21}=6, S_{22}=6, respectively. The grid line locations are calculated using functions of the type (2.5.10) and (2.5.11). The deviations in the resulting frequency response magnitude are $\delta_{s1} = 0.000365$, $\delta_p = 0.0877$, and $\delta_{s2} = 0.000421$ in the first stopband, in the passband, and in the second passband,

respectively. As it can be seen on Fig. 2.13 (d), the maximum error in the passband occurs near its corners where no samples are taken. This effect can be eliminated by taking additional rows and columns of samples passing though the transition regions and this will be demonstrated in the next subsection.

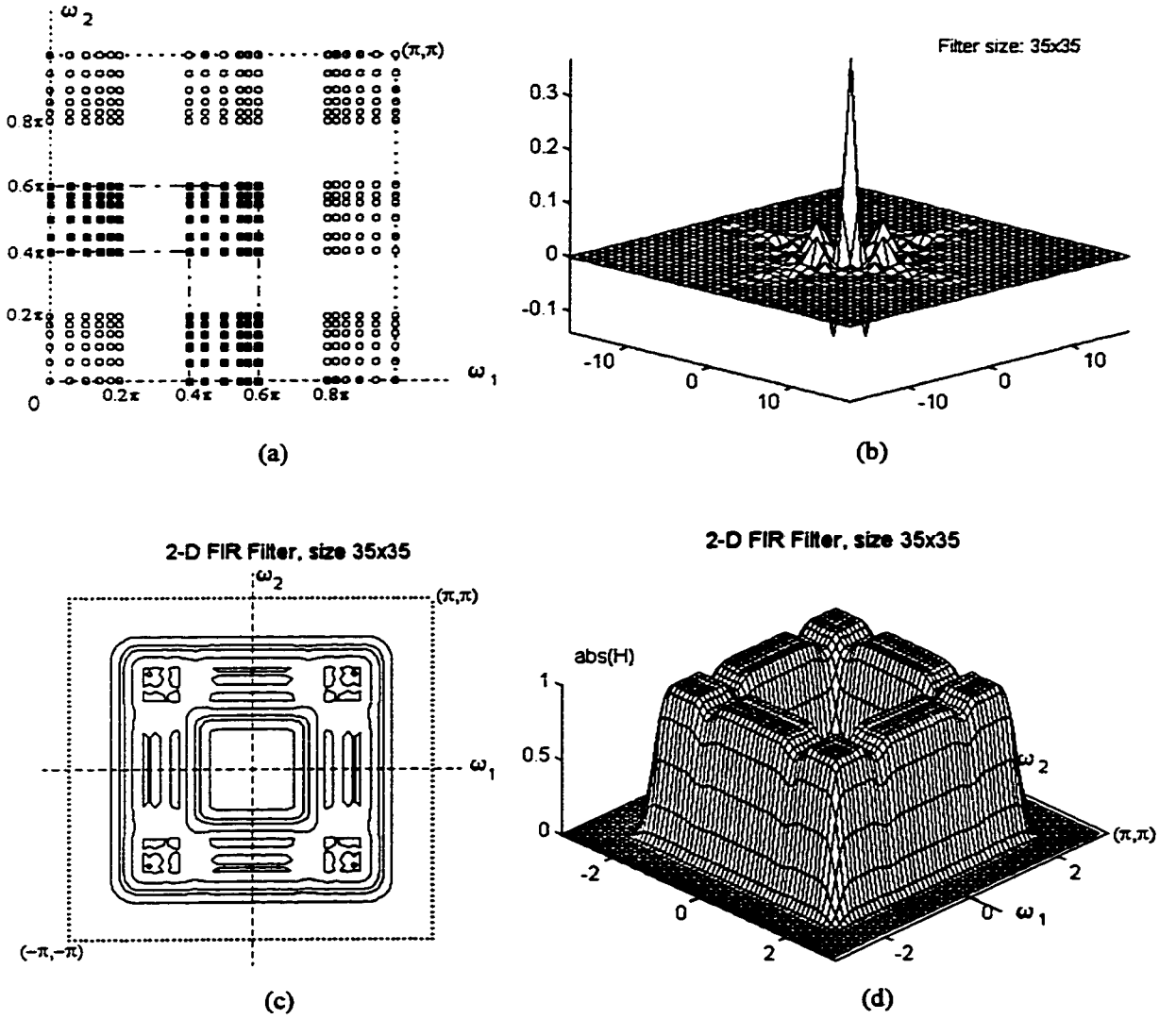


Figure 2.13 Example of a bandpass square shape, zero phase FIR filter of size 35×35 designed with exponentially distributed frequency samples. (a) Sample locations; (b) impulse response; (c) contour plot; (d) perspective plot.

A highpass zero phase FIR filter with the same frequency edge specification and the same impulse response size as a given lowpass FIR can be designed using the samples obtained for the lowpass filter by simply inverting their values, i.e., the samples with value 1 are set to 0, and vice versa. This is also true for the design of a band stop filter from given frequency samples for a bandpass filter (with the same frequency edge specs). The next example shows an highpass square shape, zero phase FIR filter designed from the sample locations used to design a lowpass filter, i. e., $\omega_{1s\ high} = \omega_{1p\ low}$ and so on.

Example 2.9 Design of a highpass square shape FIR filter using the sample locations and the inverted sample values used for designing a lowpass square shape filter, Fig. 2.14.

Lowpass filter frequency edge specifications:

$|H_{lp}(\omega_1, \omega_2)| = 1$ for $|\omega_1| \le 0.2\pi$ and $|\omega_2| \le 0.2\pi$
$|H_{lp}(\omega_1, \omega_2)| = 0$ for $0.4\pi \le |\omega_1| \le \pi$, any ω_2, $0.4\pi \le |\omega_2| \le \pi$, any ω_1

Highpass filter frequency edge specifications:

$|H_{hp}(\omega_1, \omega_2)| = 0$ for $|\omega_1| \le 0.2\pi$ and $|\omega_2| \le 0.2\pi$
$|H_{hp}(\omega_1, \omega_2)| = 1$ for $0.4\pi \le |\omega_1| \le \pi$, any ω_2, $0.4\pi \le |\omega_2| \le \pi$, any ω_1

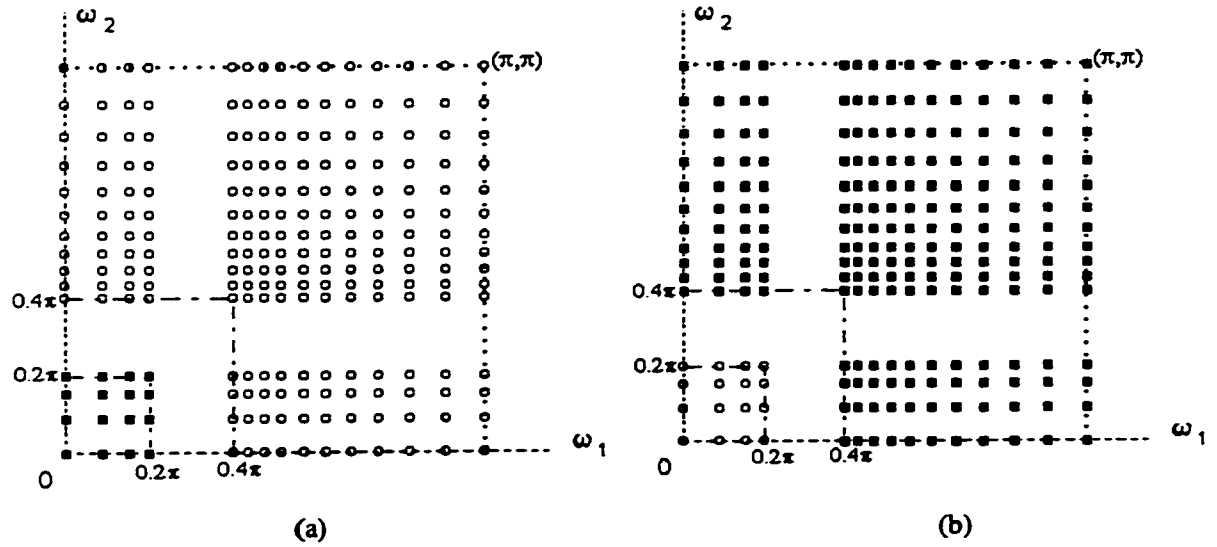


Figure 2.14 (a) Frequency samples for the design of a lowpass square shape filter; (b)the inverted samples from (a) in order to design a highpass filter. ■ -- sample value =1; o -- sample value = 0.

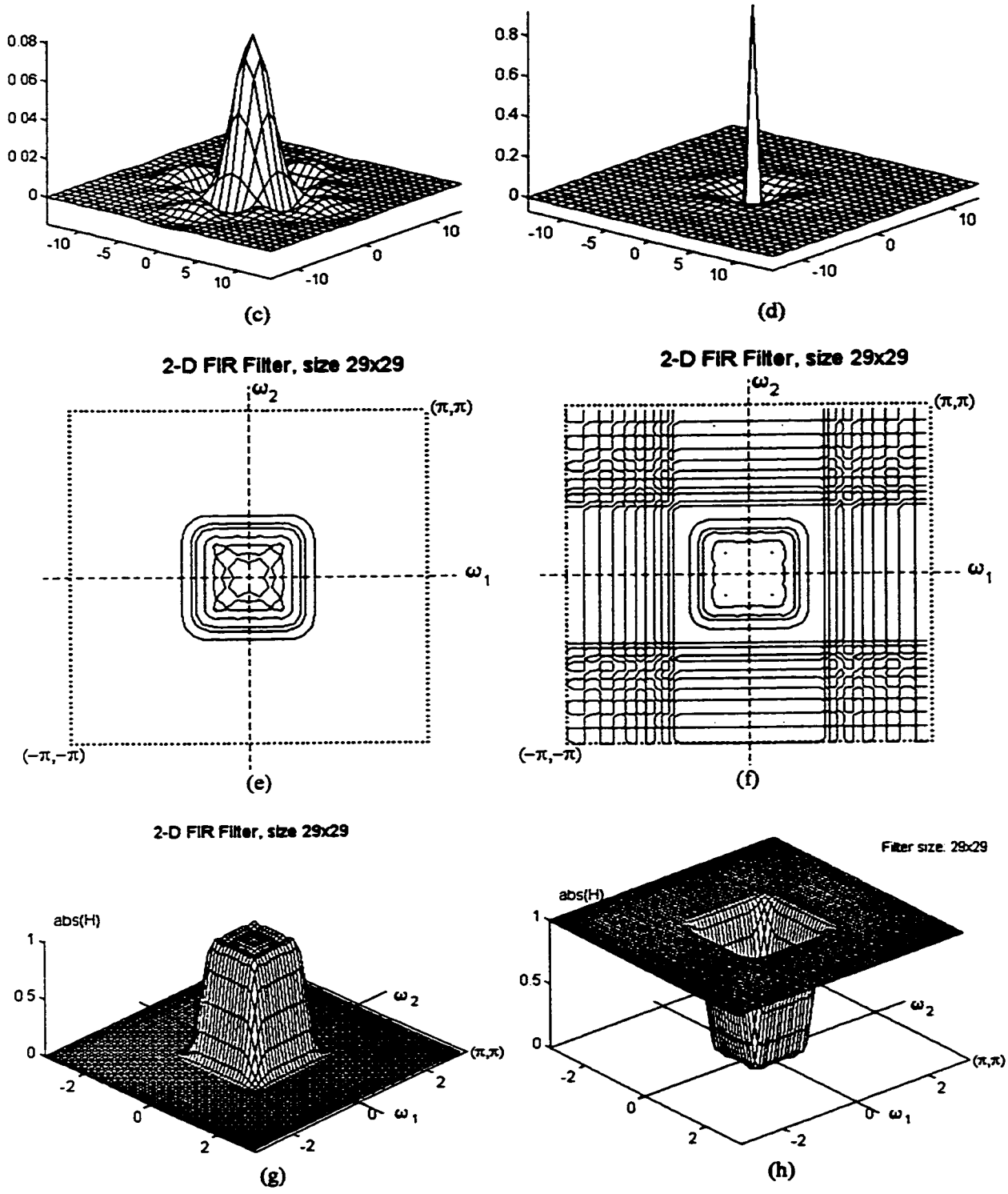


Figure 2.14 (continued) Example 2.9. (c) Impulse response of the lowpass and (d) of the highpass filter; (e) frequency response contour plot of the lowpass and (f) the highpass filters; (g) frequency response perspective plot of the lowpass and (h) the highpass filter.

A highpass 2-D digital filter can also be obtained from a given lowpass filter using the well-known relation

$$h_{hp}(n_1, n_2) = \delta(n_1, n_2) - h_{lp}(n_1, n_2) \tag{2.5.14a}$$

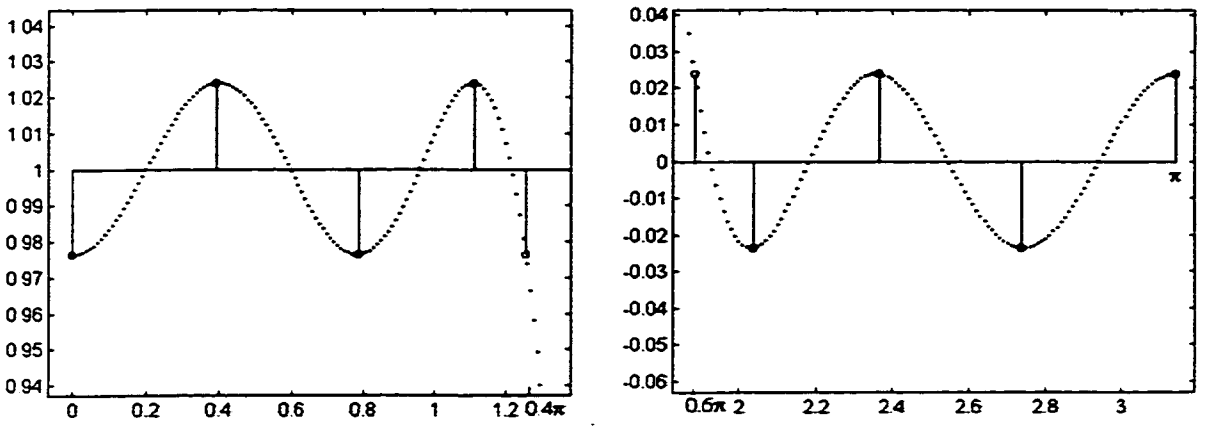
where $h_{hp}(n_1 , n_2)$ and $h_{lp}(n_1 , n_2)$ represent the highpass and lowpass filter impulse responses, respectively. The same relation can be used to obtain a bandstop filter from a bandpass filter:

$$h_{bs}(n_1, n_2) = \delta(n_1, n_2) - h_{bp}(n_1, n_2) \tag{2.5.14b}$$

Both methods, inverting the sample values with following interpolation or using (2.5.14), produce the same results. For example, the highpass filter in the last example designed from the samples shown in Fig. 2.14 (b) and the highpass filter with the same specifications but obtained using Eq. (2.5.14a) have the same performance in terms of shape and approximation error. The maximum error in both designs is $\delta_p = 0.002449$ and $\delta_s = 0.002363$ for the passband region and the stopband region, respectively. More design results are summarized in Table 2.2.

2.5.3 Nonuniform rectangular grid obtained from an 1-D optimal FIR filter.

The idea here is to use the extremal frequencies obtained with the Remez exchange algorithm as coordinates for the grid lines in the (ω_1, ω_2) plane. Using the Parks-McClellan algorithm [18] for optimal 1-D filter design (`remez.m` in Matlab), the extremal 1-D frequencies and the corresponding amplitude values at these frequencies are found. An example is shown in Fig. 2.15, where the amplitudes at the 10 extremal frequencies for an 1-D length-17 equiripple filter are denoted by circles.



(a) (b)

Figure 2.15 10 extremal points obtained with the Parks-McClellan algorithm for the design of an equiripple 1-D digital filter of order 16 and frequency edges $\omega_p = 0.4\pi$ and $\omega_s = 0.6\pi$.

As it was mentioned in Sec. 1.2, the extremal points are at least $N + 2$, where N is the order of the interpolation polynomial obtained as a linear combination of Chebyshev polynomials:

$$P(\cos\omega) = \sum_{k=0}^{N} b_k(\cos k\omega) \tag{2.5.15}$$

In the example above $N = 8$. If the length of a 1-D zero-phase prototype is N_{1D}, the number Q of extremal frequencies is (at least) [13, 18]

$$Q = N + 2 = (N_{1D}+3)/2 \tag{2.5.16}$$

A subset of $N+1$ extremal points is sufficient to design the 1-D filter. For example, if we want to complete the filter design from the extremal points given in Fig. 2.15, only 9 points are used. The program of McClellan-Parks-Rabiner [18], `remez.m` in Matlab, retains the first $N+1$ points, discarding the last one at π. However, for the design of 2-D FIR filters this extremal point at π is used also. Otherwise the error in the non-covered with samples area around $(0, \pi)$ and $(\pi, 0)$ is large. The extremal 1-D frequencies are placed along ω_1 and ω_2 axes. The frequency samples are then placed on the sides of

squares passing through the extremal frequencies, Fig. 2.16(a). All the samples on a particular square have the same value corresponding to the value of the 1-D extremal point.

Example 2.10 Design of a square shape, zero phase FIR filter using the extremal frequencies and corresponding amplitudes of the 1-D optimal filter shown in Fig. 2.15, $\omega_p = 0.4\pi$, $\omega_s = 0.6\pi$.

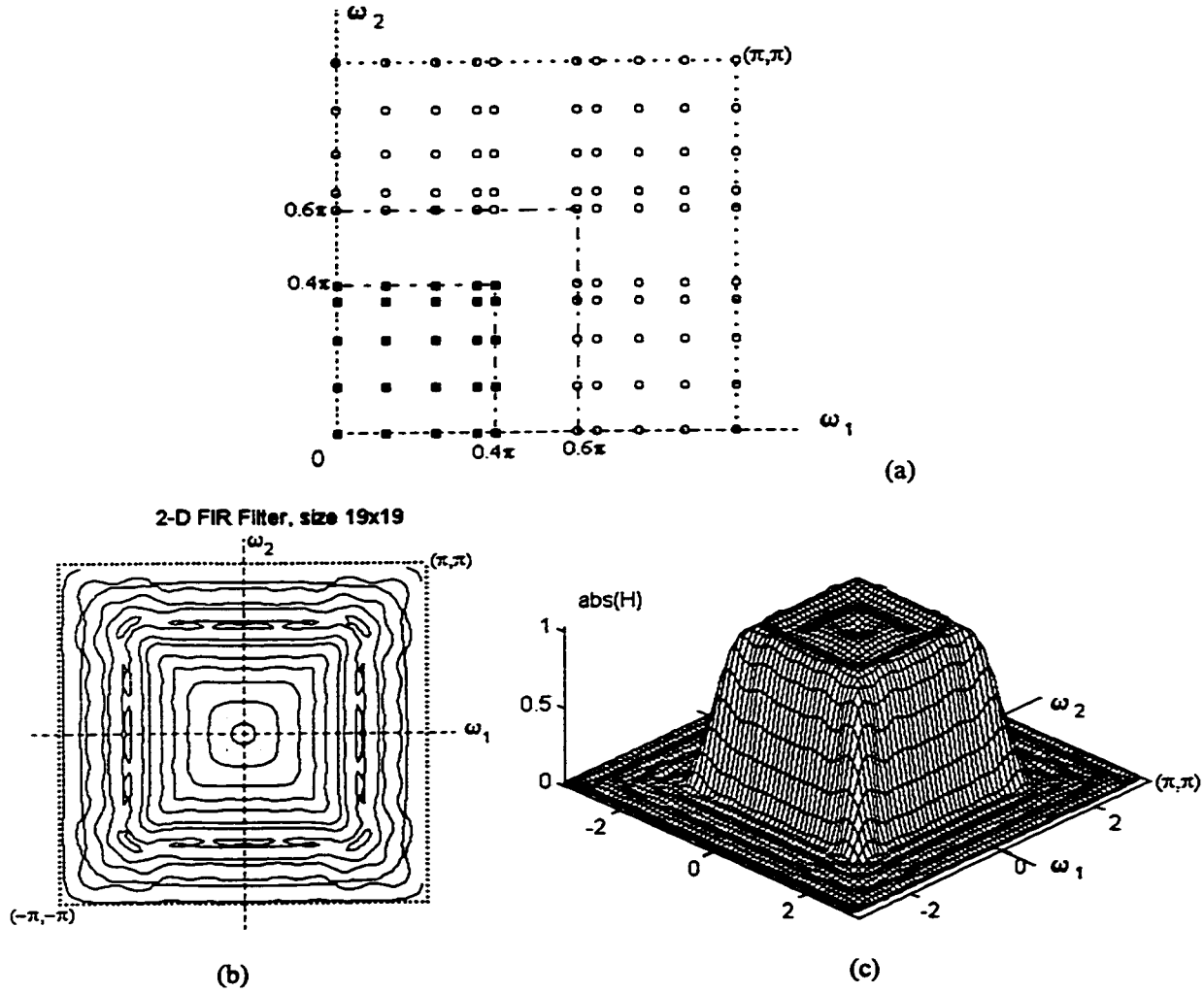


Figure 2.16 Example 2.10: (a) Samples at the vertices of a nonuniform grid obtained with the extremal frequencies for the 1-D filter, ■ -- sample value = 1±δ, o -- samples value = ±δ; (b) Contour plot of the resulting 2-D square filter; (c) perspective plot. Filter size 19×19.

The maximum ripple of the 1-D prototype filter is $\delta_p = \delta_s = \delta = 0.023763$. The total number of samples in the first quadrant is $(N+2)^2$. The resulting 2-D filter is of size 19×19 and the maximum error is $\delta_p = 0.0579$ and $\delta_s = 0.0683$ for the passband and stopband regions, respectively.

The method is computationally more intensive than the previous two methods, samples on a rectangular grid with exponential or harmonic distribution, because the Remez exchange algorithm must be used, once for square shape filters and twice for general rectangular shape. There is a small modification in the algorithm for the design of filters with rectangular shape. The extremal points for two equiripple 1-D filters are used, these of the first one are mapped along ω_1 axis, and the points of second one -- along ω_2 axis. The matrix containing the sample values at the grid vertices is obtained as an outer product of the samples of the two filters:

$$\hat{\mathbf{H}} = \mathbf{H}_1 \mathbf{H}_2^T \tag{2.5.17}$$

where $\mathbf{H}_1 = [H_{10}\, H_{11}\, .\, .\, H_{1N}\, H_{1,N+1}]^T$ and $\mathbf{H}_2 = [H_{20}\, H_{21}\, .\, .\, H_{2N}\, H_{2,N+1}]^T$ are $(1\times N+2)$ vectors containing the values at the extremal frequencies for the first and second 1-D filters, respectively. Obviously, both 1-D filters must be of the same order. The sample values are obtained using (2.5.17) because in this case the sample values lying on a given rectangular contour cannot have the same values, $H_{1k} \neq H_{2m}$, for $k \neq m$, even in most of the cases $H_{1k} \neq H_{2k}$. Consequently, the 2-D filters have increased passband and stopband peak error compared to the square shape case. The next example demonstrates this technique.

Example 2.11 2-D rectangular shape, zero phase FIR filter design with frequency samples at the vertices of a nonuniform rectangular grid obtained from the extremal frequencies for two 1-D equiripple filters. Frequency edge specifications

$$|H_d(\omega_1, \omega_2)| = 1 \quad \text{for} \quad |\omega_1| \le 0.4\pi \text{ and } |\omega_2| \le 0.2\pi$$

$$|H_d(\omega_1, \omega_2)| = 0 \quad \text{for} \quad 0.6\pi \le |\omega_1| \le \pi \text{ or } 0.4\pi \le |\omega_2| \le \pi$$

The grid line coordinates along ω_1 axis correspond to the extremal frequencies for the design of 1-D equiripple FIR filter of order 16 (length N_1 = 17) and passband and stopband edges

$$|H_1(\omega_1, \omega_2)| = 1\pm\delta \quad \text{for} \quad |\omega| \le 0.4\pi \text{ and } |H_1(\omega_1, \omega_2)| = \pm\delta \text{ for } 0.6\pi \le |\omega| \le \pi$$

Similarly, the coordinates along ω_2 axis are obtained from the extremal frequencies for the design of 1-D equiripple FIR filter of the same order 16 and passband and stopband edges

$$|H_1(\omega_1, \omega_2)| = 1\pm\delta \quad \text{for} \quad |\omega| \le 0.2\pi \text{ and } |H_1(\omega_1, \omega_2)| = \pm\delta \text{ for } 0.4\pi \le |\omega| \le \pi$$

The sample values are obtained using Eq. (2.5.17). The sample locations are presented in Fig. 2.17 (a).

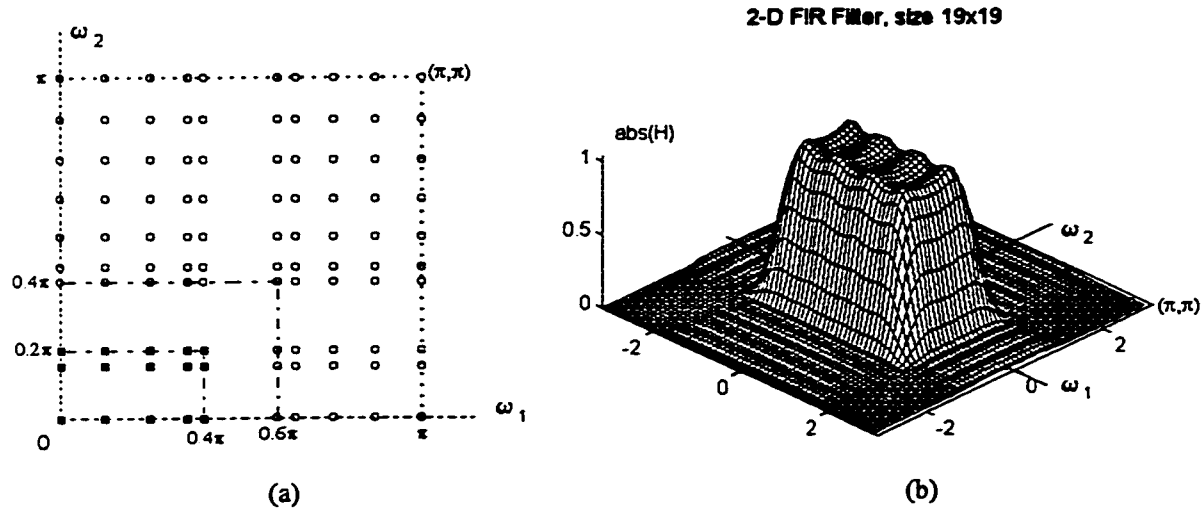


Figure 2.17 Example 2.11. (a) Sample locations; (b) Perspective plot of the resulting rectangular filter.

The impulse response of the designed filter $h(n_1, n_2)$ is determined as usual, using (2.5.4) and (2.2.10). The resulting 2-D filter is of size 19×19 points. The perspective plot of the resulting filter frequency response is shown in Fig. 2.17 (b).

An unwanted effect has been observed with this sampling technique. When the order of the 1-D prototype filters is relatively high, the filter magnitude in the transition band is not anymore monotonically decreasing. Instead, significant error peaks appear in this region. The transition band cannot be called anymore "don't care region", as in the 1-D

case. This problem can be solved by taking additional samples in the transition band. Some authors have used complicated linear programming techniques [19, 20] for finding optimal values for the transition band frequency samples in order to minimize the deviation in both pass- and stopbands. Here it will be shown that a linear interpolation of samples is reasonable and works well. Usually only one additional row and one additional column of samples is sufficient to eliminate the undesired effect. The frequency sample values on a positive slope of transition band are determined using (cf. Fig. 2.18(a))

$$\hat{H}(\omega_{1c}) = \frac{\omega_{1c} - \omega_{1a}}{\omega_{1b} - \omega_{1a}}\left[\hat{H}(\omega_{1b}) - \hat{H}(\omega_{1a})\right] + \hat{H}(\omega_{1a}), \qquad \omega_2 = const \tag{2.5.18a}$$

for $\omega_1 \geq \omega_2$. For $\omega_1 < \omega_2$, Eq. (2.5.18a) is used with interchanged places of ω_1 and ω_2. The frequency sample values on a negative slope of transition band are obtained from (cf. Fig. 2.18(b))

$$\hat{H}(\omega_{1c}) = \frac{\omega_{1b} - \omega_{1c}}{\omega_{1b} - \omega_{1a}}\left[\hat{H}(\omega_{1b}) - \hat{H}(\omega_{1a})\right] + \hat{H}(\omega_{1a}), \qquad \omega_2 = const \tag{2.5.18b}$$

for $\omega_1 \geq \omega_2$. For $\omega_1 < \omega_2$ the places of ω_1 and ω_2. interchange in (2.5.18b).

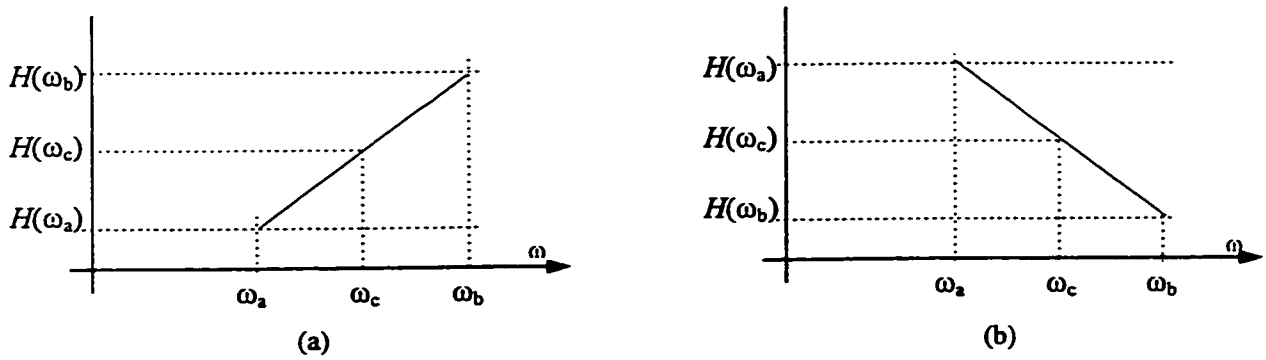


Figure 2.18 Frequency sample values in the transition band: (a) positive slope; (b) negative slope.

In the case when the additional samples are taken in the middle of the transition band their values are 0.5. The next example demonstrate the effect of additional samples in the transition band.

Example 2.12 Square shape FIR filter design with frequency samples obtained from an 1-D equiripple filter. Frequency edge specifications:

$$|H_d(\omega_1, \omega_2)| = 1 \quad \text{for} \quad |\omega_1| \le 0.2\pi \text{ and } |\omega_2| \le 0.2\pi$$
$$|H_d(\omega_1, \omega_2)| = 0 \quad \text{for} \quad 0.4\pi \le |\omega_1| \le \pi \text{ or } 0.4\pi \le |\omega_2| \le \pi$$

Filter size: 49×49 points

The sample locations and the frequency response obtained without samples in the transition band are shown in Fig. 2.19 (a), (c), and (e), respectively.

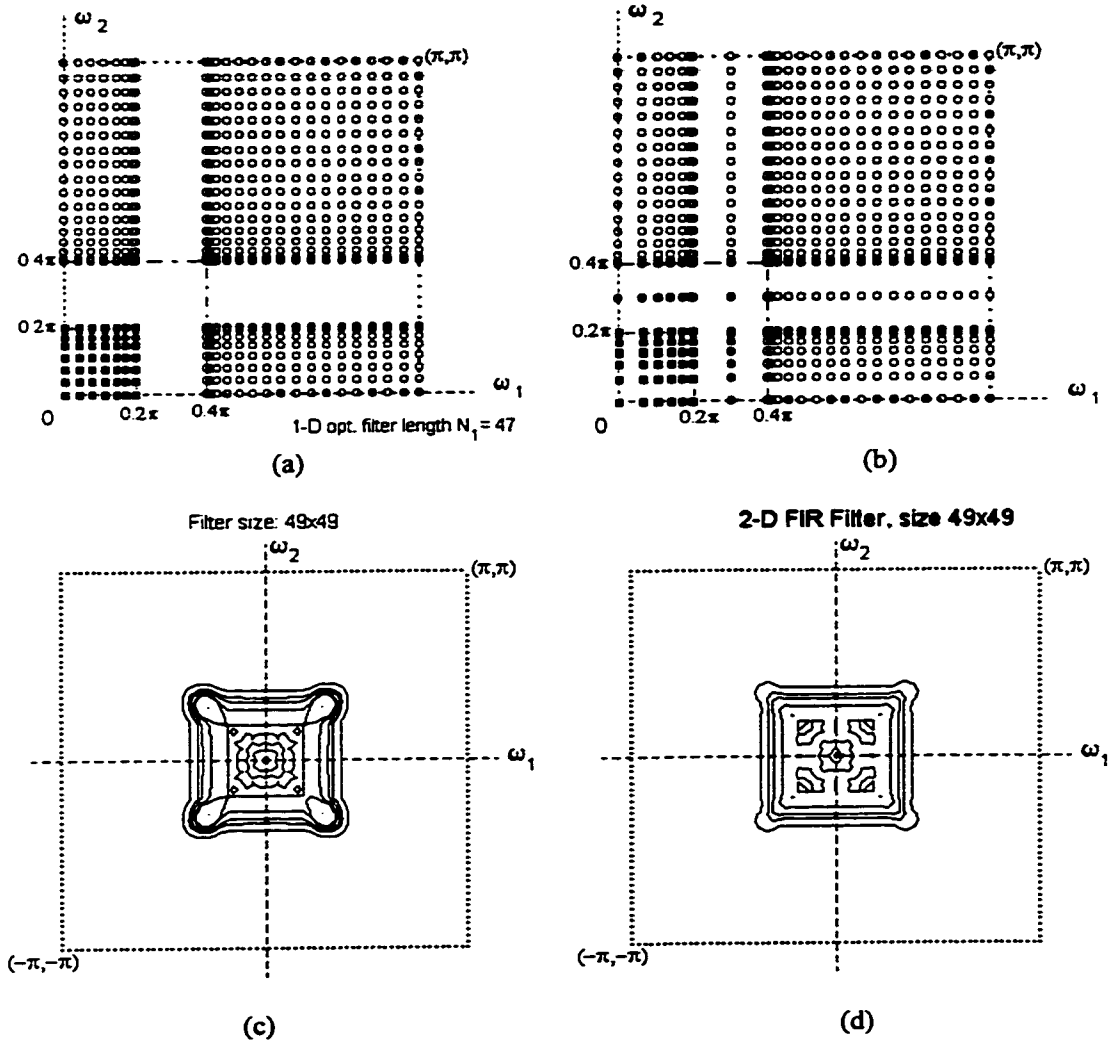


Figure 2.19 Example 2.12: Additional samples in the transition band. (a) No frequency samples in the transition band; (b) one additional row and one additional column of samples taken in the middle of the transition band. ■ -- sample value = 1±δ; o -- sample value = ±δ; • -- sample value obtained with linear interpolation of samples. (c) Frequency response contour plot of the filter designed from the samples shown in (a); (d) Contour plot of the filter designed with the additional samples in the transition band shown in (b).

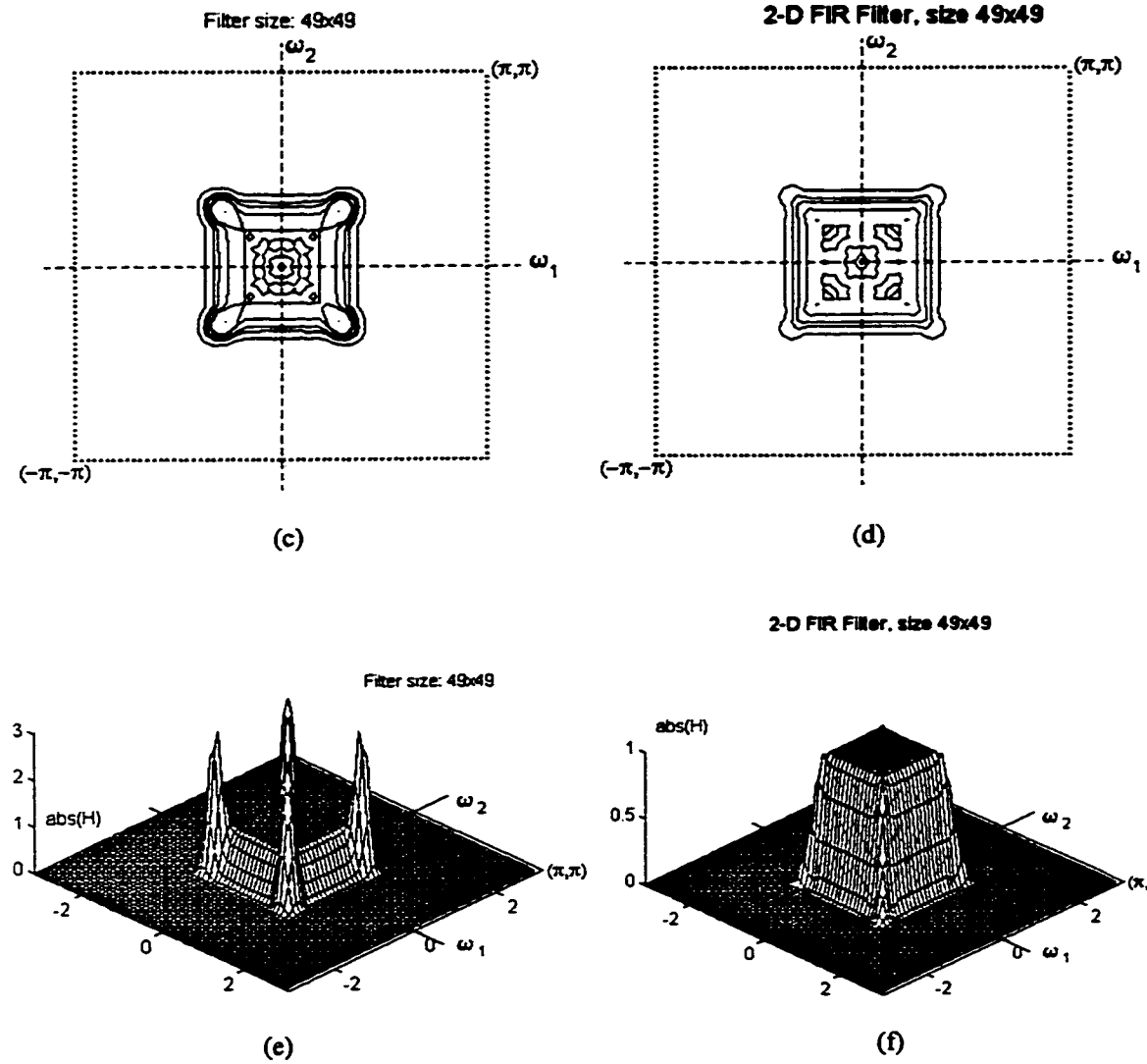


Figure 2.19 (continued) (e) Frequency response perspective plot of the filter designed without samples in the transition region;
(f) frequency response of the filter designed with the additional samples shown in (b).

In order to maintain the same 2-D filter size in the case of additional transition samples, the order of the 1-D prototype was decreased from 46 to 44. The contour and perspective plots of the filter designed with additional samples, taken according to (2.1.18), are shown in Fig. 2.19 (d) and (f). These sample locations are presented in Fig. 2.19 (b). The peak error in the second case is $\delta_p = 0.0014$ and $\delta_s = 0.0029$.

All filters in this subsection were designed using the Matlab function `grd2d.m` which was used in the previous cases of nonuniform rectangular grid. The frequency samples

were taken with a written for the purpose m file, npmcc2.m. The extremal frequencies and the corresponding amplitude values for the design of 1-D optimal filter were obtained from a modified version pmcc.m of the Matlab version remez.m of the McClellan-Parks-Rabiner computer program [18]. (By the way, in Matlab v.5 this modification is not necessary because the new remez.m returns a structure containing the extremal frequencies and the corresponding values.)

The last example with frequency samples obtained from the 1-D equiripple algorithm is to show that this sampling approach allows the design of good quality square shape, zero phase FIR filters.

Example 2.13 Square bandpass 2-D FIR filter design with frequency samples obtained from the 1-D optimal filter design algorithm. Frequency edge specifications:

$|H_d(\omega_1, \omega_2)| = 0$ for $|\omega_1| \le 0.2\pi, \quad |\omega_2| \le 0.2\pi$
$|H_d(\omega_1, \omega_2)| = 1$ for $\{0.4\pi \le |\omega_1| \le 0.6\pi, \ |\omega_2| \le 0.6\pi\} \cup \{|\omega_1| \le 0.6\pi, \ 0.4\pi \le |\omega_2| \le 0.6\pi\}$
$|H_d(\omega_1, \omega_2)| = 0$ for $\{0.8\pi \le |\omega_1| \le \pi, \text{ any } \omega_2\} \cup \{\text{any } \omega_1, \ 0.8\pi \le |\omega_2| \le \pi\}$

The frequency samples in 1-D are obtained with the following specifications:

$|H_{1d}(\omega')| = 0$ for $|\omega'| \le 0.2\pi$, $0.8\pi \le |\omega'| \le \pi$; $\quad |H_{1d}(\omega')| = 1$ for $0.4\pi \le |\omega'| \le 0.6\pi$.

1-D filter order = 34.

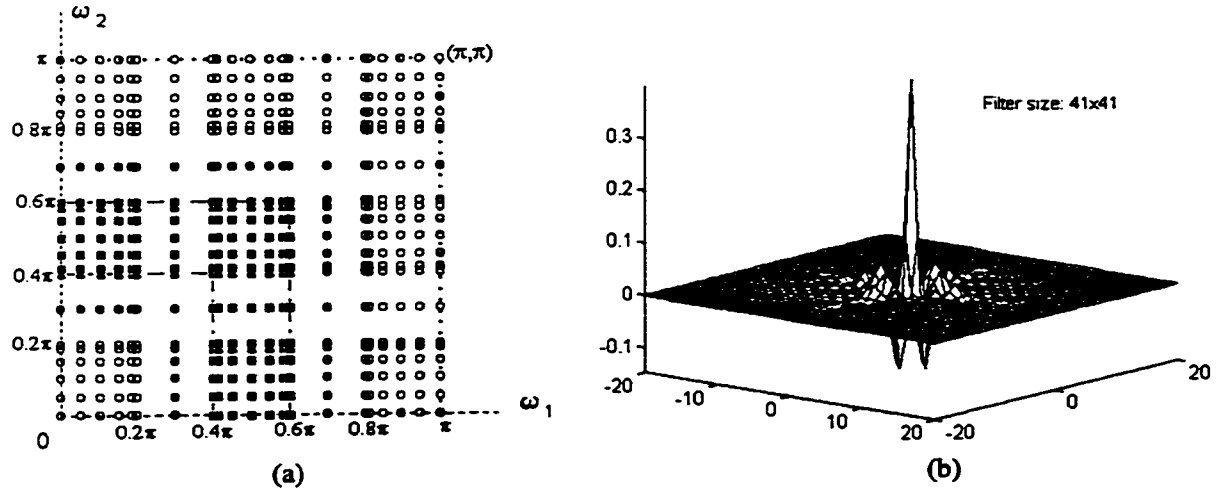
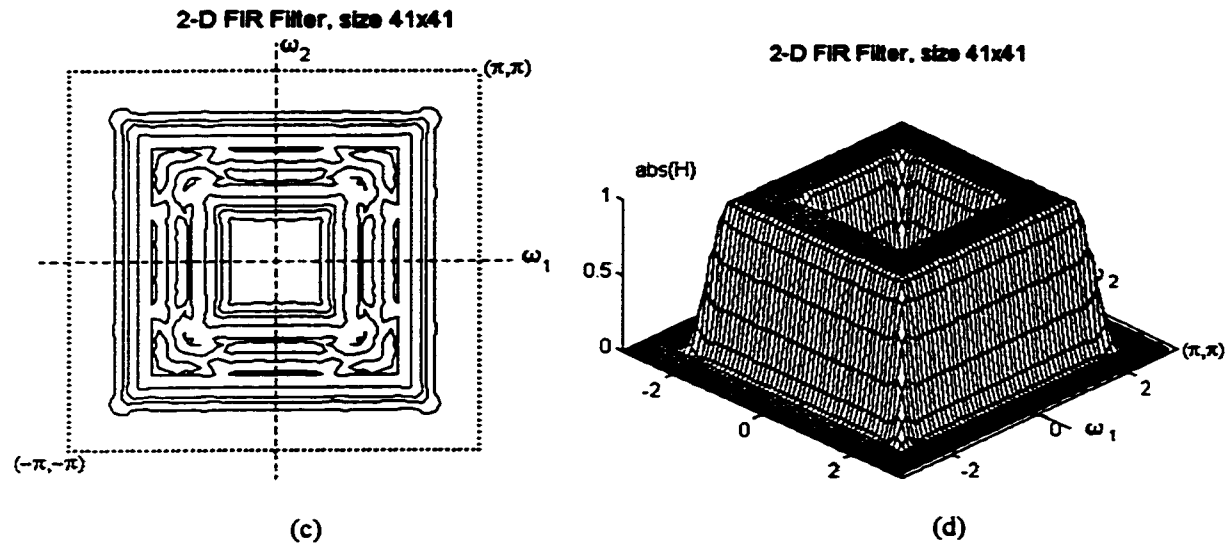


Figure 2.20 A bandpass filter designed from samples obtained using the 1-D Remez exchange algorithm. (a) Sample locations. Additional samples are taken in the two transition bands.
■ -- sample value = 1±δ, ○ -- sample value = ±δ, ● -- sample value obtained with interpolation; (b) Impulse response of the resulting 2-D FIR filter; (continues)



(c) (d)

Figure 2.20 (continued) A bandpass filter designed from samples obtained using the 1-D Remez exchange algorithm. (c) Contour plot of the frequency response; (d) perspective plot of the frequency response.

The 1-D optimal filter algorithm produces 19 extremal frequencies and corresponding frequency response values. In the middle of each transition band one additional point is taken whose value is found by linear interpolation using the neighboring samples. Thus, in 1-D the total number of frequency samples is 21, giving rise to 21×21 samples in the 2-D frequency plane. That is why the resulting 2-D filter is of size 41×41 points. The approximation error can be found in Table 2.2.

2.5.4 Sampling analytic functions based on Chebyshev polynomials.

This approach for obtaining the sample locations in the (ω_1, ω_2) plane and the sample values has been used by Bagchi and Mitra, [2]. In this subsection, the method is demonstrated again with some minor modifications. Firstly, some modifications has been done in the way of generating the 1-D analytic functions. Secondly, the frequency samples in the 2-D frequency plane are taken exactly on the vertices of a nonuniform

grid, obtained from the 1-D sample coordinates, and not as in [2], i. e., on square contours passing through the extrema (see below). Thus, the calculation of the filter coefficients can be done using (2.5.4), guaranteeing the existence and uniqueness of the solution.

The simplest way the Chebyshev polynomials $T_n(x)$ of n-th order is defined is

$$T_n(x) = \cos(n \arccos x) \qquad |x| \le 1 \tag{2.5.19a}$$

$$x = \cos\omega \quad \text{or} \quad \omega = \arccos x,$$

$$\text{and} \qquad T_n(x) = \cosh(n \operatorname{acosh} x), \qquad |x| > 1 \tag{2.5.19b}$$

The turning points of $T_n(x)$ in [-1, 1] occur at the zeros of $\sin n\omega / \sin\omega$. Together with the endpoints $x = -1$ and $x = 1$, the number of extrema in the interval [−1, 1] is n+1. They occur at

$$\omega_k = \frac{k\pi}{n}, \qquad x_k = \cos\omega_k = \cos\frac{k\pi}{n}, \qquad k = 0, 1, \ldots, n \tag{2.5.20}$$

The n zeros of $T_n(x)$ occur at

$$\omega_i = \frac{(2i+1)\pi}{2n}, \qquad x_i = \cos\frac{(2i+1)\pi}{2n}, \qquad i = 0, 1, \ldots, n-1 \tag{2.5.21}$$

The Chebyshev polynomials are orthogonal over the above two discrete sets of points.

The desired amplitude response of a zero phase 1-D FIR filter is generated as follows:

$$H(\omega) = \begin{cases} H_p(\omega) = b_1\left[1 - \delta_p T_p(x)\right], & |\omega| \le \omega_p \\ H_s(\omega) = b_2\left[\delta_s T_s(x)\right] + b_3 & \omega_s \le |\omega| \le \pi \end{cases} \tag{2.5.22}$$

where δ_p and δ_s are the peak ripples in the passband and stopband, respectively, ω_p and ω_s are the passband and stopband frequency edges, respectively. The constant b_1 is used to linearly map the interval [−1, 1] to [$-\omega_p$, ω_p], and b_2, and b_3 are used to map the interval [−1, 1] to [ω_s, $\pi+\omega_s$].

$$
\begin{aligned}
b_1 &= \omega_p \\
b_2 &= \pi - \omega_s \\
b_3 &= \pi
\end{aligned}
\tag{2.5.23}
$$

In the present work the linear transformation (2.5.22) is used, which is much simpler than the nonlinear transformation used in [2]. Both produce similar results, as it will be shown. The order p of the Chebyshev polynomial used in the passband depends of the required number of P samples in this band. And the order s of the polynomial used for the stopband is determined from S, the number of samples in the stopband. And P and S are determined† from the 1-D filter length N_{1D} which can be approximately estimated [2, 13] by choosing the lowest odd value

satisfying

$$
N_{1D} \geq \frac{-10\log\left(\delta_p \delta_s\right) - 13}{2.324\left(\omega_s - \omega_p\right)} + 1 . \tag{2.5.24}
$$

Due to the fourfold symmetry, P and S are related to N_1 by

$$
P + S = (N_1+1)/2 . \tag{2.5.25}
$$

Since P is the number of samples in $[0, \omega_p]$ and S is the number of samples in $[\omega_s, \pi]$, they are taken in proportion to the sizes of the passband and stopband and such that (2.5.25) is satisfied. The samples are chosen to be taken at the extrema of the Chebyshev polynomials in the intervals [0, 1] and [−1, 0], which correspond to the passband $[0, \omega_p]$ and the stopband $[\omega_s, \pi]$, respectively. A Chebyshev polynomial of even order n has $(n+2)/2$ extrema in the interval [0, 1] or [−1, 0], including the end points. Therefore, the orders p and s of the Chebyshev polynomials used in (2.5.22) are given by

$$
\begin{aligned}
p &= 2P - 2 \\
s &= 2S - 2
\end{aligned}
\tag{2.5.26}
$$

The choice of extrema as sample locations is motivated by the fact that in this way the sensitivity of the frequency response to a perturbation in the sample locations is

† Note that here the notations P and S are defined in a different way than in [2].

minimized. The proof can be found in [2]. The Parks-McClellan algorithm also uses the extrema of the error function as points for interpolation. According to the Alternation theorem, the use of these points for interpolation will guarantee minimization of the maximum approximation error and equiripple behavior. The difference with the considered in this subsection method is that the initial locations are improved with each iteration until the optimum is found. Here, this is performed in one step.

The values of ω at which the functions (2.5.22) are sampled are mapped one to one along ω_1 and ω_2 axes. Trough these frequencies pass the grid lines. The frequency samples in the (ω_1, ω_2) plane are taken at the vertices of this grid with equal values on each square contour and having the value of (2.5.22) for the corresponding 1-D frequency.

The next example illustrates this sampling technique.

Example 2.14 A lowpass square shape, zero phase FIR filter design with frequency samples obtained by sampling the analytic functions (2.5.22). Filter specifications:

$$|H_d(\omega_1, \omega_2)| = 1 \pm 0.01 \quad \text{for} \quad |\omega_1| \le 0.35\pi \text{ and } |\omega_2| \le 0.35\pi$$
$$|H_d(\omega_1, \omega_2)| = \pm 0.01 \quad \text{for} \quad 0.65\pi \le |\omega_1| \le \pi \text{ or } 0.65\pi \le |\omega_2| \le \pi.$$

Substituting $\delta_p = \delta_s = 0.01$, $\omega_p = 0.35\pi$, $\omega_s = 0.65\pi$ in (2.5.24) we obtain $N_{1D} \ge 13.35$. The lowest odd value is $N_{1D} = 15$.

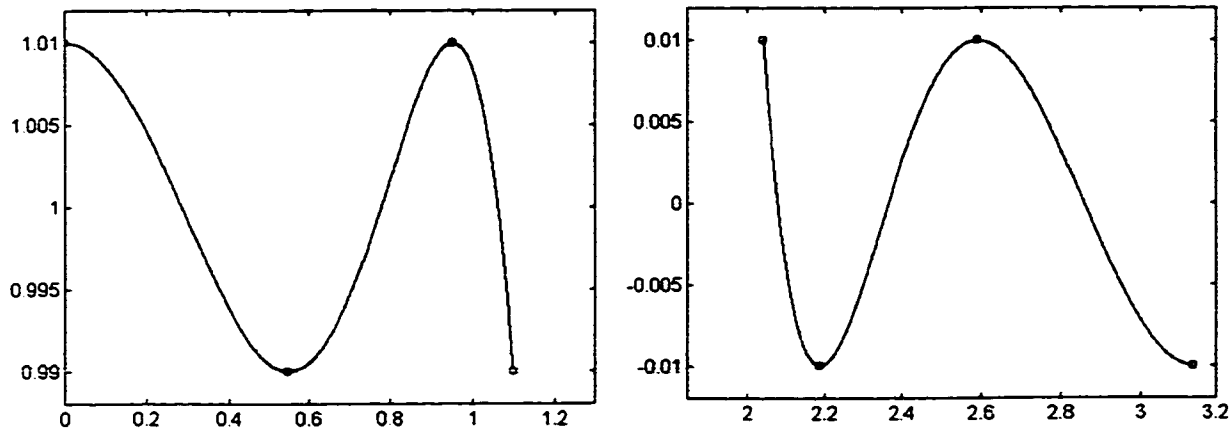
$P + S = 8$ and since the passband and stopband have equal width of 0.35π,

$P = 4$ and $S = 4$. The orders of the Chebyshev polynomials are

$p = 2P - 2 = 6$ and $s = 2S - 2 = 6$. Therefore,

$$H_p(\omega) = 0.35\pi[1 - 0.01T_6(x)], \qquad |\omega| \le 0.35\pi, \qquad |x| \le 1$$
$$H_s(\omega) = (\pi - 0.65\pi)[0.01\ T_6(x)] + \pi, \qquad 0.65\pi \le |\omega| \le \pi, \quad |x| \le 1$$

These functions are presented in Fig. 2.21 (a) and (b), respectively. They are sampled at the extrema of the corresponding intervals, denoted with circles.



(a) (b)

Figure 2.21 The functions $H_p(\omega)$, (a) and $H_s(\omega)$, (b) for Example 2.14. Chebyshev polynomials of 6-th order are used. The samples at the extrema are denoted with "o".

The resulting frequency vector which is mapped onto ω_1 and ω_2 axes is ω_k = [0, 0.5045 0.8906, 1.0996, 2.042, 2.251, 2.6371, π]^T , and the corresponding sample values are H_k = [1.01, 0.99, 1.01, 0.99, 0.01, -0.01, 0.01, -0.01]^T. Therefore, the number of grid lines in the (ω_1, ω_2) plane is 8×8, the total number of samples is 64, and the resulting 2-D FIR filter is of size 15×15 points. The sample locations are shown in Fig. 2.22 (a). The impulse response of the filter is shown in Fig. 2.22 (b), and the frequency response contour and perspective plots are presented in (c) and (d), respectively. The resulting peak passband error is δ_p = 0.0147 and the peak stopband error is δ_s = 0.0149. As it was mentioned above, the use of a linear mapping of the interval [−1, 1] to [$-\omega_p$, ω_p], and to [ω_s, $\pi+\omega_s$] produces similar results as the more elaborated nonlinear mapping used in [2]. For example, a 9×9 filter with frequency specifications as in Example 2.14 designed using the mapping proposed in [2] has peak errors δ_p = 0.0322 and δ_p = 0.0471, while the same filter designed with the mapping proposed in this work exhibits peak errors δ_p = 0.0250 δ_s = 0.0804.

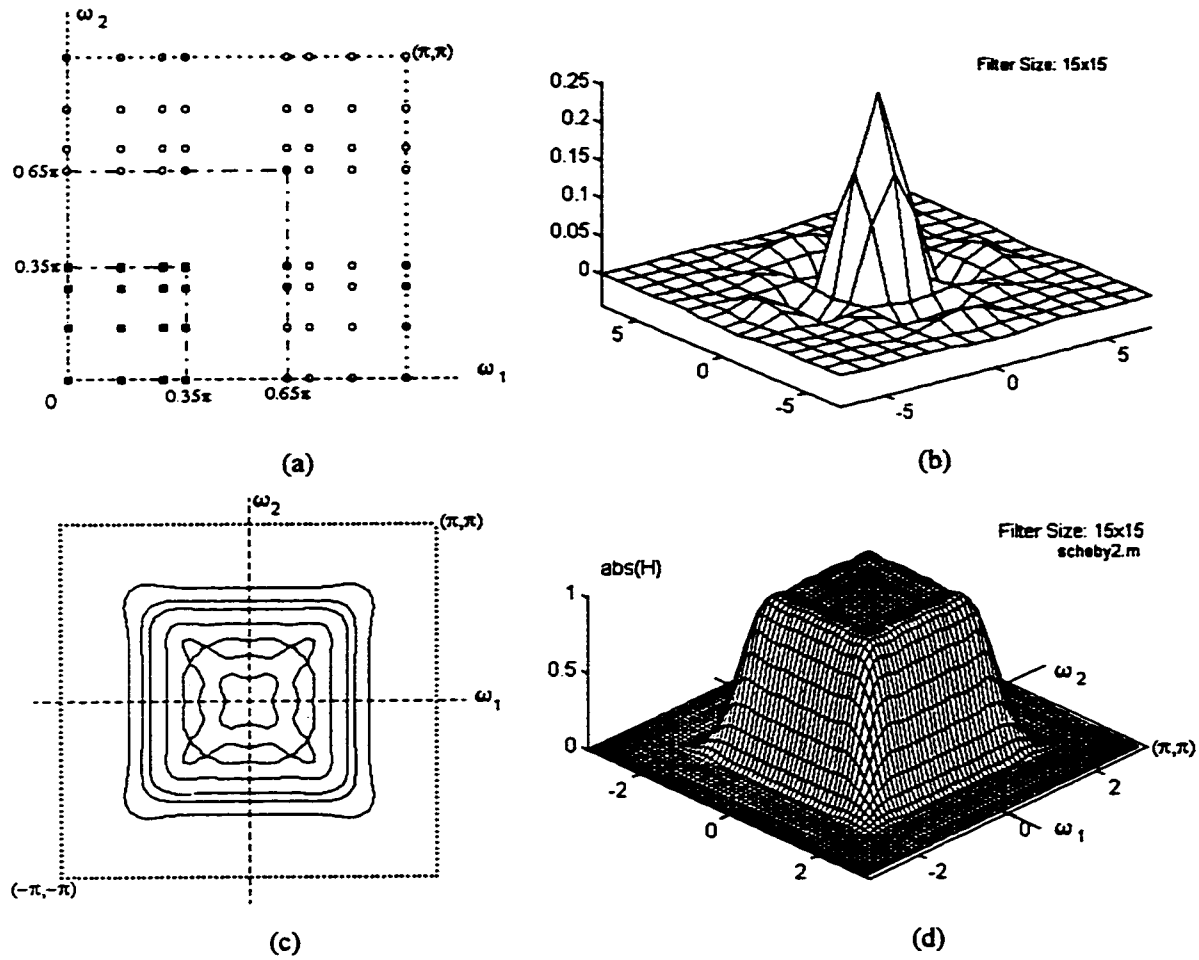


Figure 2.22 Example 2.14: a lowpass 15×15 filter designed by sampling 1-D analytic functions, Eq.(2.5.22) (a) Sample locations. ■ -- sample value = 1±δ, o -- sample value = ±δ, (b) Impulse response of the resulting 2-D FIR filter; (c) Contour plot of the frequency response; (d) perspective plot of the frequency response.

The design results obtained with frequency samples at the vertices of a nonuniform rectangular grid are summarized in Table 2.2. It can be seen that the performance of exponential distribution of samples is slightly over that of harmonic distribution. The best performance with respect to maximum deviations in the pass- and stopbands produces the technique using Chebyshev polynomials. However, the good results of this technique are only for square shape filters. For rectangularly (not square) shaped filters better results can be obtained with the exponentially and harmonically distributed frequency samples. This is due to the fact that in the case of non-square shape the use of extrema of Chebyshev polynomials or the extrema of 1-D optimal filter the sample values are

determined as an outer product, which is equivalent to a separable design. A separable 2-D square filter $h(n_1, n_2)$ is designed from two 1-D filters $h_1(n)$ and $h_2(n)$, $h(n_1, n_2) = h_1(n)$ $h_2(n) \Leftrightarrow H(\omega_1, \omega_2) = H_1(\omega)\, H_2(\omega)$.

Filter Size	Passb. edges ω_{1p}, ω_{2p} (ω_{1p1}, ω_{1p2} ....)	Stopb. edges ω_{1s}, ω_{2s} (ω_{1s1}, ω_{1s2} ...)	Passband dev., δ_p	Stopband dev., δ_s (δ_{s1} δ_{s2})	Algorithm (Sampling)	α (β) (N_1)	Matlab m. File
9×9	0.4π	0.6π	0.1912	0.2937	Uniform	-	fsampdr
9×9	0.4π	0.6π	0.1036	0.3035	Harm .	1.0	sharm3
9×9	0.35π	0.65π	0.0317	0.2462	Harm .	1.0	sharm3
9×9	0.35π	0.65π	0.0314	0.2460	Exp	1.25	nexp2g
9×9	0.45π	0.75π	0.1087	0.0761	Exp	1.25	nexp2g
9×9	0.45π	0.75π	0.1352	0.1133	1-D Opt	7	pmcc2
9×9	0.35π	0.65π	0.0250	0.0804	Cheb. poly	-	scheby2
9×9	0.35π	0.65π	0.0322	0.0471	Cheb. poly	-	Ref.[2]
11×11	0.35π	0.65π	0.2027	0.1496	Uniform	-	fsampdr
11×11	0.35π	0.65π	0.0586	0.0349	Harm	1.0	sharm3
11×11	0.35π	0.65π	0.0582	0.0346	Exp	1.25	nexp2g
11×11	0.45π	0.75π	0.5075	0.3510	1-D Opt	9	pmcc2
11×11	0.35π	0.65π	0.0452	0.0288	Cheb. poly	-	scheby2
15×15	0.25π, 0.45π	0.55π, 0.75π	0.1993	0.1341	Uniform	-	fsampdr
15×15	0.25π, 0.45π	0.55π, 0.75π	0.0239	0.0205	Harm	1.0	sharm3
15×15	0.25π, 0.45π	0.55π, 0.75π	0.0240	0.0198	Exp	1.25	nexp2g
15×15	0.35π	0.65π	0.0147	0.0149	Cheb. poly	-	scheby2
19×19	0.2π	0.4π	0.2150	0.1171	Uniform	-	fsampdr
19×19	0.2π	0.4π	0.0581	0.0221	Exp	1.25	nexp2g
19×19	0.2π	0.4π	0.0552	0.0460	1-D Opt	17	npmcc2
19×19	0.4π	0.6π	0.0354	0.0178	Exp	1.25	nexp2g
21×21	0.35π	0.65π	0.1034	0.1115	Uniform	-	fsampdr
21×21	0.35π	0.65π	0.0011	0.0063	Harm	1.0	sharm3
21×19	0.6π, 0.3π	0.8π, 0.5π	0.0283	0.0397	Exp	1.25	nexp2g
29×29	0.2π	0.4π	0.02363	0.00245	Exp	1.25	nexp2g
29×29H	0.2π	0.4π	0.00245	0.02363	Exp	1.25	nexp2g
31×31	0.45π	0.55π	0.2633	0.1155	Uniform	-	fsampdr
31×31	0.45π	0.55π	0.1069	0.0548	Harm	0.1	sharm3
31×31	0.45π	0.55π	0.0701	0.0364	Exp	0.75	nexp2g
31×31	0.45π	0.55π	0.0951	0.0657	1-D Opt	29	npmcc2
35×35B	0.4π, 0.6π	0.2π, 0.8π	0.0876	0.000365	Exp	1.25	nexp2b
	0.4π, 0.6π	0.2π, 0.8π		0.000421			
35×35B	0.4π, 0.6π	0.2π, 0.8π	0.0612	0.0225	1-D Opt	33	npmcc2b
	0.4π, 0.6π	0.2π, 0.8π		0.0272			
41×41B	0.4π, 0.6π	0.2π, 0.8π	0.0104	0.0037	1-D Opt.	35	npmcc2b
s. in TB	0.4π, 0.6π	0.2π, 0.8π		0.0252			

Table 2.2 Performance comparison (caption on the next page).

Table 2.2 Performance comparison of rectangular shape FIR filters designed with frequency samples at the vertices of a nonuniform rectangular grid..

Uniform -- uniform sampling;

Harm -- sampling using harmonic series, Sec. 2.5.1;

Exp -- sampling using exponential functions, Sec. 2.5.2;

1-D Opt -- sample coordinates from extrema of 1-D optimal design; Sec. 2.5.3,

N_1 -- 1-D prototype length;

Cheb. poly -- sample coordinates from the extrema of shifted and scaled Chebyshev polynomials.

No letter after the size -- lowpass filter,

H -- highpass filter; **B** -- bandpass filter;

s. in **TB** -- additional samples in transition bands.

2.6 Summary

In this chapter several nonuniform sampling techniques have been proposed that are pertinent for the design of 2-D rectangular shape, zero phase FIR filters. These techniques can be grouped into two categories: (i) frequency sampling on parallel lines and (ii) frequency sampling at the vertices of a nonuniform rectangular grid. A common feature of both techniques is that the interpolation problem has a unique solution, or stated otherwise, the NDFT matrix is nonsingular. The two approaches, parallel lines and nonuniform grid, are very pertinent for the rectangular filter design because in these cases the frequency sample locations "naturally" match the shape of the filter. The practice has shown that the best results with the frequency sampling techniques are obtained when the samples are taken on contour lines that match the desired filter shape.

The sampling techniques proposed here, exponential and harmonic distribution of frequency samples, have shown that the flexibility of choosing the sample locations is not the only advantage of the nonuniform sampling. Rectangular shape filters with good

performance can be designed using simple and fast algorithms. It should be noted that parameters α and β (cf. Eqs. (2.4.7), (2.4.8), (2.5.6)) can be optimized and then even increased filter performance can be expected. The method based on Chebyshev polynomials gives superior filter performance in terms of maximum passband and stopband deviations. However, not much can be done to improve this technique, except maybe the application of some nonlinear transformation of the region $[-1, 1]$ to $[-\omega_p, \omega_p]$ or to $[-\omega_s, \omega_s]$. Unfortunately, there are not too much design examples of rectangular shape filters in the literature and a comprehensive comparison of the results with other methods for 2-D FIR filter design cannot be made.

It is an interesting fact that the design based on the extrema of 1-D optimal prototype does not give the expected performance. This simply is due to the fact that the Alternation theorem does not hold in 2-D and the implication is that the sample locations for the best in Chebyshev sense design in 2-D cannot be obtained by simply extending the 1-D case to 2-D. The critical point sets in 2-D are further investigated in the next chapter dealing with 2-D circularly symmetric FIR filters.

Chapter 3
Design of Circularly Symmetric FIR Filters

In this chapter the design of 2-D circular shape FIR filters with nonuniform frequency sampling is considered. After a brief foreword, the existing symmetries in the frequency response and impulse response of circular shape FIR filters are discussed in Section 3.1. Also in this section, the generalized polynomial expression of the frequency response is given which will be used to solve the 2-D interpolation problem. The minimum number of frequency samples, tolerance scheme specifications, domain of approximation, and sampling parameters are also discussed. In Section 3.2, a sampling technique for designing circular shape FIR filters is introduced. This technique is based on exponential distribution of circular contours on which the frequency samples are taken. A class of sampling techniques based on extrema obtained from a 1-D optimal FIR filter design are proposed and investigated in the next section, Section 3.3. These techniques are presented in subsections 3.3.1 to 3.3.3, respectively. Some considerations concerning the sampling density and the interpolation method are presented in Section 3.4. A least-squares approach is proposed and demonstrated in this section as a way to improve the solution accuracy. At the end, the results of the proposed sampling methods are summarized and compared to results obtained with conventional 2-D FIR filter design methods.

The 2-D FIR filters with circular symmetry have enjoyed special interest in the past decades. It is primarily due to the fact that the circular symmetry does not give

preferential treatment in any particular direction in the frequency domain. Also, the circular symmetry is very attractive since the frequency response is isotropic for a constant radius and that suggests easy extension of some of 1-D filter design methods to 2-D, e.g. the rotated window method [25]. In this method, a 2-D window $w(n_1, n_2)$ is obtained by rotating a 1-D analog window $w_a(t)$ and sampling the resulting 2-D analog window $w(t_1, t_2)$. One of the most popular windows is the rotated Kaiser window which will be used in this chapter for comparison. The early studies of 2-D FIR filter design with uniform frequency sampling consider FIR filters with circularly symmetric frequency response, see for example Hu and Rabiner [20] and Fiasconaro [19]. The McClellan transformation [24] proved to be very suitable for circular FIR filter design. This transformation was extensively studied and further developed, e.g. [A1]. It produces high performance filters, in some restrictive cases even optimal filters [10], and this method is usually considered in elaborated applications requiring high performance. The iterative design techniques for optimal 2-D linear phase FIR filters [21, 22], too, are based on filters with circular shape frequency response.

The design of circularly symmetric FIR filters is attractive because the eightfold symmetry suggests great savings and the circular shape hints for simplifications using rotations. On the other hand the design of circular filters with frequency samples is challenging since a rectangular grid, even nonuniform, does not allow samples to be taken exactly along the shape contours. The methods from the previous chapter do not produce the desired results with circular filters in terms of regular circular shape. As it was already mentioned, the best shape with the frequency sampling is produced when the samples are taken along shape isocontours. However, if the samples are taken on circles, 2-D matrix handling, as for example in Eq. (2.5.4), is impossible. Vector representations of the frequency samples and 2-D impulse response sequence should be used instead, as in (1.1.16). Intuitively, the problem lies into the more appropriate polar representation of sample coordinates in this case and the Cartesian coordinate representation of the DFT , NDFT, and the point locations of the filter impulse responses.

3.1 Circular shape FIR filter symmetry constraints and sampling parameters.

3.1.1 Symmetry constraints.

Before delving into 2-D circular FIR filter design with frequency samples, it is useful to review the symmetry properties of these filters and to determine what design savings can be realized.

In this chapter, as in Chapter 2, only zero phase 2-D filters will be considered, i.e., filters whose frequency response is a real function, $H(\omega_1 , \omega_2) = H^*(\omega_1 , \omega_2)$. In Sec. 2.2 it was pointed that for real valued impulse response filters this implies symmetry about the space domain origin, $h(n_1 , n_2) = h(-n_1 , -n_2)$. It is frequently of interest to explore new 2-D design techniques using FIR filters with an eightfold (octant) symmetry. Digital filters with circularly symmetric frequency response characteristic are filters with octal symmetry. The frequency response is symmetric with respect to the ω_1 axis, the ω_2 axis, and both diagonals $\omega_2 = \omega_1$, and the $\omega_2 = -\omega_1$:

$$H(\omega_1,\omega_2) = H(-\omega_1,\omega_2) = H(\omega_1,-\omega_2) = H(\omega_2,\omega_1) \tag{3.1.1}$$

From the symmetry properties of the Fourier transform, the eightfold symmetry in the Fourier domain implies eightfold symmetry in the space domain.

$$h(n_1, n_2) = h(-n_1,n_2) = h(n_1,-n_2) = h(n_2,n_1) \tag{3.1.2}$$

The only independent points of $h(n_1, n_2)$ are these in the half first quadrant, as depicted in Fig. 3.1(a).

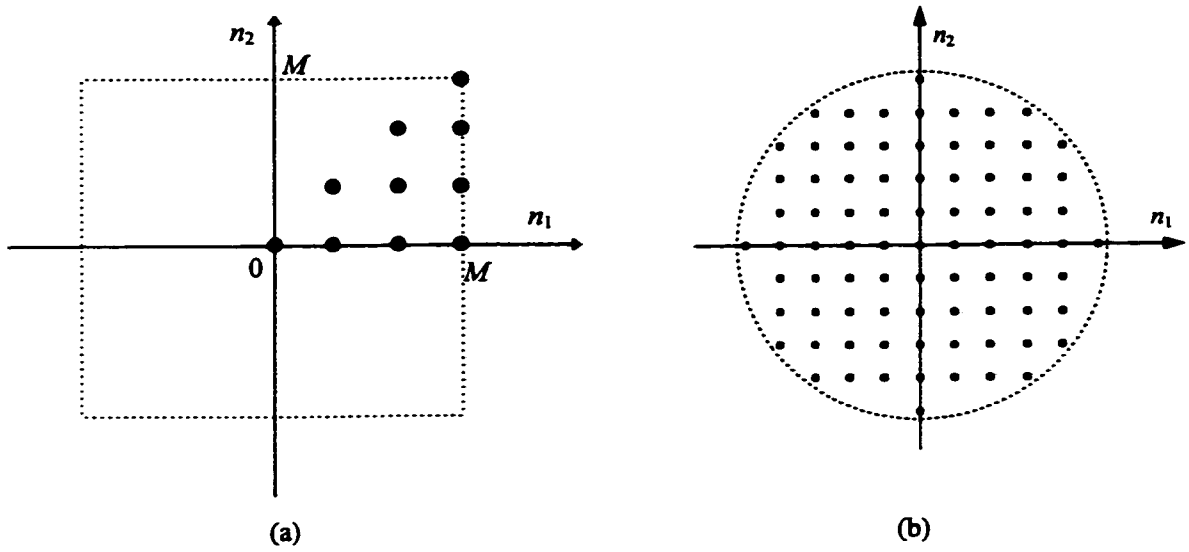


Figure 3.1 (a) Independent points of an 7×7-point impulse response with eightfold symmetry and square region of support; (b) ideal region of support of an 9×9-point impulse response of a circularly symmetric FIR filter.

In fact, the circular symmetry of the filter frequency response $H(\omega_1,\omega_2)$ implies circular symmetry of the impulse response $h(n_1, n_2)$ [10]. An impulse response has circular symmetry if it is a function of $n_1^2 + n_2^2$, see below Eqs. (3.1.12)-(3.1.15). The support region of such an impulse response of size 9×9 is shown in Fig. 3.1 (b). Such support region has advantages as smaller number of independent points than the square region of support and, therefore, the number of arithmetic operations per output sample in the implementation stage is smaller, too. However, the shape of support region of independent coefficients is different for the different filter sizes, making the matrix-vector manipulation difficult. That is why in the present work only filters with square impulse response region of support will be considered. As the experiments showed, the performance gain of circular impulse response support region is almost insignificant.

For a zero-phase filter with a finite-extent impulse response of size $N_1 \times N_2$, where

$N_1 = 2M_1+1$, $N_2 = 2M_2+1$, the transfer function is

$$H(z_1, z_2) = \sum_{n_1=-M_1}^{M_1} \sum_{n_2=-M_2}^{M_2} h(n_1, n_2) z^{-n_1} z^{-n_2} \tag{3.1.3}$$

and the corresponding frequency response is

$$H(\omega_1, \omega_2) = \sum_{n_1=-M_1}^{M_1} \sum_{n_2=-M_2}^{M_2} h(n_1, n_2) e^{-j\omega_1 n_1} e^{-j\omega_2 n_2} \tag{3.1.4}$$

Since for circular symmetry

$$N_1 = N_2 = N, \quad M_1 = M_2 = M, \quad N = 2M+1, \tag{3.1.4a}$$

and using the symmetries given by (3.1.2), this expression can be rewritten as

$$H(\omega_1, \omega_2) = \sum_{n_1=0}^{M} \sum_{n_2=0}^{M} a(n_1, n_2) \cos n_1\omega_1 \cos n_2\omega_2 \tag{3.1.5}$$

where

$$a(0, 0) = h(0, 0), \tag{3.1.5a}$$

$$a(n_1, 0) = 2h(n_1, 0), \tag{3.1.5b}$$

$$a(0, n_2) = 2h(0, n_2), \tag{3.1.5c}$$

and $$a(n_1, n_2) = 4h(n_1, n_2) \quad \text{for } n_1, n_2 \neq 0. \tag{3.1.5d}$$

Since $h(n_1, n_2) = h(n_2, n_1)$ implies $a(n_1, n_2) = a(n_2, n_1)$, hence,

$$H(\omega_1, \omega_2) = \sum_{n_1=0}^{M} \sum_{n_2=0}^{n_1} b(n_1, n_2)\left[\cos(n_1\omega_1)\cos(n_2\omega_2) + \cos(n_2\omega_1)\cos(n_1\omega_2)\right] \tag{3.1.6}$$

where

$$b(n_1, n_2) = a(n_1, n_2) \quad \text{for} \quad n_1 \neq n_2,$$

and $$b(n_1, n_2) = 0.5a(n_1, n_2) \quad \text{for} \quad n_1 = n_2. \tag{3.1.6a}$$

With further simplification of the notations, (3.1.5) can be expressed as

$$H(\omega_1, \omega_2) = \sum_{k=1}^{L} b(k)\phi_k(\omega_1, \omega_2) \tag{3.1.7}$$

where

$$\phi_k(\omega_1, \omega_2) = \cos(\omega_1 n_1)\cos(\omega_2 n_2) + \cos(\omega_1 n_2)\cos(\omega_2 n_1) \tag{3.1.8}$$

and

$$b(k) = b(n_1, n_2) \quad \text{for} \quad k = \frac{n_1(n_1+1)}{2} + n_2 + 1 \tag{3.1.9}$$

The number of independent filter coefficients, therefore, is

$$L = \frac{(M+1)(M+2)}{2} = \frac{(N+1)(N+3)}{8} = \sum_{i=1}^{(N+1)/2} i \tag{3.1.10}$$

Hence, L frequency sampling points should be sufficient to determine the coefficients $b(k)$, respectively $a(n_1, n_2)$ and $h(n_1, n_2)$. The size of the inverse problem is reduced by factor of $8N^2/(N+1)(N+3)$, which is between 4 and 6.5 for low and medium order filters.

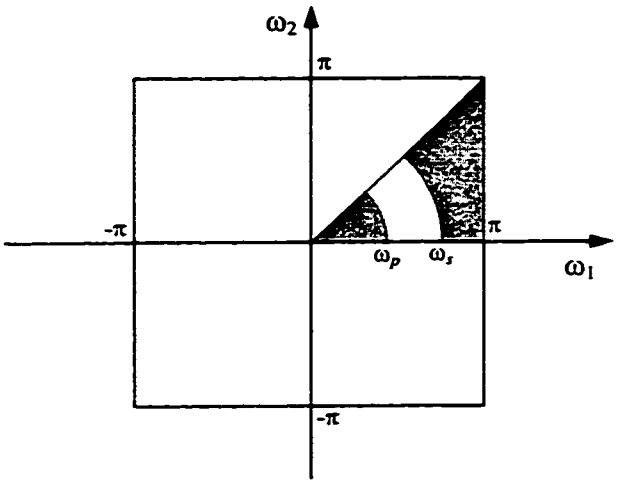


Figure 3.2 Approximation domain.

Since the frequency response (3.1.7) is symmetric with respect to both axes and both diagonals of the frequency plane, it will be sufficient to consider for sampling only the half of the first quadrant, $K = \{(\omega_1, \omega_2): 0 \le \omega_1 \le \pi; 0 \le \omega_2 \le \omega_1\}$, the shaded region in Fig. 3.2. In general, the samples can appear anywhere in this continuous domain. For now we shall consider that the set of sampling points is $\{\omega_1, ..., \omega_k, ..., \omega_L\}$, i. e., L sampling points, where $\omega_k = (\omega_{1k}, \omega_{2k})$. The basis functions $\phi_k(\omega) = \phi_k(\omega_1, \omega_2)$ can be defined as in (1.2.11) but in the case of eightfold symmetry it is more convenient to use the form (3.1.8) above (or as suggested by (3.1.5)). The characteristic vector, associated with a particular point ω_k is the vector

$$\Phi(\omega_k) = \left[\phi_1(\omega_k) \quad \phi_2(\omega_k) \quad \cdots \quad \phi_L(\omega_k)\right]^T. \tag{3.1.10.a}$$

The existence and uniqueness of solution to the approximation problem depends on the linear independence of these L characteristic vectors. In Sec. 2.1 it has been mentioned that L characteristic vectors randomly chosen from K, but for nonoverlapping ω_k , are almost always independent. Although this condition cannot be guaranteed, the practice shows that this is the case. Much more serious issue is the ill-conditioning, which will be considered later.

The frequency response specifications of zero phase circular FIR filters are given using similar tolerance scheme as for rectangular filters. Tolerance schemes for the four basic types circular filters are given in Fig. 3.3 and Fig. 3.4. In this text the radius of the passband boundary is denoted by ω_p, and the stopband radius by ω_s. To avoid ambiguity, the 1-D frequency edges are denoted by ω_p' and ω_s', respectively.

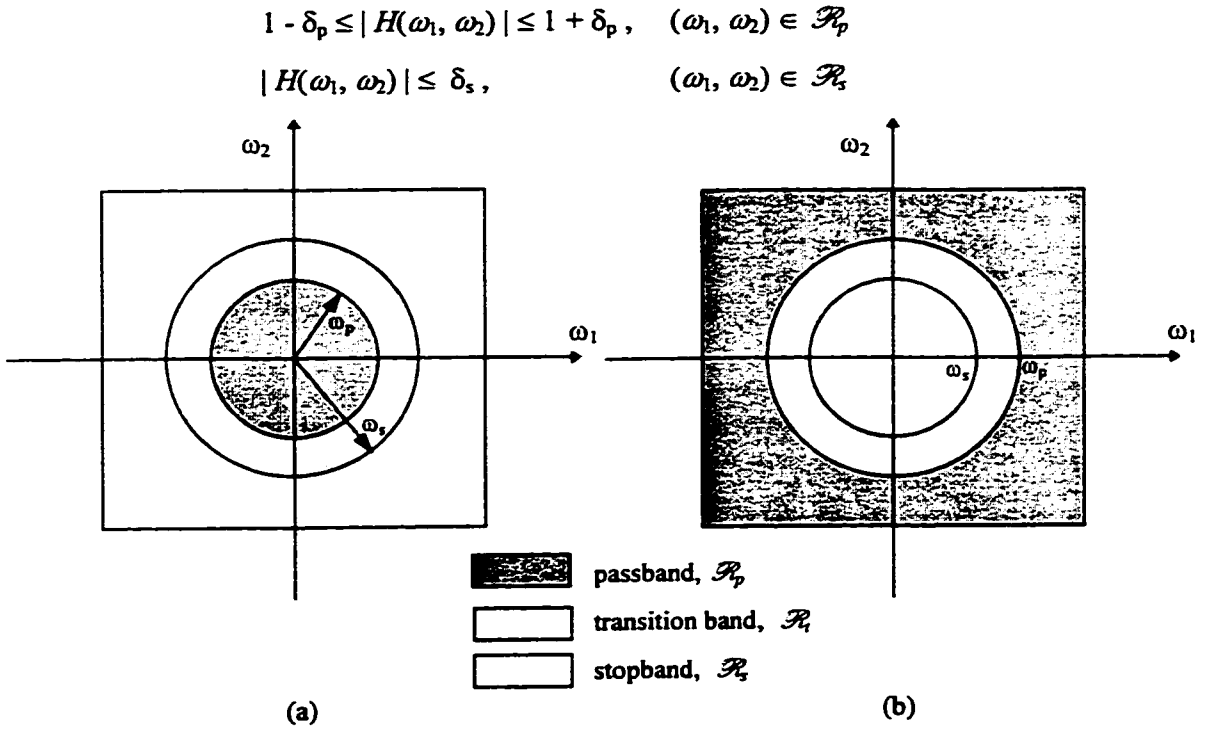


Figure 3.3 2-D circular shape filter specifications using a tolerance scheme for (a) lowpass and (b) highpass filters.

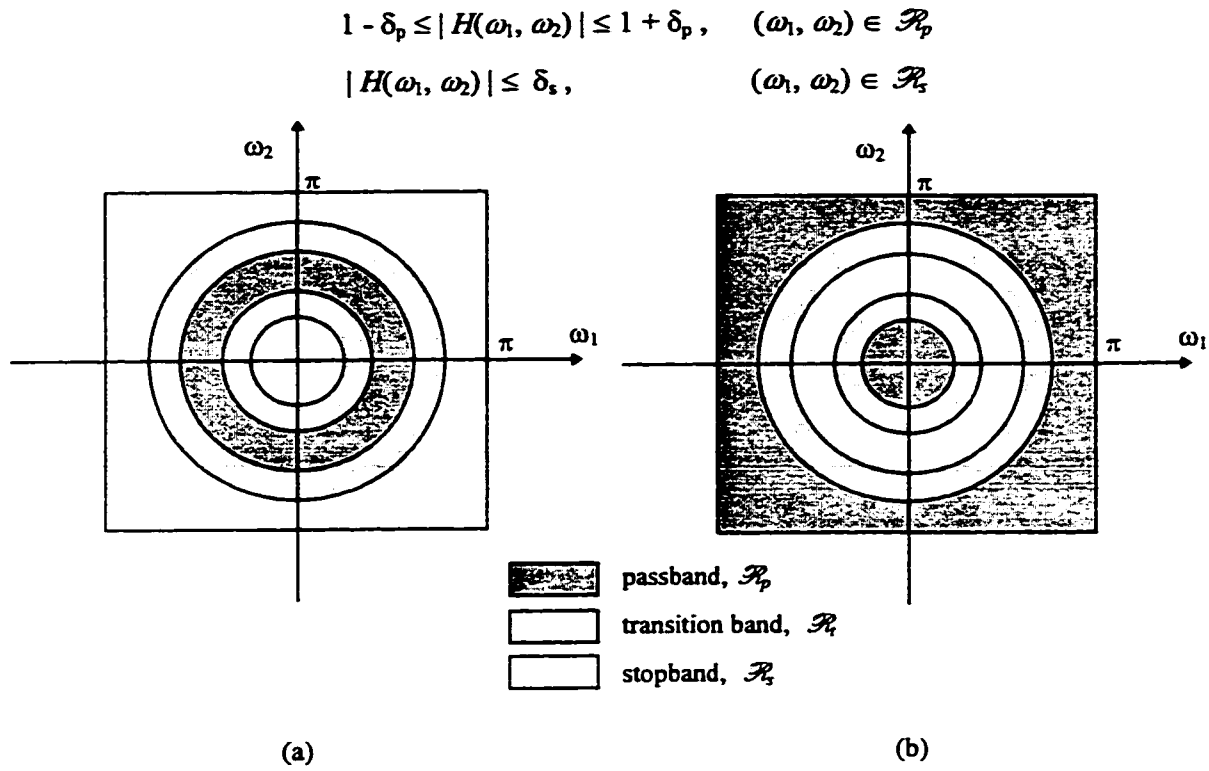


Figure 3.4 2-D circular filter specifications using a tolerance scheme for (a) bandpass and (b) bandstop filters.

It has been shown [10, 11] that the impulse response corresponding to the ideal lowpass circular frequency response

$$H_{lp}(\omega_1,\omega_2) = \begin{cases} 1, & \sqrt{\omega_1^2+\omega_2^2} \le \omega_C \\ 0, & \sqrt{\omega_1^2+\omega_2^2} > \omega_C \end{cases} \quad \text{and} \quad |\omega_1|,|\omega_1| \le \pi \tag{3.1.11}$$

is given by

$$h_{lp}(n_1,n_2) = \frac{\omega_C}{2\pi\sqrt{n_1^2+n_2^2}} J_1\left(\omega_C\sqrt{n_1^2+n_2^2}\right) \tag{3.1.12}$$

where J_1 is the first order Bessel function of the first kind and ω_C is the cutoff frequency. The impulse response of the ideal circular highpass filter is obtained by simply subtracting (3.1.12) from the impulse,

$$h_{hp}(n_1,n_2) = \delta(n_1,n_2) - h_{lp}(n_1,n_2) \tag{3.1.13}$$

The expressions for a bandpass and bandstop filters are similar:

$$h_{bp}(n_1,n_2)=\frac{\omega_{C2}}{2\pi\sqrt{n_1^2+n_2^2}}J_1\left(\omega_{C2}\sqrt{n_1^2+n_2^2}\right)-\frac{\omega_{C1}}{2\pi\sqrt{n_1^2+n_2^2}}J_1\left(\omega_{C1}\sqrt{n_1^2+n_2^2}\right) \quad (3.1.14)$$

$$h_{bs}(n_1,n_2)=\delta(n_1,n_2)-h_{bp}(n_1,n_2) \quad (3.1.15)$$

It is worth to note that the frequency response given by (3.1.5) (and (3.1.8)) can be expressed using Chebyshev polynomials. Introducing the notation $x=\cos\omega_1$, $y=\cos\omega_2$, $|x|\le 1, |y|\le 1$, (3.1.5) may be written as

$$H(\omega_1,\omega_2)=\sum_{n_1=0}^{M}\sum_{n_2=0}^{M}a(n_1,n_2)T_{n_1}(x)T_{n_2}(y) \quad (3.1.16)$$

where $T_k(\circ)$ is the k-th Chebyshev polynomial. Using the trigonometric identity

$$\cos(n+1)\omega+\cos(n-1)\omega=2\cos\omega\cos n\omega$$

and $T_0(x)=1$, $T_1(x)=x$, Eq. (3.1.16) can be written as

$$H(\omega_1,\omega_2)=\sum_{n_1=0}^{M}\sum_{n_2=0}^{M}c(n_1,n_2)x^{n_1}y^{n_2} \quad (3.1.17)$$

This expression of the frequency response is convenient to check the conditions of Theorem 1.5, Sec. 1.2.

3.1.2 Sampling Parameters

It was already mentioned that the practice has best results in terms of shape regularity are obtained when the frequency samples are lie on contours that mach the desired filter shape. Such is the case with the circularly symmetric filters, too. However, while the

shape in the region R_π : $\sqrt{\omega_1^2+\omega_2^2} \le \pi$ is well described by circular contours, this is not necessarily true for the corner region R_C : $\left\{\sqrt{\omega_1^2+\omega_2^2} > \pi\right\} \cap \left\{|\omega_1| \le \pi, |\omega_2| \le \pi\right\}$. Taking into account the symmetries considered above, in this text R_π and R_C will be limited to one half of the first quadrant of the frequency plane, as shown in the figure below.

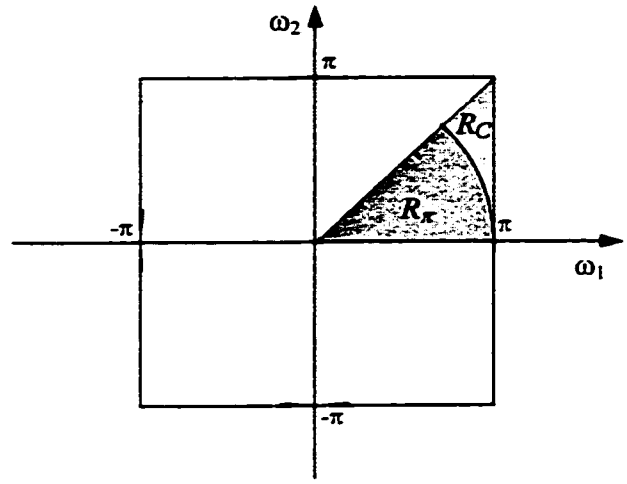


Figure 3.5 Definition of the regions R_π and R_C.

The sampling parameters to be determined are:

1) The number of contours and their distribution in R_π ;
2) the number of samples on each contour and the spacing between them;
3) the sample values;
4) the shape and the locations of contours in the region R_C
5) the sample density on each contour and sample values in this region

Now, the design problem of a 2-D circular shape, zero phase FIR filter with nonuniform frequency samples can be stated.

Problem Definition: Determine parameters 1) - 5) such that the design result is a circular filter with small passband and stopband deviations (nearly equiripple behavior, if

possible) and shape close to the ideal circular shape.

From bivariate interpolation point of view the problem can be defined as: Given a set of points $(\omega_{1k}, \omega_{2k})$, $k = 1, ..., L$, and a set of corresponding values at these points $\hat{H}_D(\omega_{1k}, \omega_{2k})$, find the coefficients $b(n_1, n_2)$, Eq. (1.3.6) of the interpolating function $H(\omega_1, \omega_2)$ such that the interpolation conditions

$$H(\omega_{1k}, \omega_{2k}) = \hat{H}_D(\omega_{1k}, \omega_{2k}), \quad \text{for} \quad k = 1, ..., L,$$

are satisfied.

Once the parameters 1) - 5) have been determined, the filter coefficients can be found by evaluating Eq. (3.1.6) at each sampling point and solving the resulting system of linear equations. This can be expressed in matrix form as

$$\hat{\mathbf{H}} = \mathbf{C}\mathbf{b} \tag{3.1.18}$$

where $\hat{\mathbf{H}}$ is an $(L\times 1)$ vector containing the sample values, $\mathbf{b}$ is an $(L\times 1)$ vector representing the sequence $b(k)$ given by (3.1.9). $\mathbf{C}$ is an $(L\times L)$ matrix having the form

$$\begin{bmatrix} 2 & \cos\omega_{11} + \cos\omega_{21} & 2\cos\omega_{11}\cos\omega_{21} & \cos 2\omega_{11} + \cos 2\omega_{21} & \cdots & 2\cos M_1\omega_{11}\cos M_2\omega_{21} \\ 2 & \cos\omega_{12} + \cos\omega_{22} & 2\cos\omega_{12}\cos\omega_{22} & \cos 2\omega_{12} + \cos 2\omega_{22} & \cdots & 2\cos M_1\omega_{12}\cos M_2\omega_{22} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 2 & \cos\omega_{1L} + \cos\omega_{2L} & \cos\omega_{1L}\cos\omega_{2L} & \cos 2\omega_{1L} + \cos 2\omega_{2L} & \cdots & 2\cos M_1\omega_{1L}\cos M_2\omega_{2L} \end{bmatrix} \tag{3.1.19}$$

Obviously, the solution depends on the nonsingularity and the conditioning of $\mathbf{C}$. The sequence $b(k)$ is found from

$$\mathbf{b} = \mathbf{C}^{-1}\hat{\mathbf{H}} \tag{3.1.20}$$

Then the impulse response $h(n_1, n_2)$ can be determined using (3.1.6a) and (3.1.5).

The investigation of the sampling parameters 1) - 5) will begin by employing some of the sampling techniques used in the previous chapter and analyzing the results. In all of the design examples the frequency samples of the desired filter frequency response are taken on circular contours in the region R_π. The shape of contours in the corner region R_C has been varied in a search of optimal shape.

3.2 Exponential distribution of the circular contours.

The design examples will begin as usual with the design of lowpass filters. The frequency samples in the region R_π lie on circular contours. The radii of these circles are the scaled and shifted samples of exponential functions similar to those used in Sec. 2.4 and 2.5. Namely, the radius of the k-th circle r_k, $k = 0, 1, \ldots, P+S$, is calculated as follows. For the passband the exponential function used is

$$f_p(x) = 1 - e^{-\alpha x}, \qquad x \in [0, 1], \tag{3.2.1}$$

and for the stopband the function used is

$$f_s(x) = e^{\alpha x} - 1, \qquad x \in [0, 1], \tag{3.2.2}$$

In both cases α is a positive constant. The functions f_s and f_p are sampled uniformly in the interval [0, 1]. The function $f_p(x)$ is sampled uniformly at P points in the interval [0, 1]. The scaled sample values are then used as radii in the interval [0, ω_p] where ω_p is the filter passband edge.

$$r_i = \frac{\omega_p}{f_{pm}} f_p(x_i), \tag{3.2.3}$$

where $x_i = i/(P-1)$ for $i = 0, 1, ..., P-1$, and $f_{pm} = f_p(1)$ (3.2.3a)

Similarly, the function $f_s(x)$ is sampled uniformly at S points and the radii in the stopband are obtained according to the following transformation

$$r_{(i+P)} = \frac{\pi - \omega_s}{f_{sm}} f_s(x_i) + \omega_s, \tag{3.2.4}$$

where $x_j = i/(S-1)$ for $i = 0, 1, ..., S-1$, and $f_{sm} = f_s(1)$. (3.2.4a)

The numbers P and S are chosen in proportion to the passband and stopband width, respectively, and such that $P + S$ is approximately $(N+1)/2$, and $N \times N$ is the filter impulse response size. The value of the frequency samples in the passband region is set to 1, and in the stopband region it is 0.

Example 3.1 A 2-D lowpass circular shape zero phase FIR filter. Edge specifications:

$|H_d(\omega_1, \omega_2)| = 1$ for $\omega_1^2 + \omega_2^2 \le (0.4\pi)^2$
$|H_d(\omega_1, \omega_2)| = 0$ for $(0.6\pi)^2 \le \omega_1^2 + \omega_2^2 \le \pi^2$

Filter size: 21×21 points.

Only 66 of the 441 filter coefficients are independent and 66 samples are taken in the half of the first quadrant of the frequency plane, Fig. 3.6(a). The samples in the region R_π lie on circular contours with radii according to (3.2.3) and (3.2.4) for the passband and stopband, respectively. A total of 55 samples are taken in this region. The rest 11 samples are taken in R_C and lie on straight lines with slope -1. The frequency response contour and perspective plots of the designed filter are shown in Fig. 3.6 (c) and (d), respectively. The peak error in the passband is $\delta_p = 0.0053$, and in the stopband $\delta_s = 0.0782$.

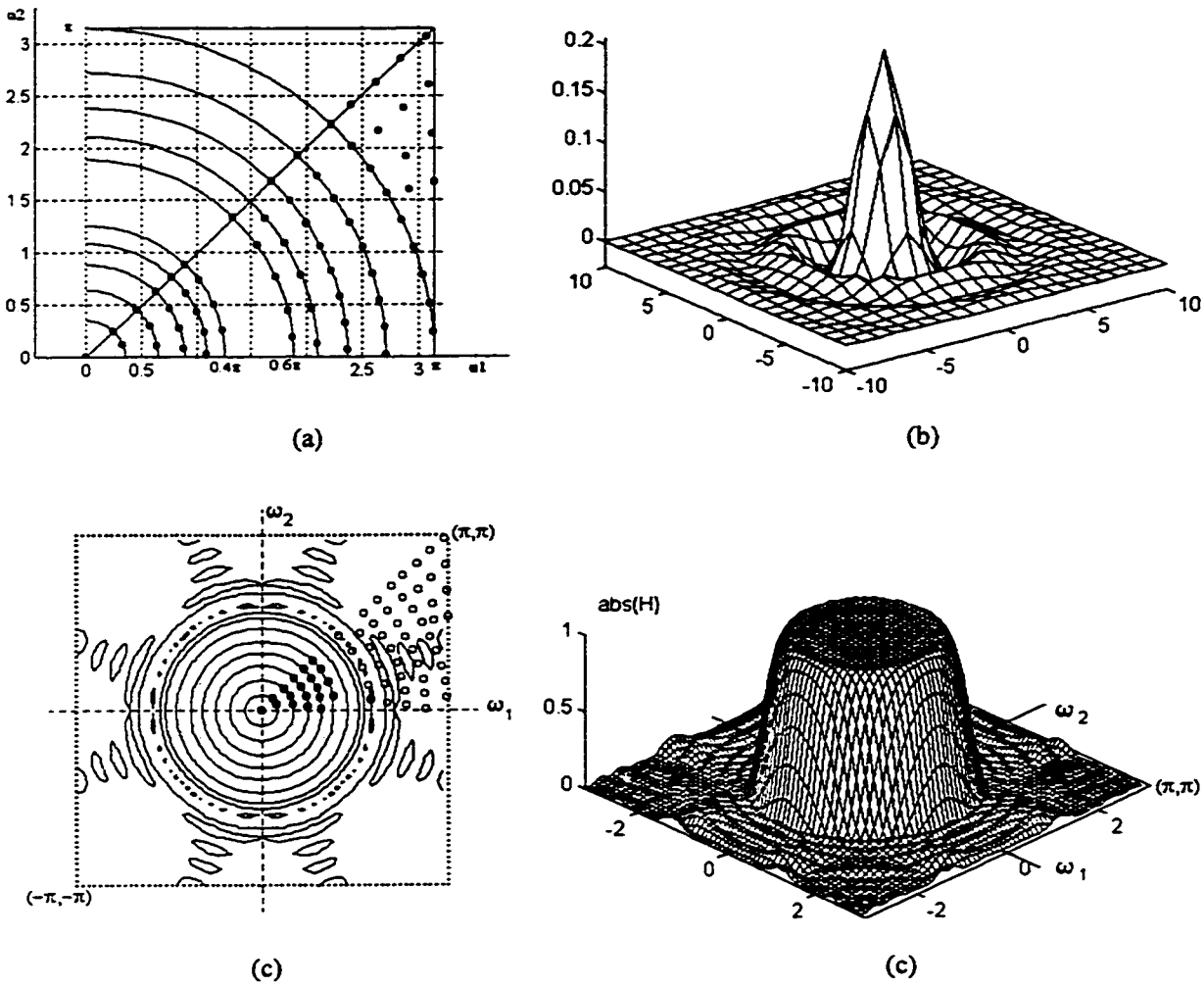


Figure 3.6 A 21×21-point circularly shaped filter designed with frequency samples taken on circular contours with exponential distribution in the passband and stopband, (a). The samples in R_C lie on straight lines. (b) Impulse response of the designed filter; (c) frequency response contour plot; (d) perspective plot.

The samples on each circle are equidistant but with unequal angular spacing between two different circles. The sampling density in the frequency domain agrees with some theoretical consideration which will be given later. In Fig. 3.6 (c) the samples having value of unity are denoted with •, and these with value zero with o.

As a practical matter of fact, the sample uniformity on each circle deteriorate the condition number of matrix (3.1.19). This effect has been alleviated by taking N_a more

samples than necessary to design an $N \times N$ filter and then removing randomly N_a samples in the region R_C . A drawback of this sampling technique is the approximately high ill-conditioning caused mostly by the sample arrangement in R_C.

3.3 Circular contours passing through the extrema of 1-D optimal filter.

In the region R_π : $\left\{\sqrt{\omega_1^2 + \omega_2^2} \le \pi\right\} \cap \{0 \le \omega_1 \le \pi,\ 0 \le \omega_2 \le \omega_1\}$ the samples lie on circular contours. Each circular contour is centered at the origin of the (ω_1 , ω_2) plane and has a radius equal to the corresponding 1-D extremal frequency obtained from the algorithm for optimal 1-D digital filter design. The sample on each contour have the same values and are equal to the corresponding 1-D samples used for the optimal filter design (the Parks-McClellan algorithm). In the corner region R_C the samples have been taken using different approaches as extrapolation of the 1-D extremal frequencies to $\pi\sqrt{2}$, samples lying on straight lines, scaling the 1-D frequency axis, and samples on hyperbolic contours, etc. Only some of these techniques deserve illustration here since they produce relatively good results in terms of small peak approximation error and shape regularity..

3.3.1 No samples in the corner region R_C

In some cases it is possible to design a circular shape FIR filter of good performance from samples taken only in the region R_π : $\left\{\sqrt{\omega_1^2 + \omega_2^2} \le \pi\right\} \cap \{0 \le \omega_1 \le \pi,\ 0 \le \omega_2 \le \omega_1\}$.

This technique can be employed for the design of filters of relatively low order (filter size up to 15×15 points) and relatively wide transition band (0.2π or wider). The radii of the circles on which the frequency samples are taken are equal to the extremal frequencies ω_k obtained with the Remez exchange algorithm, i.e., $r_k = \omega'_k$, $k = 1, ..., Q$, where Q is the

number of 1-D extremal frequencies. The technique will be illustrated with the following examples.

Example 3.2 A lowpass circularly shaped zero phase FIR filter with the following frequency edge specifications:

$$|H_d(\omega_1, \omega_2)| = 1 \quad \text{for} \quad \omega_1^2 + \omega_2^2 \le (0.35\pi)^2$$
$$|H_d(\omega_1, \omega_2)| = 0 \quad \text{for} \quad (0.65\pi)^2 \le \omega_1^2 + \omega_2^2 \le \pi^2$$

Filter size: 15×15 points.

A circularly shaped filter of this size has 36 independent coefficients out of 225. Therefore, to solve for those coefficients 36 frequency samples will be sufficient. We have, see Eqs. (3.1.4a) and (3.1.10), N=15, M= 8, L=36. A 1-D optimal design that produces Q=8 extremal frequencies is the design of a length N_{1D} = 13 equiripple filter, cf. Eq. (2.5.16). The Parks-McClellan algorithm with specifications N_{1D}=13, ω_p=0.35π, ω_s=0.65π, weighting = [1, 1], produces 8 frequencies which are

$\omega_k = [0 \quad 0.1607\pi \quad 0.2946\pi \quad 0.35\pi \quad 0.65\pi \quad 0.7125\pi \quad 0.8464\pi \quad \pi]^T = r_k$.

The samples in the 2-D frequency plane are taken on circles with in the above radii and centered at the origin. Only the first octant is covered, as shown in Fig. 3.7 (a).

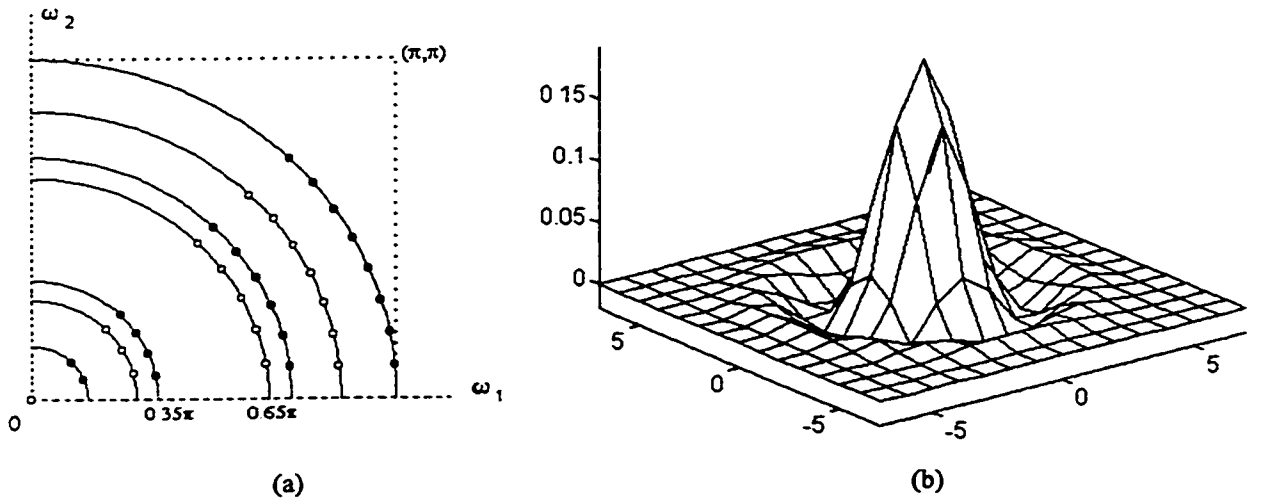


Figure 3.7 A 15×15-point zero phase FIR filter designed with samples taken only in R_π. (a) Sampling locations; (b) impulse response; (continued on the next page)

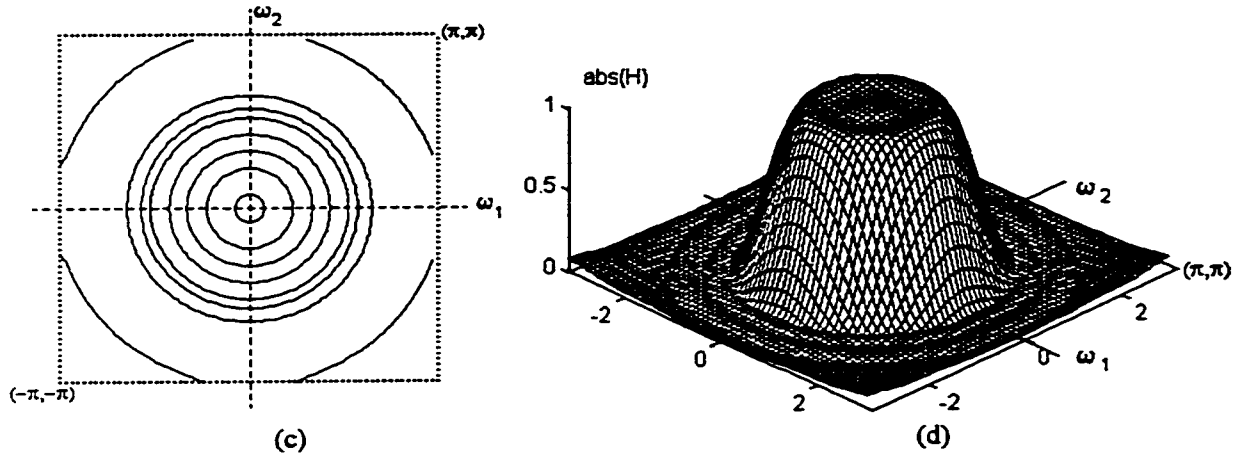


Figure 3.7 (continued) A 15×15-point zero phase FIR filter designed with samples taken only in R_x. (c) frequency response contour plot; (d) frequency response magnitude perspective plot.

The number of samples on each circle is k, where k is the circle number, k=1, ... , Q . This is a good choice since in this way the requirement of Eq. (3.1.10) is satisfied. The positive and negative samples are denoted by o and •, respectively, see (1.2.14). The 1-D ripple obtained with this design is δ_{1D} = 0.0173, both for the passband and stopband, naturally. The contour plot and perspective plot of the frequency response are shown in Fig. 3.7 (c) and (d), respectively. The peak error of the designed filter is δ_p=0.0174 in the passband, and δ_s= 0.0787 in the stopband. The passband peak error is almost the same as the ripple of the 1-D optimal filter. The maximum error in the stopband is at the corner (π, π). Several design results using the same technique are presented in Table 3.1.

Filter size	N_{1D}	# of Samples	ω_p	ω_s	δ_{1D}	δ_p	δ_s
11×11	9	21	0.3π	0.7π	0.0257	0.0258	0.0433
13×13	11	28	0.25π	0.45π	0.0642	0.0642	0.0726
15×15	13	36	0.35π	0.65π	0.0173	0.0174	0.0787
15×15	13	36	0.38π	0.62π	0.0332	0.0332	0.0768
19×19	17	55	0.3π	0.6π	0.0037	0.0037	0.2275
23×23	19	78	0.38π	0.62π	0.0056	0.0057	6.4339

Table 3.1 Performance of the technique with no samples in R_C.
N_{1D} – 1-D prototype length; ω_p, ω_s – radii of the 2-D passband, stopband regions;
δ_{1D} – 1-D ripple. δ_p, δ_s – 2-D passband and stopband deviations.

As it can be seen from the table, the maximum error in the stopband increases rapidly for filter sizes larger than 15×15 points. For the last item in the table, a 23×23-point filter this error is unacceptably large, $\delta_s = 6.4339$. This peak error occurs in the corner region R_C where no samples were taken. The next design example shows this unwanted effect of increasing the peak stopband error in the R_C region with the increase of sample density.

Example 3.3 A lowpass circularly shaped zero phase FIR filter with the following frequency edge specifications:

$$|H_d(\omega_1, \omega_2)| = 1 \quad \text{for} \quad \omega_1^2 + \omega_2^2 \le (0.3\pi)^2$$
$$|H_d(\omega_1, \omega_2)| = 0 \quad \text{for} \quad (0.6\pi)^2 \le \omega_1^2 + \omega_2^2 \le \pi^2$$

Filter size: 19×19 points.

The 1-D optimal prototype whose 10 extremal frequencies and corresponding amplitude values were used has length $N_{1D} = 17$ points. Its maximum ripple is $\delta_{1D} = 0.0037$. The 2-D sample arrangement is the same as in the previous example and it is shown in Fig. 3.7 (a). A total of 55 samples were taken, necessary to solve for the 55 independent filter points.

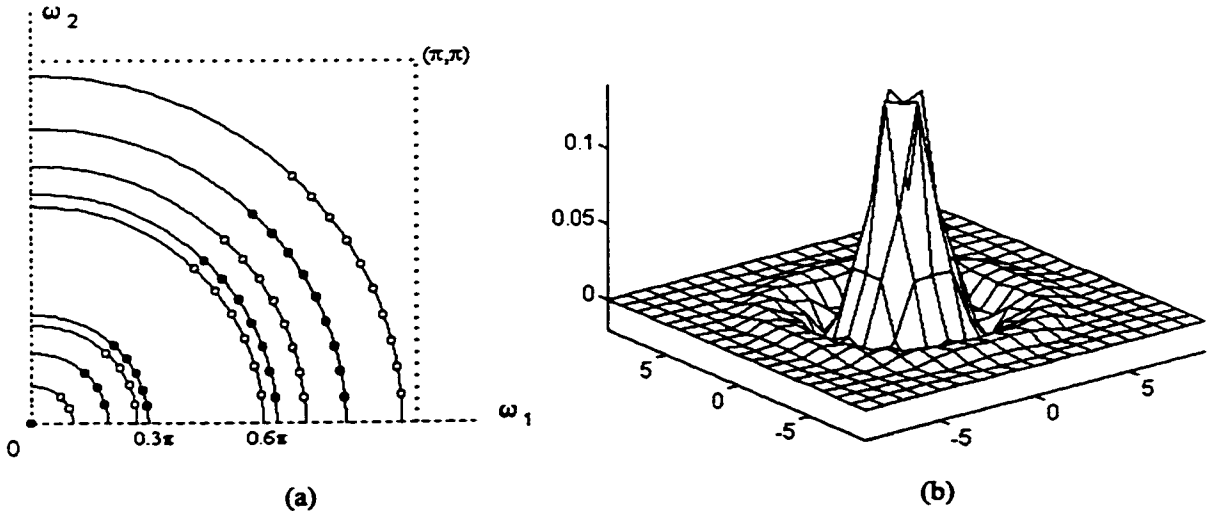
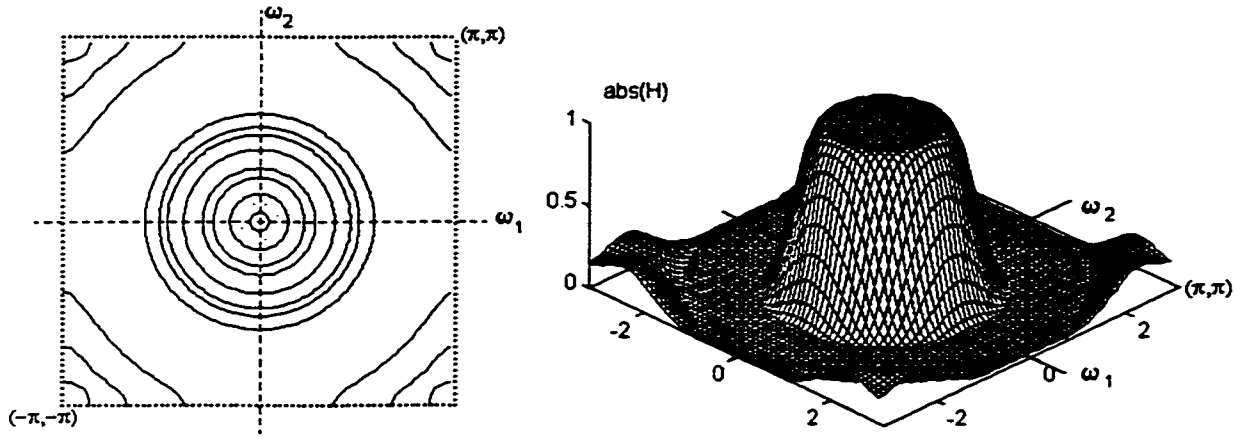


Figure 3.8 A 19×19-point zero phase FIR filter designed with 55 samples taken only in R_π. 1-D filter length $N_{1D} = 17$. (a) Sampling locations; (b) impulse response, 19×19; On the next page: (c) frequency response contour plot; (d) frequency response magnitude perspective plot.



(c) (d)

Figure 3.8 (continued)

Now the sampling density is increased, we have 19 samples more in the same area R_π. This increases considerably the error in the corner region, as it can be seen on the perspective plot, Fig. 3.7 (d). The maximum deviation in the stopband now is $\delta_s = 0.2275$. However, the passband maintains almost equiripple behavior and the low ripple of the 1-D prototype, namely $\delta_p = \delta_{1D} = 0.0037$.

It is interesting fact that for each filter size increase, e.g. from 15×15 to 17×17, the condition number of matrix **C**, Eq. (3.1.19), increases approximately by factor of 10. In other words, the ill-conditioning rapidly increases.

3.3.2 Scaling the 1-D frequency axis.

This technique uses a very simple trick in order to cover with samples the corner region R_C: $\left\{\sqrt{\omega_1^2+\omega_2^2} > \pi\right\} \cap \left\{0 \le \omega_1 \le \pi,\ 0 \le \omega_2 \le \omega_1\right\}$. The 1-D extremal frequencies are obtained by supplying to the Remez exchange algorithm frequency edge specifications of the desired filter which are compressed by factor of $1/\sqrt{2}$. The obtained extremal

frequencies are then expanded by factor of $\sqrt{2}$. In this way the interval [0, π] of the 1-D frequency axis is mapped into the diagonal [(0,0), (π,π)] of the (ω_1, ω_2) plane. This can be viewed as scaling the whole (ω_1, ω_2) plane. Now the circular contours passing through the extremal frequencies reach the point (π, π). Of course, this point is not always covered since the point at π in 1-D is not always a member of the extremal point set. The following two examples illustrate this sampling technique.

Example 3.4 A lowpass circularly shaped zero phase FIR filter with

unity gain in the passband region $\mathcal{R}_p$: $\omega_1^2 + \omega_2^2 \le (0.42\pi)^2$,

zero gain in the stopband region $\mathcal{R}_s$: $(0.58\pi)^2 \le \omega_1^2 + \omega_2^2 \le \pi^2$

Impulse response region of support: 21×21 points.

For this design `remez` was called with the following specifications: $\omega_p = (0.42\pi)/\sqrt{2}$, $\omega_s = 0.58\pi/\sqrt{2}$, filter order = 20. With these specifications the resulting 1-D ripple is δ_{1D}=0.024. The frequency samples in the (ω_1, ω_2) plane were taken as before: on each circle the samples are at equal angles with respect to the origin and with equal values equal to the 1-D amplitude of the corresponding extremal frequency. A total of 66 samples were taken in the domain of approximation in order to solve for the 66 independent impulse response points. These samples are shown in Fig. 3.9 (a). As before, the positive and negative samples are denoted by o and •, respectively. The mesh plot of the resulting impulse response is presented in Fig. 3.9 (b). The frequency response magnitude contour plot and perspective plot are shown in Fig. 3.9 (c) and (d), respectively. The maximum deviation in the passband of the 2-D filter is δ_p = 0.0256, and the stopband deviation is δ_s = 0.0794. Obviously, the passband nearly maintains the low ripple of the 1-D prototype filter. The problem is in the stopband where the deviation is increased. The maximum error in the stopband is at the point (π, 0.7812π).

As a matter of fact, with this sampling technique there is no significant improvement of the conditioning of matrix **C**. However, now it is possible to design filters with narrower transition bands and with better deviation performance.

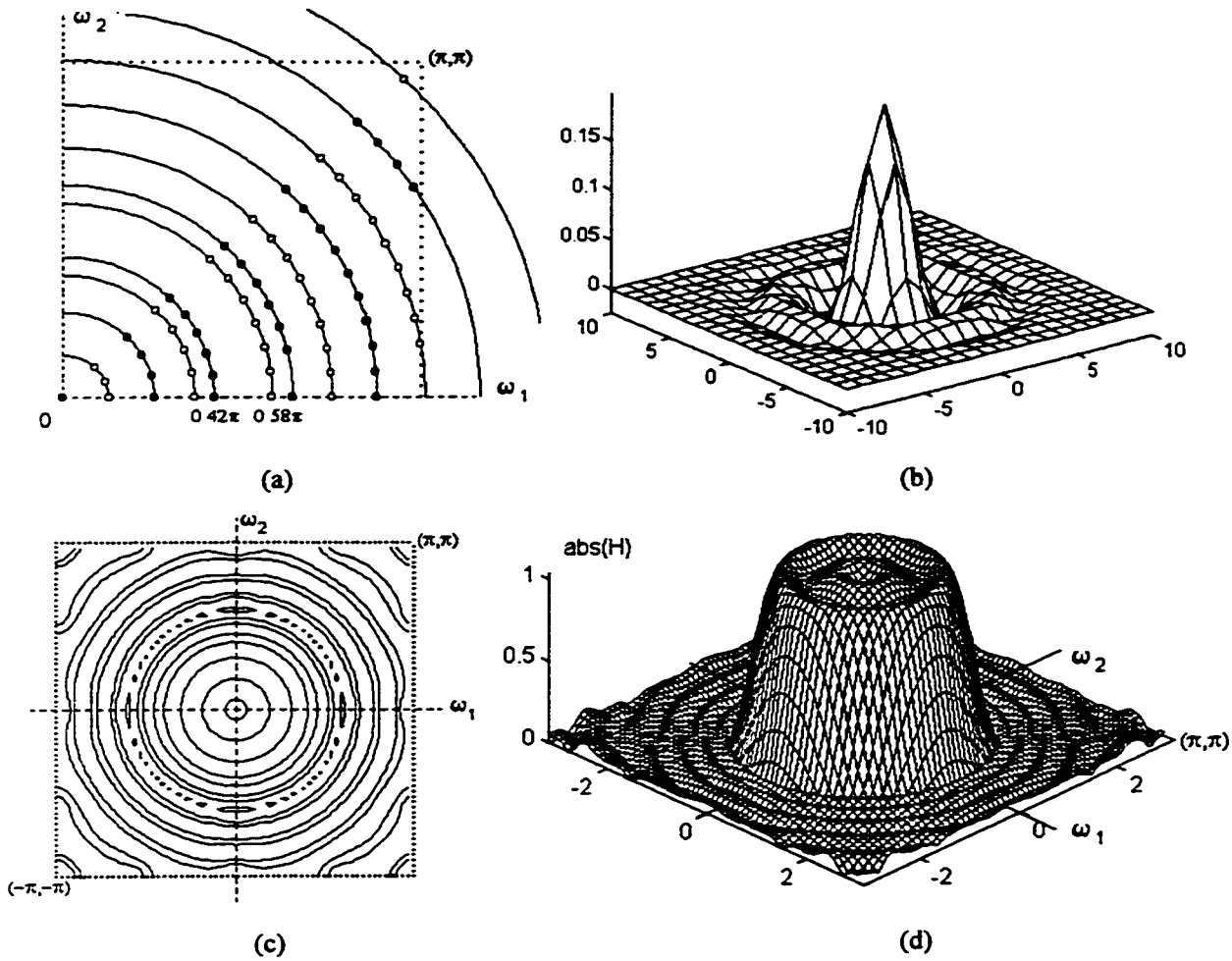


Figure 3.9 Design Example 3.4. A 21×21-point filter designed by scaling the 1-D extremal frequencies. (a) Frequency samples; (b) impulse response; (c) frequency response magnitude contour plot; (d) frequency response magnitude perspective plot.

Naturally, this technique can be also used for the design of highpass, bandpass, and bandstop 2-D circularly shaped, zero phase digital FIR filters. This is done by simply using the corresponding extremal points for the design of 1-D equiripple highpass, bandpass, and bandstop filters, respectively. Alternatively, a highpass or bandstop filter can be obtained from a lowpass or bandpass filter using the relations (3.1.13) and (3.1.15), respectively. The next example demonstrates the design of a bandpass filter.

Example 3.5 A bandpass circular shaped FIR filter with specifications:

unity gain in the passband region $\mathcal{R}_p$: $(0.3\pi)^2 \le \omega_1^2 + \omega_2^2 \le (0.6\pi)^2$,

zero gain in the stopband region $\mathcal{R}_s$: $0 \le \omega_1^2 + \omega_2^2 \le (0.1\pi)^2 \cup (0.8\pi)^2 \le \omega_1^2 + \omega_2^2 \le \pi^2$

Impulse response region of support: 21×21 points.

The 1-D extremal frequencies were obtained from the modified Matlab program `remez.m` supplied with the above frequency edge specifications translated to 1-D and filter order 20. For a 21×21-point filter 66 frequency samples are necessary to be taken. The locations of these samples are shown in Fig. 3.10 (a). The impulse response and frequency response magnitude contour and perspective plots of the designed filter are shown in Fig. 3.10 (b), (c), and (d), respectively.

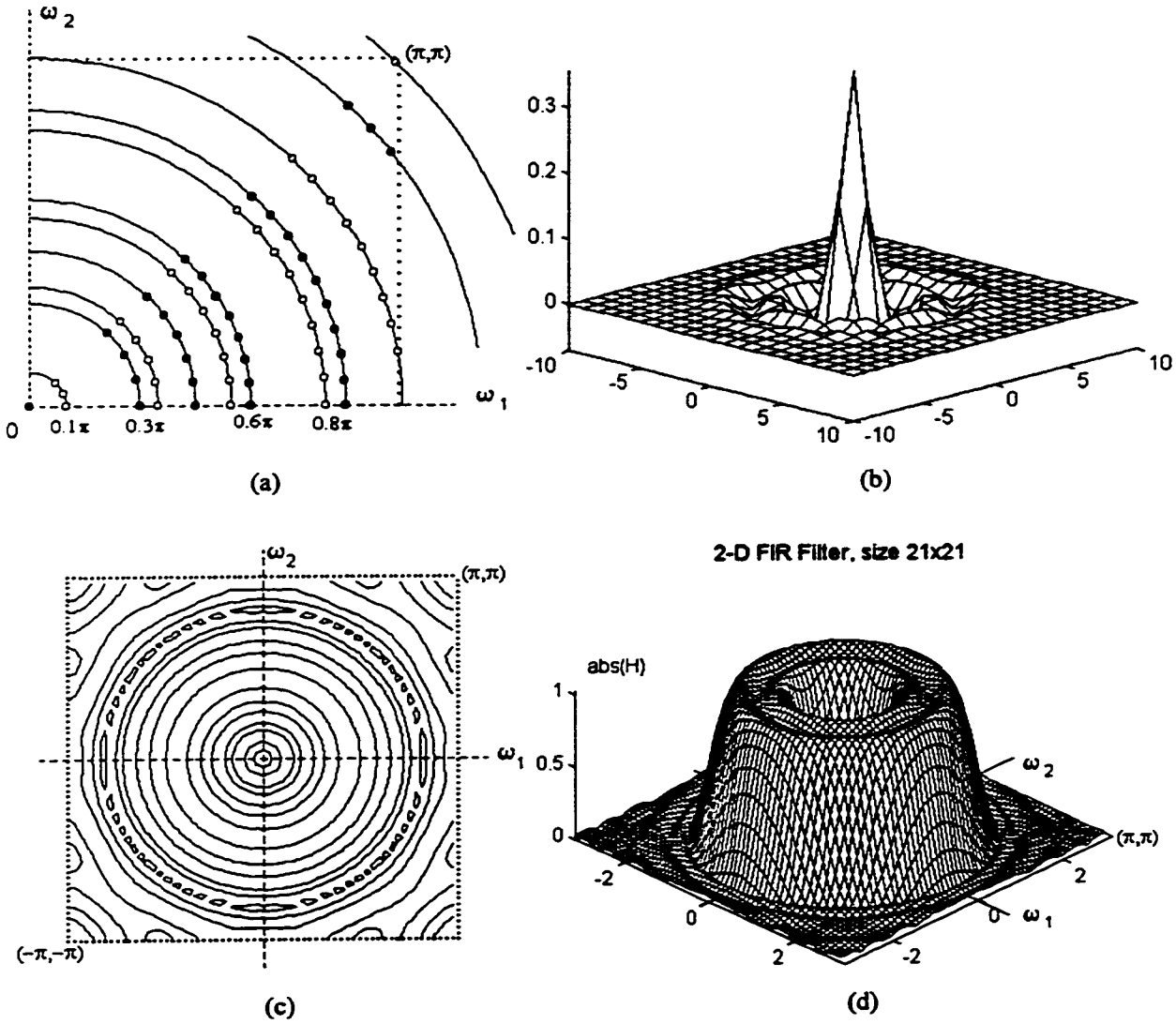


Figure 3.10 Design Example 3.5. A 21×21-point bandpass filter designed by scaling the 1-D extremal frequencies. (a) Frequency samples; (b) impulse response; (c) frequency response magnitude contour plot; (d) frequency response magnitude perspective plot.

The resulting 2-D zero phase bandpass FIR filter frequency response magnitude exhibits deviations $\delta_{s1} = 0.0296$, $\delta_p = 0.0299$, and $\delta_{s2} = 0.0552$ in the first stopband region, in the passband, and in the second stopband, respectively. As before, the maximum error is in the corner region R_C, which shows that the sampling locations and values in the corner are not the optimal. In other words, the guess for the frequency response behavior in this region is not exact. Further improvement of the sampling arrangement, producing filters with better performance in terms of deviations, is proposed in the next subsection.

3.3.3 Hyperbolic contours in the region R_C.

This technique has been motivated by the fact that the frequency response contours of a filter designed with samples taken only in R_π , Fig. 3.10 (a), "naturally" bend in the shape of hyperbolas in the corner region R_C, Fig. 3.6 (b) and (c). This is the case even if a small number of samples is taken in this region.

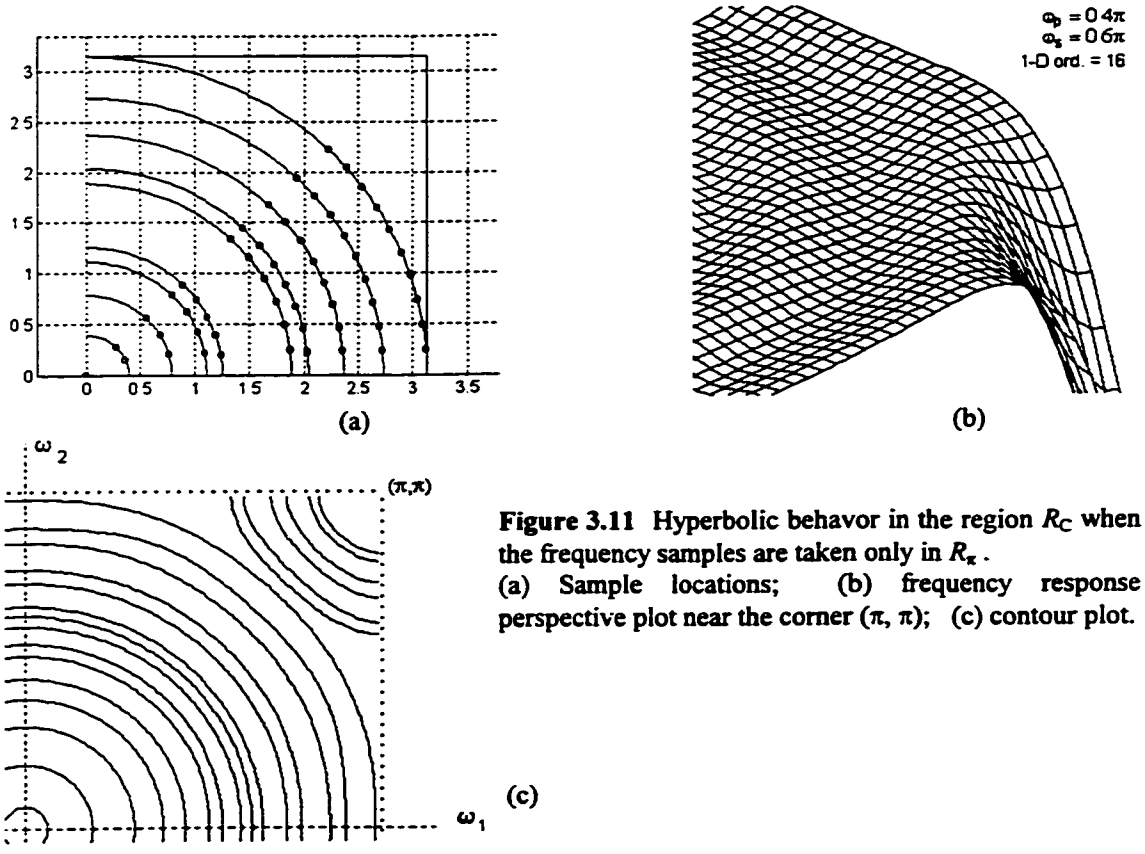


Figure 3.11 Hyperbolic behavor in the region R_C when the frequency samples are taken only in R_π .
(a) Sample locations; (b) frequency response perspective plot near the corner (π, π); (c) contour plot.

The hyperbolic contours on which the samples will be taken have the same orientation as the frequency response contours in Fig. 3.6 (c), i.e., these are hyperbolas rotated by $\pi/4$. The density of the samples in the corner region R_C is chosen to be approximately the same as in R_π. The frequency samples in R_π are taken as described in Sec. 3.3.1, on circular contours with radii equal to the 1-D extremal frequencies. The effectiveness of this sampling technique will be demonstrated with several design examples.

Example 3.6 A 2-D lowpass circularly shaped zero-phase FIR filter with impulse response of size 23×23 points, and with the following frequency edges

passband region $\mathcal{R}_p$: $0 \le \omega_1^2 + \omega_2^2 \le (0.38\pi)^2$,

stopband region $\mathcal{R}_s$: $(0.62\pi)^2 \le \omega_1^2 + \omega_2^2 \le \pi^2$

The 1-D extremal frequencies were obtained from the modified Matlab program `remez.m` supplied with the above frequency edge specifications translated to 1-D, i.e., $\omega_p'=0.38\pi$, $\omega_s'=0.62\pi$, and 1-D filter order = 20. A 23×23-point eightfold symmetric filter has 78 independent impulse response points. In order to find them, 78 samples of the desired frequency response amplitude were taken in the first octant of the (ω_1, ω_2) plane, Fig. 3.12 (a). As in the previous subsections, the error-positive samples are denoted by • , and the positive points by o. The samples on each contour have the value of the corresponding 1-D extremal sample. The pattern is continued in the corner region R_C, i.e., the samples alternate in sign on successive contours and have magnitude equal to the 1-D ripple. The distance between the hyperbolic contours is taken to be equal to the average distance between the 1-D extremal frequencies in the last band, the stopband in this example. The perspective plot of the designed filter frequency response magnitude is shown in Fig. 3.12 (d). The equiripple behavior and the low 1-D ripple are inherited in the passband region of the designed filter; $\delta_{1D} = 0.0056492$, $\delta_p = 0.0056607$. The error in the stopband is about 3 times larger, $\delta_s = 0.017425$. This can be somehow corrected by introducing a weighting function in the Remez exchange algorithm and thus making the error minimization more important in the 1-D stopband. In turn, the error in the 2-D filter stopband region will also be decreased.

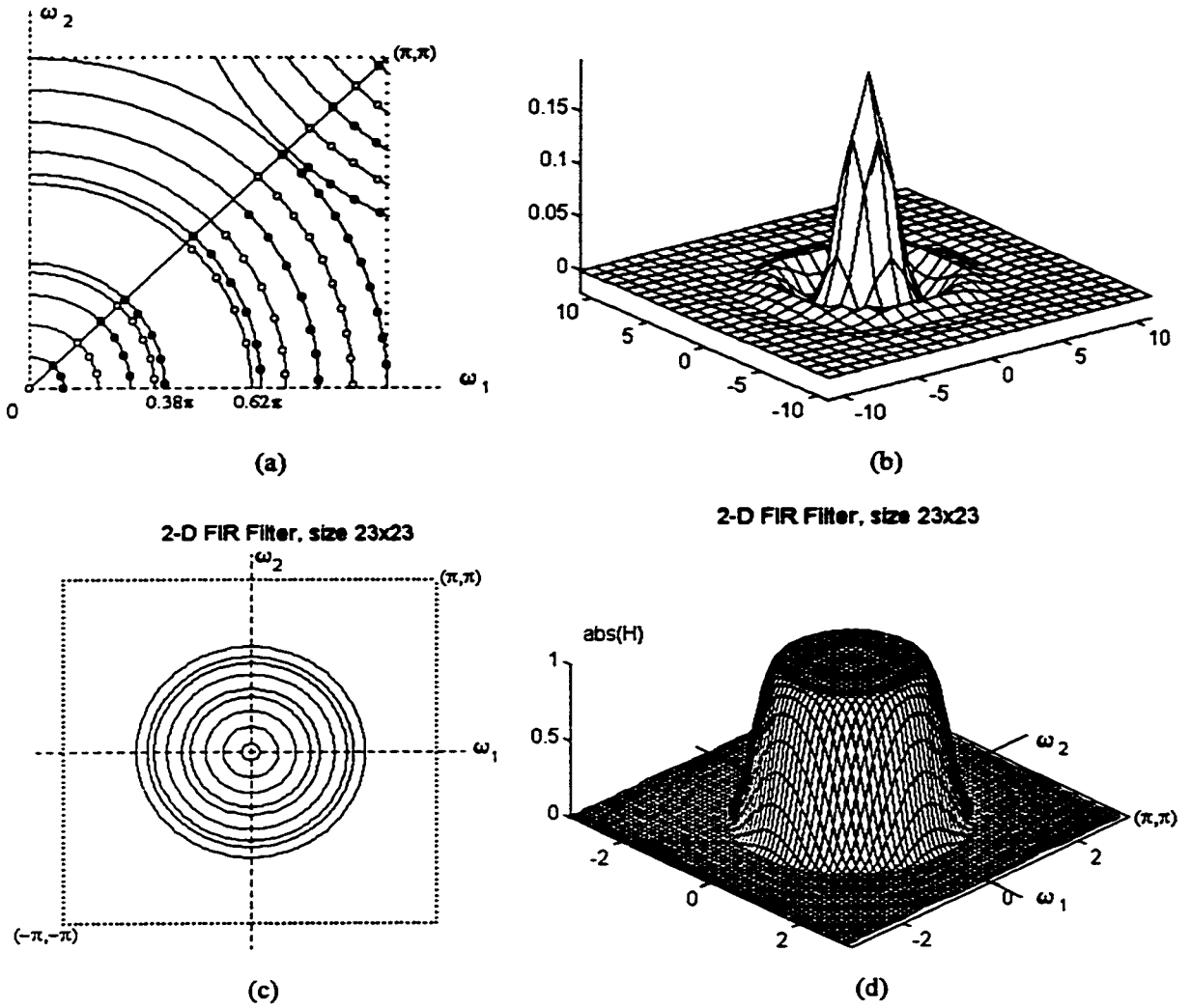


Figure 3.12 Example 3.6: A 23×23-point circular filter designed from frequency samples in the corner region taken on hyperbolic contours. The extremal frequencies and the corresponding amplitude values of 1-D length-21 optimal filter were used. (a) Sample locations; (b) impulse response of the designed filter; (c) and (d) frequency response magnitude contour and perspective plots, respectively.

Using the modified `remez.m` with a weight vector w = [0.5, 1], the error in the 2-D design was somehow balanced: δ_p = 0.007623 in the passband region and δ_s = 0.007445 in the stopband region, respectively. The method allows other types of filters, as passband, stopband, and highpass filters, to be designed. This is done by simply using the extremal points of the corresponding type 1-D optimal filter (again using the modified `remez.m`).

Example 3.7 A 2-D circularly shaped bandstop zero phase FIR filter with impulse response of size 25×25 points, and with the following frequency edges

passband region $\mathcal{R}_p$: $0 \le \omega_1^2 + \omega_2^2 \le (0.2\pi)^2$ and $(0.8\pi)^2 \le \omega_1^2 + \omega_2^2 \le \pi^2$

stopband region $\mathcal{R}_s$: $(0.4\pi)^2 \le \omega_1^2 + \omega_2^2 \le (0.6\pi)^2$.

The 1-D extremal frequencies and corresponding amplitudes were obtained with filter order of 22 and the above frequency specifications translated to 1-D, i. e., $\omega'_{p1} = 0.2\pi$, $\omega'_{s1} = 0.4\pi$, $\omega'_{s2} = 0.6\pi$, and $\omega'_{p2} = 0.8\pi$. The locations of 91 frequency samples, necessary to find the 91 independent filter points, are shown in Fig. 3.13 (a).

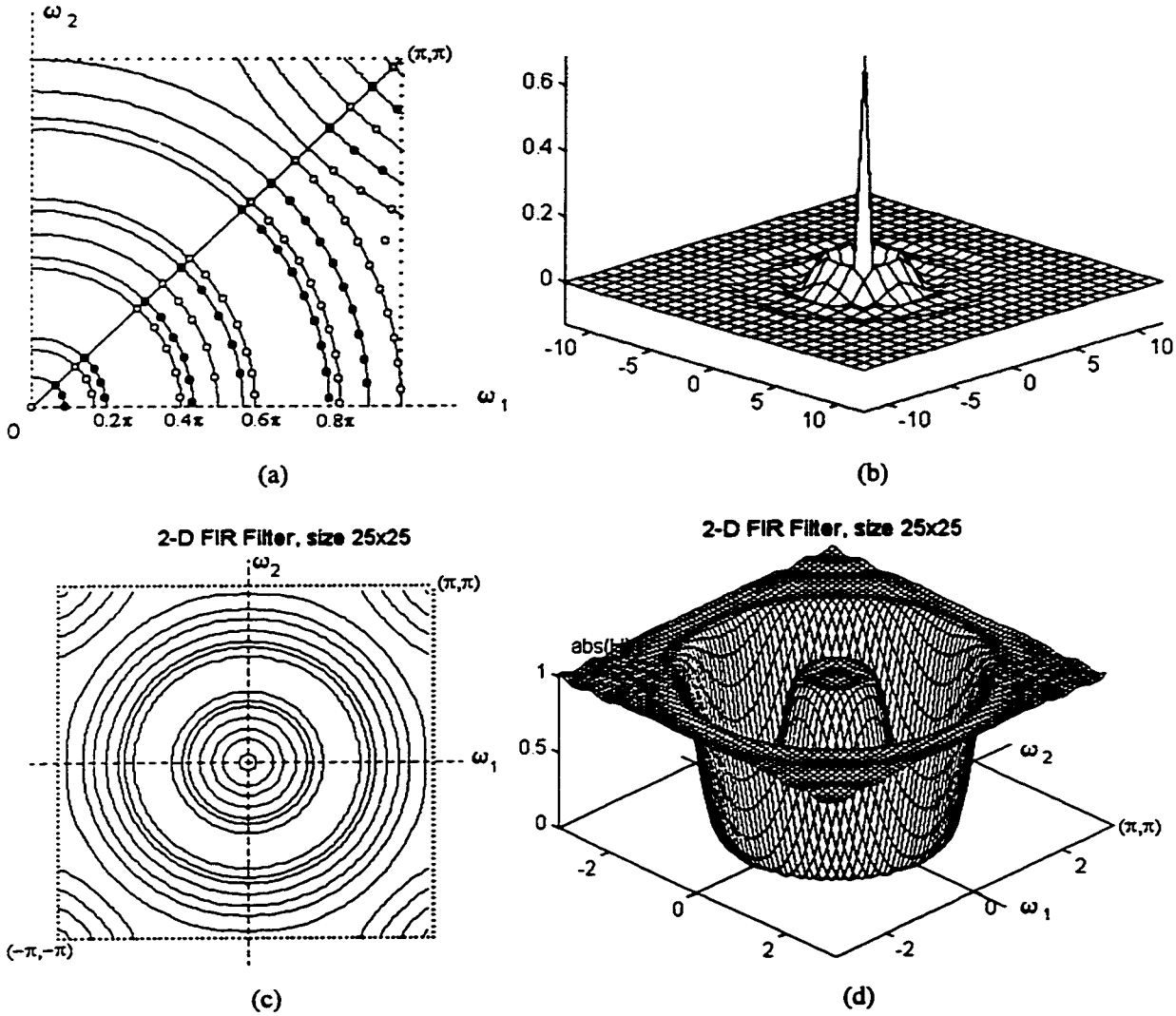


Figure 3.13 Example 3.7: A 25×25-point bandstop FIR filter designed from frequency samples in the corner region taken on hyperbolic contours. (a) Sample locations; (b) impulse response of the designed filter; (c) and (d) frequency response magnitude contour and perspective plots, respectively.

The 1-D ripple of an order-22 optimal filter without weighting is $\delta_{1D} = 0.011513$. The

resulting deviations of the frequency response magnitude of the designed 2-D filter are $\delta_{p1} = 0.011557$, $\delta_s = 0.011611$, and $\delta_{p2} = 0.028292$ in the first passband, the stopband, and in the second passband, respectively. This error can be compared to the error that return bandstop filters designed with the window method and frequency transformation method. The results are given in Table 3.2. The both filters have the same region of support, 25×25, and the input frequency edges were the same as the filter designed in Example 3.7. For this example and in the following examples the shape parameter in the rotated Kaiser window method was chosen such that the minimum possible passband and stopband deviations were obtained for the given filter size. As for the transformation method, the bandstop filter was designed using the original (McClellan) transformation

$$\cos\omega = -\tfrac{1}{2} + \tfrac{1}{2}\cos\omega_1 + \tfrac{1}{2}\cos\omega_2 + \tfrac{1}{2}\cos\omega_1\cos\omega_2. \qquad (3.3.1)$$

Since the 2-D filter should be of size 25×25 and the transformation sequence is 3×3, the 1-D optimal prototype has length $N_{1D} = 25$. As it can be seen from the table, the performance of the bandstop filter designed with the proposed method for nonuniform frequency sampling is better in the first passband compared to the first passband of the filter designed with the window method. The roles are interchanged in the second passband. The transformation method gives much better deviation performance and this is natural since a higher order 1-D prototype was used with smaller ripple, δ_{1D}=0.004045. However, the transformation method filter failed with the second passband edge and that is why the deviation in the second passband is so large, δ_{p2}= 0.242. The true second passband edge is at radius $\omega = 0.9\pi$ instead at $\omega = 0.8\pi$. And if we want the comparison to be more fair, the transformation filter should be designed with the 1-D prototype used for the frequency sampling method. In this case the deviations of both filters are very similar, yet the transformation filter fails again with the second passband edge.

The shape performance of the 2-D FIR filters designed with the proposed nonuniform frequency sampling method should not be underestimated either. Fig. 3.14 shows the contour plots of three filters designed with the same specifications:

passband region $\mathcal{R}_p$: $0 \le \omega_1^2 + \omega_2^2 \le (0.75\pi)^2$
stopband region $\mathcal{R}_s$: $(0.95\pi)^2 \le \omega_1^2 + \omega_2^2 \le \pi^2$

impulse response region of support size: 23×23 points.

The first frequency response contour plot, Fig. 3.14 (a), is the filter designed using the rotated Kaiser window method. The second plot, (b), represents the filter designed with the frequency transformation method, and the third one, (c), is the contour plot of the filter designed with the method under consideration. The dotted circle is an ideal circle drawn with radius 0.85π, the cutoff, and it is not a part of the contours. As expected, the window method gives the most regular circular shape. The shape obtained with the nonuniform frequency sampling method is not better but it is pretty much similar. By the way, the shape regularity can also be seen in the contour plots of all the previous examples in this section.

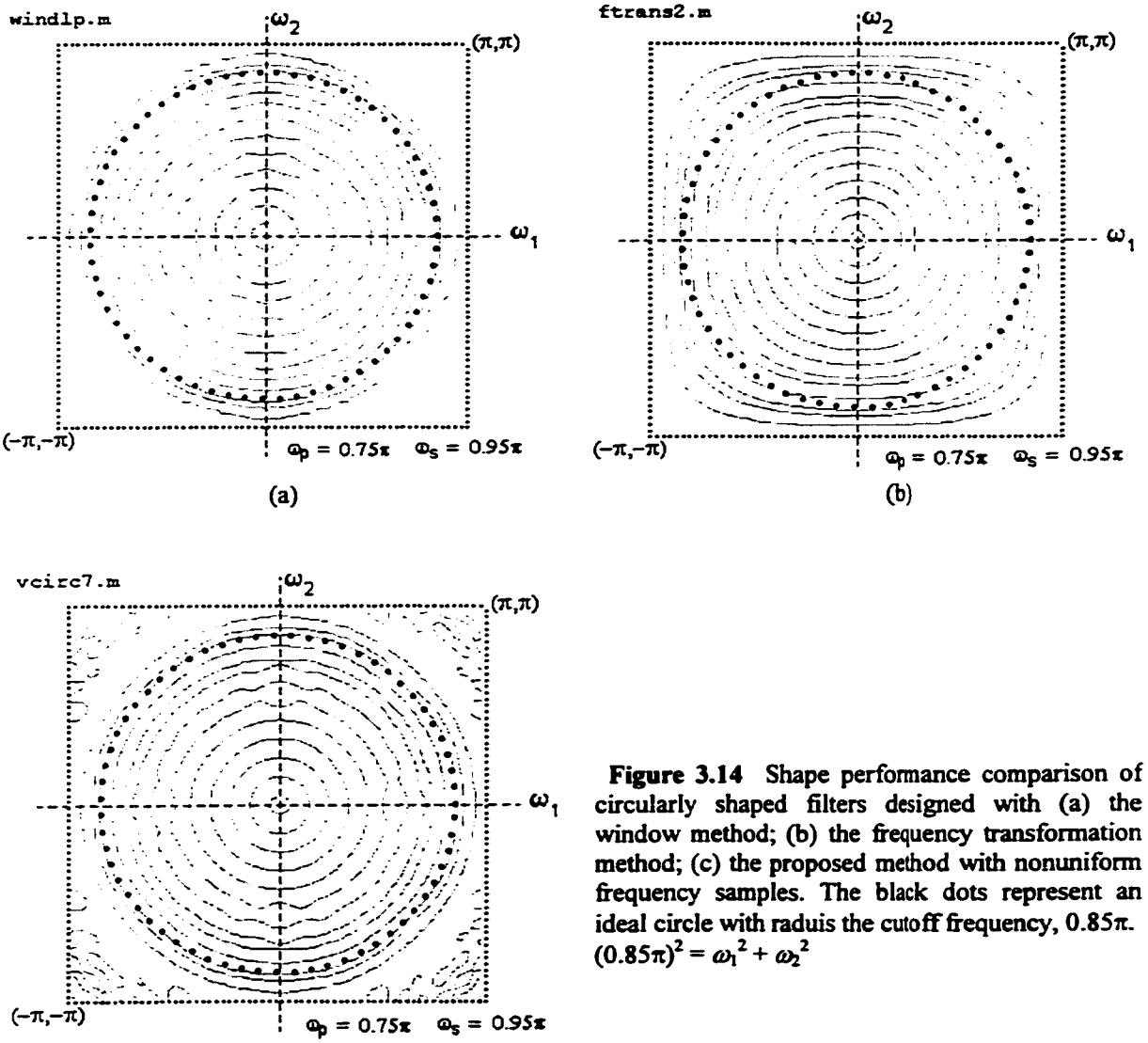


Figure 3.14 Shape performance comparison of circularly shaped filters designed with (a) the window method; (b) the frequency transformation method; (c) the proposed method with nonuniform frequency samples. The black dots represent an ideal circle with raduis the cutoff frequency, 0.85π. $(0.85\pi)^2 = \omega_1^2 + \omega_2^2$

The last example in this section is to show the limitations of the proposed technique for design of circularly shaped zero phase FIR filters with nonuniform frequency sampling. Although the approximation error is zero at the sampling points, it increases unacceptably between them in cases of very sharp transition bands and/or the filter order is relatively high and, therefore, the number of samples in the domain of approximation is large.

Example 3.8 A lowpass circularly shaped zero phase FIR filter with impulse response support size 25×25 points;

passband region $\mathcal{R}_p$: $0 \le \omega_1^2 + \omega_2^2 \le (0.48\pi)^2$;

stopband region $\mathcal{R}_s$: $(0.52\pi)^2 \le \omega_1^2 + \omega_2^2 \le \pi^2$.

The transition band now is only 0.04π. The 1-D extremal frequencies and corresponding amplitudes were obtained with 1-D filter order 22 and the above frequency specifications translated to 1-D, i.e., $\omega'_p = 0.48\pi$, $\omega'_s = 0.52\pi$. For a 25×25-point filter 91 frequency samples are necessary and their locations are shown in Fig. 3.15 (a).

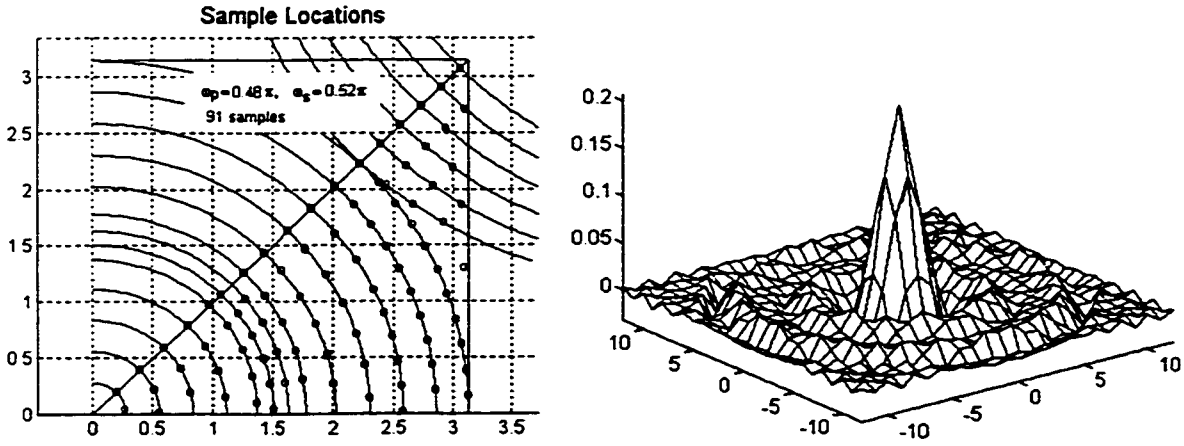


Figure 3.15 Example 3.8: A 25×25-point lowpass FIR filter with transition band width = 0.04π.
(a) Locations of the 91 samples; (b) impulse response of the designed filter.
On the next page: (c) and (d) frequency response magnitude contour and perspective plots.

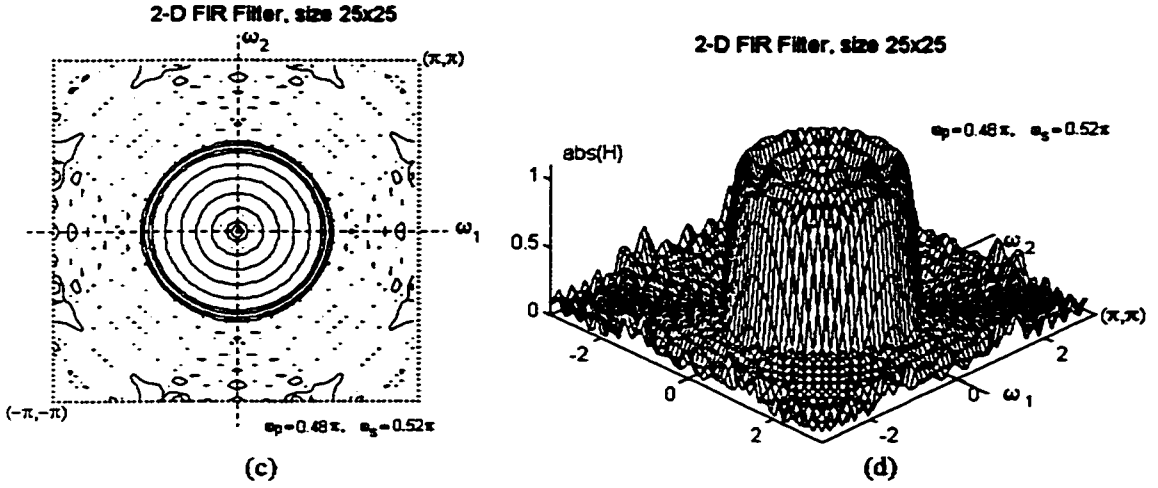


Figure 3.15 (continued) Example 3.8.

The deviations in this case are pretty large, especially in the stopband: $\delta_p = 0.1148$, $\delta_s = 0.3554$. The ripple of 1-D prototype (filter length $N_{1D} = 23$, $\omega_p' = 0.48\pi$, $\omega_s' = 0.52\pi$) is $\delta_{1D} = 0.0307$. Note that a circularly symmetric filter with the same impulse response support size, 25×25 points, and with the same frequency boundaries, $\omega_p = 0.48\pi$ and $\omega_s = 0.52\pi$, is difficult to design even with the "standard" methods considered previously: the window method and the frequency transformation method. This can be seen from the frequency response plots of these filters, shown in Fig. 3.16, (a) and (b).

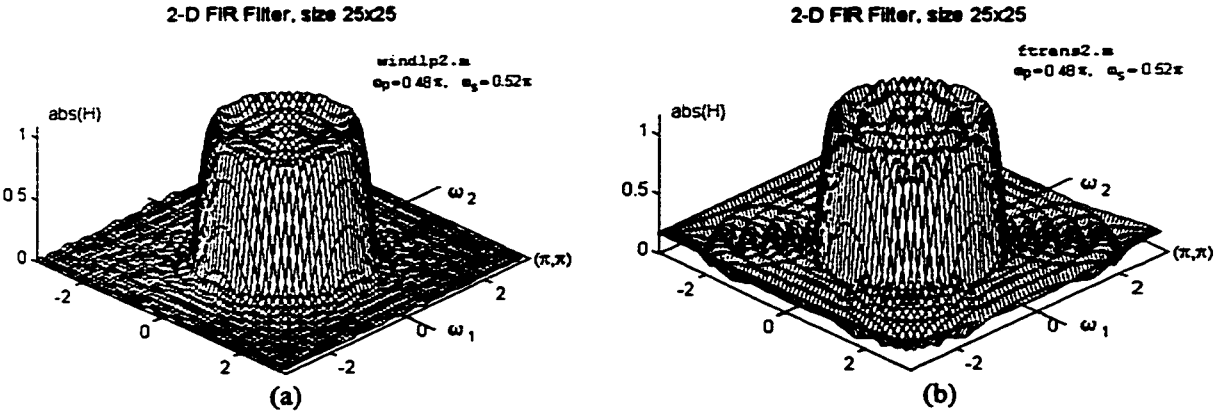


Figure 3.16. 25×25-point circularly shaped FIR filters designed using
(a) the rotated Kaiser window method and
(b) the McClellan transformation. In both cases $\omega_p = 0.48\pi$, $\omega_s = 0.52\pi$.
Deviations: (a) $\delta_p = 0.0966$, $\delta_s = 0.3756$ ($\delta_s = 0.0984$ for $\omega_1^2+\omega_2^2 \geq 0.6\pi$);
(b) $\delta_p = 0.1755$, $\delta_s = 0.5654$ ($\delta_s = 0.1749$ for $\omega_1^2+\omega_2^2 \geq 0.6\pi$);

The large stopband deviations are due to the fact that these designs fail with the stopband boundary. The true stopband boundary is actually at about 0.6π, see the note in parenthesis.

It is clear from the above examples what is the potential of the proposed sampling technique of frequency samples taken on circular contours in the region R_π having radii equal to 1-D extremal frequencies and frequency samples taken on hyperbolic contours in the region R_C. The number of samples on each contour in R_π, the density of contours and the number of samples on each contour in R_C are parameters that deserve optimization. In the next section some issues concerning these parameters are considered.

3.4 Some considerations concerning the sample density and interpolation method.

3.4.1 Sample density and locations.

It has been shown in the last section that, with the proposed techniques for nonuniform frequency sampling on circular contours, relatively high quality approximations of the four basic types of ideal piecewise constant, circularly symmetric FIR can be designed. Up to impulse response support sizes of about 25×25 points and transition bands as narrow as 0.15π, these filters are comparable, in terms of error and shape performance, to the two major "standard" approaches for designing 2-D circular zero phase FIR filters, the rotated Kaiser window and the McClellan's frequency transformation. Especially the technique with sampling on hyperbolic contours in the corner region R_C deserves attention. For filter sizes larger than 25×25 points and transition bans narrower than 0.15π the frequency samples happened to be too densely located and the bad effect of ill-conditioning is strong. Also, an effect is present, similar to the well-known in the numeric analysis example of Runge [29, 5] in the univariate interpolation. In general, the sequence of interpolation polynomials $\{P_N(x)\}$ obtained with equidistant points diverge. In other

words, the error $E(x)$ of interpolating a function $f(x)$ on equally spaced nodes in an interval $[a, b]$, $E(x) = f(x) - P_N(x)$, does not tend to zero as the polynomial order N increases. This nonconvergence is called the *Runge phenomenon*, [5]. An example is presented in Fig. 3.17 (a). The function

$$f(x) = \frac{1}{1+20x^4} \tag{3.4.1}$$

is approximated by a polynomial constructed on 9 equally spaced samples of this function in the interval [– 1, 1].

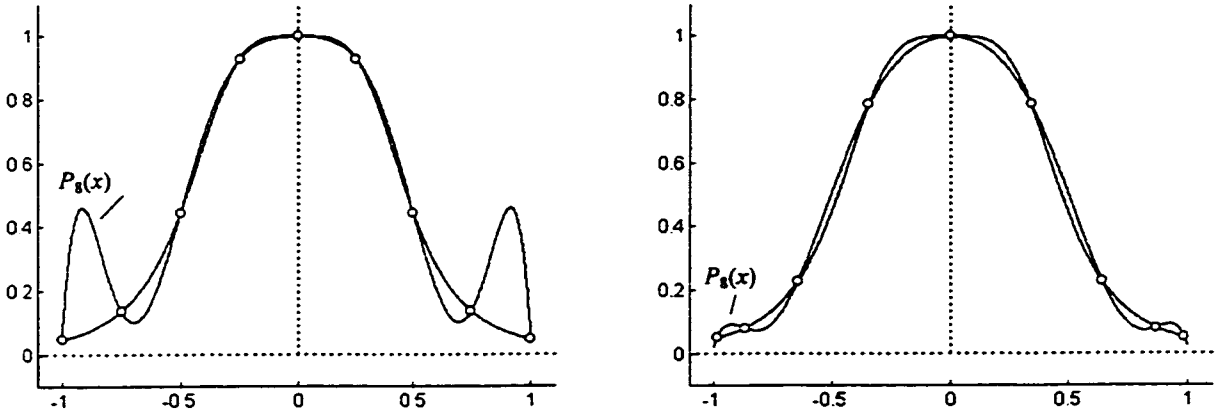


Figure 3.17 (a) Polynomial approximation to $f(x) = 1/(1+20x^4)$ based on 9 uniformly spaced nodes over [–1, 1]. (b) Polynomial approximation of the same function based on 9 Chebyshev nodes over [–1, 1].

The remedy to the Runge's phenomenon in 1-D is sampling at Chebyshev nodes. The Chebyshev nodes are the zeros of a Chebyshev polynomial $T_{N+1}(x)$ of order $N+1$, given by

$$x_k = \cos\frac{(2k+1)\pi}{2N+2} \qquad \text{for} \quad k = 0, 1, ..., N+1 \tag{3.4.2}$$

Fig. 3.17(b) shows how the error is decreased by constructing an interpolation polynomial of the same degree on Chebyshev nodes. In general, if $f(x)$ and its first derivative are continuous on [–1, 1], then it can be proven that Chebyshev interpolation will produce a sequence of polynomials $\{P_N(x)\}$ that converge uniformly to $f(x)$ over [–1, 1].

However, in 2-D the things are different. The straightforward extension of the above sampling approach in the frequency plane seems to be sampling at the zero locations of $T_M(x)T_M(y)$, where $x = \cos\omega_1$, $y = \cos\omega_2$. Some design simulations have been conducted with samples taken at the zero locations in the (ω_1, ω_2) plane, however the design results were not satisfactory.

From the design simulations conducted in this chapter it became clear that there is a lower and upper limit of the distance between the contours on which the frequency samples are taken. Several results in sampling and reconstruction with polar coordinates might prove to be very useful in the case of designing circularly symmetric FIR filters with frequency sampling. For example, Heideman and Veldhuis [33] considered reconstruction of two-dimensional continuous function of finite circular extent from samples of its Fourier transform. This study is closely related to the projection-slice theorem and computer tomography, cf. [A4]. It was shown that polar coordinate representation $F(\rho, \theta)$ of the Fourier transform $F(\Omega_1, \Omega_2)$ of a function $f(r, \varphi)$ of finite circular extent with radius a can be represented by a countably infinite set of one-dimensional functions $F_m(\theta)$. These functions describe $F(\rho, \theta)$ on concentric equidistant circles, with radii $\frac{m}{2a}$, $m = 0, 1, 2, \ldots$. Also, it was shown that every function F_m may be reconstructed from a finite number of samples N_m . Using the projection-slice theorem, the authors showed that the Fourier transform $F(\Omega_1, \Omega_2)$ may be sampled on circular contours with a minimum sampling distance $\frac{1}{2a}$ and the number of samples on each contour should satisfy $N_m = 2\pi m + 5$. Although the projection-slice theorem holds only for continuous signals, this result was used with a slight modification in the algorithms taking samples on the circular contours in the region R_π. The number of samples on the m-th circle is

$$N_m = \text{int}\left[2\pi m + 10\right], \qquad m = 0, \ldots, Q-1, \tag{3.4.3}$$

where int[ ] denotes the nearest integer and Q is the number of 1-D extremal frequencies. With a number of samples on each contour determined by (3.4.3) the performance

improvement was significant. The design examples showed that the minimum distance between adjacent contours is even more important than the number of samples on each contour. However, the projection-slice theorem does not hold for discrete-space signals. Moreover, it can be argued that $\frac{1}{2a}$ is the minimum sampling distance. Therefore, the question about the minimum distance between sampling contours in the discrete-space case remains open.

The interpolation matrix **C**, Eq. (3.1.19), becomes ill-conditioned in case of narrow transition bands and relatively high filter order primarily because of the basis functions used, Eq. (3.1.8), and its large size in such cases. Improvement of the solution accuracy can be obtained, for example, by using residual correction or also known as mixed precision iterative improvement, [32]. If the linear system (3.1.18), $\hat{\mathbf{H}} = \mathbf{Cb}$, is solved with partial pivoting factorization $\boldsymbol{P}\hat{\mathbf{H}} = \boldsymbol{LU}$, the accuracy can be improved by repeating the following algorithm

$$\begin{aligned} &\mathbf{r} = \hat{\mathbf{H}} - \mathbf{Cb} \ \text{(double precision)} \\ &\text{Solve} \quad \boldsymbol{Ly} = \boldsymbol{Pr} \quad \text{for } \boldsymbol{y}. \\ &\text{Solve} \quad \boldsymbol{Ux} = \boldsymbol{y} \quad \text{for } \boldsymbol{x}. \\ &\mathbf{b} = \mathbf{b} + \boldsymbol{x} \end{aligned} \tag{3.4.4}$$

If t-digit arithmetic is used to compute $\boldsymbol{P}\hat{\mathbf{H}} = \boldsymbol{LU}$, **b**, $\boldsymbol{x}$, and $\boldsymbol{y}$, then $2t$-digit arithmetic is to be used for $\hat{\mathbf{H}} - \mathbf{Cb}$. The number of iterations depends on the machine precision and the condition number of **C**, [32]. The method is relatively computationally cheap. Each iteration costs $O(L^2)$, to be compared with the original $O(L^3)$ expenses in the factorization $\boldsymbol{P}\hat{\mathbf{H}} = \boldsymbol{LU}$. However, if **C** is badly enough conditioned with respect to the machine precision, then no improvement may result. The mixed precision iterative improvement is machine-dependent and this limits its use in software for wide distribution.

Another improvement of the interpolation problem solution can be obtained with the least squares solution of an overdetermined system of equations, i.e., the minimization of $\left\|\mathbf{Cb} - \hat{\mathbf{H}}\right\|_2$. This method is considered in the next subsection.

3.4.2 Linear least-squares solution.

Another possible solution to the large deviations in case of large sample density and narrow transition bands is the least-squares solution of an overdetermined linear system. Since the maximum deviations are in the form of sharp peaks in between the sample locations, it can be expected that a least squares fit will smooth off these peaks. This is done by taking more frequency samples than the number of independent filter points and an overdetermined system of linear equations is solved. If the number of frequency samples is N_s, and the number of independent filter coefficients is L, where $N_s \geq L$, then the least squares solution is in fact minimization of (cf. Eq. 3.1.18)

$$\left\| \mathbf{Cb} - \hat{\mathbf{H}} \right\|_2 \tag{3.4.5}$$

where the matrix $\mathbf{C}$ now is $(N_s \times L)$ ant the vector containing the sample values $\hat{\mathbf{H}}$ is $(N_s \times 1)$. The most reliable solution procedure for this problem involve the reduction of $\mathbf{C}$ to various canonical forms via orthogonal transformations [32]. One of the most popular factorizations is the $\boldsymbol{QR}$ factorization $\mathbf{C} = \boldsymbol{QR}$ where $\boldsymbol{Q}$ is an orthogonal matrix of size $(N_s \times N_s)$ and $\boldsymbol{R}$ is an upper triangular matrix of size $(L \times L)$. There are several methods for computing an orthonormal basis for a set of vectors. Along with the classical Gram-Schmidt orthogonalization process and the numerically more stable modified Gram-Schmidt, methods based on Householder, block Hauseholder, Givens and fast Givens transformations exist. Information about these methods can be found for example in Golub and Van Loan [32]. In case when $\mathbf{C}$ is rank deficient, i.e., rank($\mathbf{C}$) < L, $\boldsymbol{QR}$ with column pivoting should be used. In our case the matrix $\mathbf{C}$ is almost always full rank. For the solution in the least squares sense of the overdetermined system of equations $\mathbf{Cb} = \hat{\mathbf{H}}$, the effective rank of $\mathbf{C}$ is determined from a $\boldsymbol{QR}$ decomposition with pivoting. A drawback of the method is the large number of floating point operations for the linear system solution. The Householder $\boldsymbol{QR}$ decomposition with pivoting requires $4N_sL - 2r^2(N_s + L) + 4r^3/3$ floating point operations where r = rank($\mathbf{C}$), [32].

Example 3.9: A bandpass zero phase FIR filter of size 33×33 points and with frequency boundaries as follows:

passband $\mathcal{R}_p$: $|H_d(\omega_1, \omega_2)| = 1$ for $(0.4\pi)^2 \le \omega_1^2 + \omega_2^2 \le (0.6\pi)^2$,

stopband $\mathcal{R}_s$: $|H_d(\omega_1, \omega_2)| = 0$ for $0 \le \omega_1^2 + \omega_2^2 \le (0.2\pi)^2$ and $(0.8\pi)^2 \le \omega_1^2 + \omega_2^2 \le \pi^2$.

A total of 162 samples were taken, Fig. 3.18(a), using the approach of the previous section. An eightfold symmetric filter of size 33×33 has 153 independent points. Therefore, only 9 samples were taken in excess.

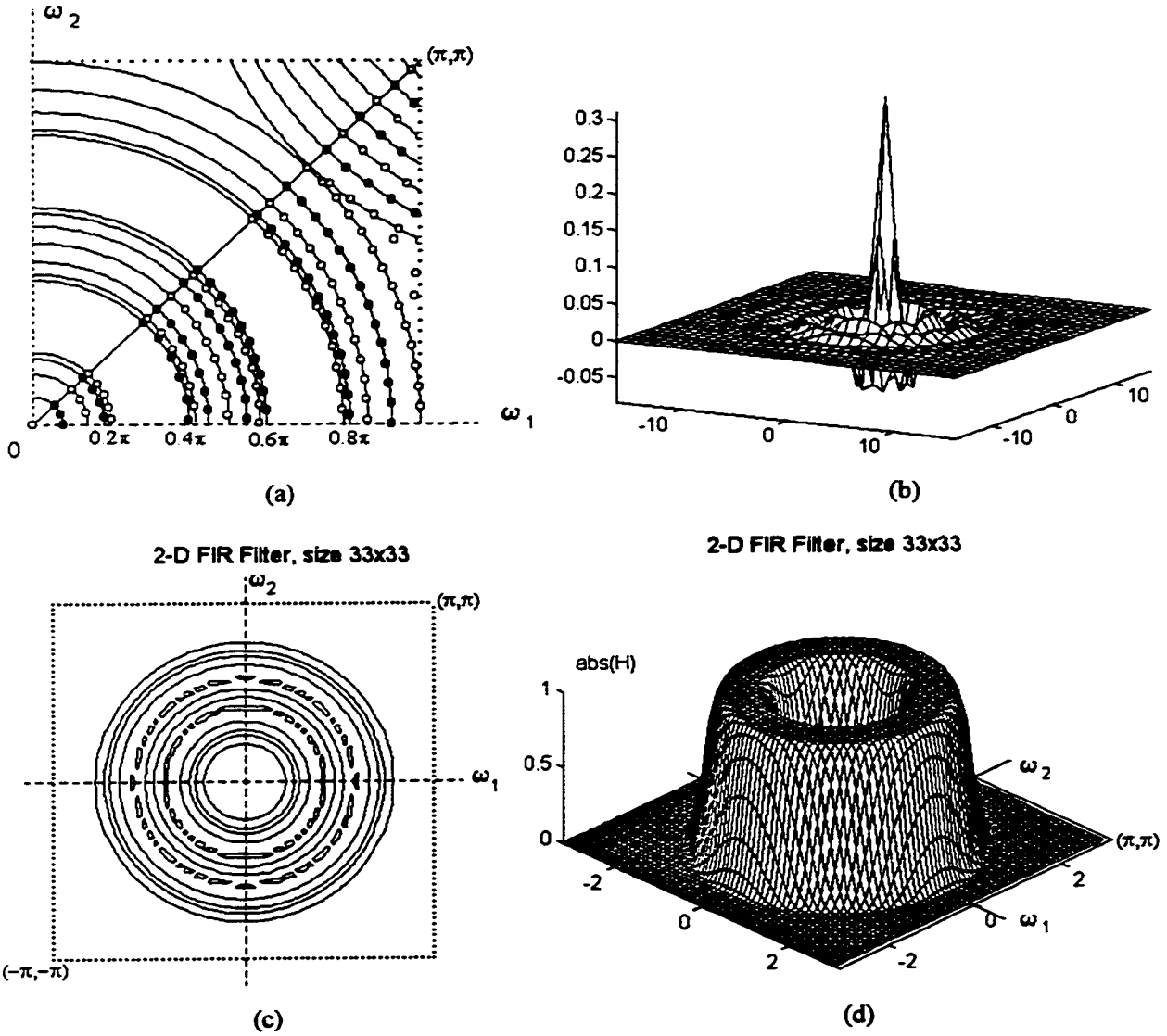


Figure 3.18 A circularly shaped bandpass zero phase FIR filter designed using the sampling technique described in Sec. 3.3.3 and using a least-squares solution.

Although the residual correction and least squares solution methods can improve the solution precision to a certain degree, they have their limitations and drawbacks. Some more elegant solution should be considered as, for example, interpolating in polar coordinates. An interesting method for reconstruction of a bandlimited 2-D signals from a set of nonuniform samples lying on concentric circles centered at the origin is presented in [31]. The function and its samples are represented in polar coordinates. The polar coordinate representation seem to be the right approach in our case, too.

Some design results obtained with the sampling techniques proposed in this chapter are shown in Table 3.2 for comparison. Comparison is made also with the standard design methods for circular shape FIR filters: the rotated 1-D Kaiser window and the frequency transformation method. The abbreviations used in the table stand for, as follows:

Exp Samples on circular contours with exponential distribution in R_π and on straight lines in R_C, Sec. 3.3.1;
R_π Samples taken only in the region R_π, Sec. 3.3.2;
Sc Scaling the 1-D frequency axis, Sec. 3.3.3;
HC Samples on hyperbolic contours in R_C, Sec. 3.3.4;
Win Rotated 1-D Kaiser window;
FT Frequency transformation using the McClellan's transformation sequence.
LS Least squares solution.

Lowpass Filters							
Filter size	1-D filter length, N_{1D}	Passband edge, ω_p	Stopband edge, ω_s	1-D ripple, δ_{1D}	Passband dev., δ_p	Stopband dev., δ_s	Method (sampling)
11×11	-	0.35π	0.65π	-	0.0398	0.0538	Exp†
	9			*0.0549*	0.0551	0.0891	R_π
	11			*0.0581*	0.0581	0.0932	Sc
	9			*0.0549*	0.0551	0.0791	HC
	-			-	0.0611	0.0515	Win
	11			*0.0174*	0.0174	0.0674	FT
15×15	-	0.35π	0.65π	-	0.0148	0.0165	Exp
	13			*0.0173*	0.0174	0.0787	R_π
	17			*0.0181*	0.0181	0.0280	Sc
	13			*0.0173*	0.0174	0.0312	HC
	-			-	0.0461	0.0168	Win
	15			*0.0057*	0.0057	0.0403	FT

Table 3.2 Performance comparison of circular shape FIR filter design with the proposed nonuniform sampling techniques. **(a)** Lowpass filters.

Filter size	1-D filter length, N_{1D}	Passband edge, ω_p	Stopband edge, ω_s	1-D ripple, δ_{1D}	Passband dev., δ_p	Stopband dev., δ_s	Method (sampling)
21×21	-	0.4π	0.6π	-	0.0084	0.0345	Exp
	23			*0.02363*	0.0315	0.0754	Sc
	17			*0.02376*	0.0238	0.0849	HC
	-			-	0.0265	0.0288	Win
	21			*0.01136*	0.01141	0.1040* (0.01140)	FT
23×23	-	0.75π	0.95π	-	0.0045	0.1843	Exp
	21			*0.0106*	0.0109	13.5789	R_π
	21			*0.0106*	0.0259	0.0562	HC
	-			-	0.0294	0.0192	Win
	23			*0.0056*	0.0057	0.5336* (0.0552)	FT
23×23	21	0.38π	0.62π	*0.00565*‡	0.00762	0.00745	HC
23×23	21	0.42π	0.58π	*0.0307*‡	0.0308	0.0341	HC
25×25	27	0.38π	0.62π	*0.0058*‡	0.0064	0.0107	Sc
	23	0.38π	0.62π	*0.0024*	0.0027	0.0077	HC
	23	0.48π	0.52π	*0.1746*	0.1818	0.3970	HC
	-	0.48π	0.52π	-	0.0966	0.3756* (0.0984)	Win
	25	0.48π	0.52π	*0.1746*	0.1755	0.5654* (0.1749)	FT

Table 3.2 (a) .Lowpass filters (continued).

Bandpass Filters						
$N \times N$, N_{1D}	Frequency edge radii, ω_{s1} , ω_{p1} , ω_{p2}, ω_{s2}	1-D ripple, δ_{1D}	Passband δ_p	Stopbands δ_{s1},	δ_{s2}	Method (sampling)
23×23, 21	0.1π, 0.3π, 0.6π, 0.8π	*0.013467*	0.014109	0.013467,	0.032416	HC
21×21, 21	0.1π, 0.3π, 0.6π, 0.8π	*0.029594*	0.029855	0.029594,	0.055167	Sc
23×23, –	0.1π, 0.3π, 0.6π, 0.8π	-	0.020189	0.053733,	0.020382	Win
21×21, 21			0.013543	0.013467,	0.273272*	FT

Table 3.2 (b) .Bandpass filters.

Bandstop Filters					
$N \times N$, N_{1D}	Frequency edge radii, ω_{s1} , ω_{p1} , ω_{p2}, ω_{s2}	1-D ripple, δ_{1D}	Stopband δ_s	Passbands δ_{p1}, δ_{p2}	Method (sampling)
25×25, 23	0.2π, 0.4π, 0.6π, 0.8π	*0.011513*	0.011611	0.011557, 0.028892	HC
25×25, 27	0.2π, 0.4π, 0.6π, 0.8π	*0.011861*	0.011861	0.012207, 0.038974	Sc
25×25, --	0.2π, 0.4π, 0.6π, 0.8π	-	0.017434	0.054256, 0.016468	Win
25×25, 25	0.2π, 0.4π, 0.6π, 0.8π	0.004045	0.010134	0.004050, 0.242094*	FT

Table 3.2 (c) .Bandstop filters.

† All designs with exponential contour distributions use $\alpha = 0.4$.

‡ means that the 1-D optimal filter was designed using weights, usually 1:2 in favor of the stopband error minimization. The larger ripple is shown.

* The deviation values marked with an asterisk, mostly for the frequency transformation method, are large since the design fails with the stopband or passband boundaries and this deviation is actually measured in the transition band. The design should be done by choosing appropriate 1-D pass- and stopband edges, such that the 2-D passband region is completely inside $\cos \omega_p' = F(\omega_1,\omega_2)$, and the stopband region is completely outside $\omega_s' = F(\omega_1,\omega_2)$, see Eq. (3.3.1). The deviation in the actual pass- or stopband is given in parenthesis.

Lowpass Filters, Least-Squares method							
$N \times N$, N_{1D}	Passband edge, ω_p	Stopband edge, ω_s	Number of samples, N_s	1-D ripple, δ_{1D}	Passband dev., δ_p	Stopband dev., δ_s	Method (sampling)
31×31, 29	0.35π	0.65π	142	*0.000234*	0.000235	0.001422	LS
31×31, 29	0.40π	0.60π	140	*0.002713*	0.002731	0.015097	LS
31×31, 37	0.45π	0.55π	216	*0.014726*	0.046693	0.048541	LS
37×37, 37	0.45π	0.55π	197	*0.014726*	0.015788	0.056534	LS
39×39, 37	0.46π	0.54π	216	*0.027785*	0.029628	0.049505	LS

Table 3.2 (d) .Lowpass filters, LS method.

Bandpass Filters						
$N \times N$, N_{1D}	Frequency edge radii, ω_{s1} , ω_{p1} , ω_{p2}, ω_{s2}	1-D ripple, δ_{1D}	Passband δ_s	Stopbands δ_{p1},	δ_{p2}	Method (sampling)
33×33, 30, 153 samples	0.2π, 0.4π, 0.6π, 0.8π	*0.001815*	0.001890	0.001821,	0.205364	HC
33×33, 30, 162 samples	0.2π, 0.4π, 0.6π, 0.8π	*0.001815*	0.001833	0.001821,	0.008588	LS
33×33, –	0.2π, 0.4π, 0.6π, 0.8π	-	0.002982	0.003033,	0.003712	Win
33×33, 33	0.2π, 0.4π, 0.6π, 0.8π	*0.001227*	0.001247	0.001233,	0.209816*	FT

Table 3.2 (e) Bandpass filters, least squares method is compared with other methods.

All filters, including the window method, were designed using Matlab programs written by the author. Exception is the frequency transformation method, for which the Matlab program `ftrans2.m`, included in the Image Processing Toolbox, was used. The routine `vcirc3.m` is for taking samples in the region R_π only. The sampling by scaling is performed with `vcirc4.m`, and the sampling on hyperbolic contours in R_C is done with `vcirc7.m`. The system of linear equations is solved and the filter coefficient were determined in all cases with `cnint2d.m`, except for the LS solution, where `cnint8.m` was used. Listings of these programs can be found in Appendix A.

3.5 Summary and Conclusions.

In this chapter sampling techniques for designing circularly shaped zero-phase FIR filters with nonuniform frequency sampling were proposed. These techniques are simple conceptually and for implementation. With them a circularly shaped FIR filters with very regular shape and low passband and stopband deviations can be designed. The range of the impulse response support size and the range of transition bands covered with this design method are sufficient for the needs of most practical applications.

The proposed techniques produce circularly shaped filters whose performance is comparable and sometimes even superior to the existing methods. As the design results showed and it can be seen from Table 2.1, in many cases the deviations of the resulting 2-D FIR filter frequency response are almost the same as the ripple of the 1-D optimal filter. This is especially true for the circular region around the origin of the (ω_1, ω_2) plane and for filter with transition bands of 0.2π and wider. This fact suggests that it is possible to design 2-D FIR circularly shaped filters with deviations as low as the ripple of the corresponding 1-D optimal filter.

A drawback of the proposed methods are the comparatively large amount of computation, mostly for the solution of the system of linear equations. Another weak point is the ill-conditioning which is very large for medium and high order filters. These limitations can be overcome using some efficient bivariate interpolation techniques. However, this implies further limitations to the frequency sample locations and, therefore, possible deterioration in shape performance.

At the end of this chapter, it was shown that circular shape FIR filters with narrow transition band (about 0.05π) and increased impulse response support size can be designed with nonuniform frequency sampling using the proposed techniques with the number of samples greater than the number of independent filter coefficients, i.e., by solving an overdetermined system of linear equations. A drawback of this approach is the increased number of arithmetic operations. And the result does not always justify the expenses, i. e., the filters designed using the least-squares fit have deviations of magnitude comparable with the deviations obtained with the conventional approach with slightly wider transition band(s) and lower filter order.

Chapter 4
Diamond and Fan FIR Filters

This chapter is primarily concerned with the design of two-dimensional half-band FIR filters with nonuniform frequency sampling. The outline of this chapter is as follows. In Section 4.1, we discuss the existing symmetries in the frequency response and impulse response of half-band diamond and 90° fan shaped FIR filters and the basis functions used in the generalized bivariate polynomial expression of their frequency response. Also in this section, the minimum number of frequency samples is considered and the tolerance scheme specifications for these filters are presented. In Section 4.2, two classes of sampling techniques are introduced which can be used to design diamond and 90° fan filters with nonuniform frequency sampling. The first technique is based on sampling on parallel lines passing through scaled extremal frequencies obtained from a 1-D optimal FIR filter design. The second technique utilizes sampling on parallel lines with exponential distribution. Both techniques are illustrated with design examples. At the end of this section, design results obtained with the proposed techniques are summarized in tables and compared with results obtained with the McClellan frequency transformation method. In the next section of this chapter, Section 4.3, the potential of the proposed nonuniform sampling techniques is demonstrated with the design of 2-D zero-phase FIR filters with various shapes of their frequency response, e.g., X-shaped filters, cross shaped filters, etc. The conclusions are presented in Section 4.4.

The diamond and fan filters are of special importance for practical applications. The 2-

D half-band diamond FIR filters can be used as a downsampling filters for quincunxially[†] sampled 2-D data. In order to reduce the data rate for digital transmission of HDTV signals, for example, a quincunx downsampling followed by quincunx upsampling at the receiving end is performed. This method is preferred to orthogonal downsampling since it does not reduce the resolution in the horizontal and vertical directions in which the human visual system is more sensitive. Half-band diamond filters have applications in such downsampling schemes as pre- and postfilters. Another application of these filters is as an interlace-to-noninterlace scanning converters of digital TV signals.

Fan filters are used for the discrimination of seismic waves based on their velocity of arrival at a 1-D array of sensors. Most of the published design work has been on the 90° fan filters. These filters are attractive since their symmetries and properties of the impulse response lead to efficient filter designs and implementations. Also, a 90° fan filter can be very easily derived form a halfband diamond filter and vice versa by simply shifting the frequency response in ω_1 or ω_2 direction by π. However, the 90° fan filter has some limitations as normalized speed cutoff of 1, impossibility to discriminate two waves arriving at the array from opposite directions. Other fan filters, as 60° and 30° fan filters, prove to be more useful.

Several methods have been proposed to design half-band diamond and fan FIR filters. In [28], an l_p-norm minimization technique is used to design such filters. The McClellan frequency transformation method [24] is also well known. These design methods produce filters with passband ripples. Yoshida *et al.* [27] have proposed a design method for maximally flat 2-D half-band FIR filters. Mitra *et al.* [1] and Bagchi and Mitra [2] introduce a nonuniform frequency sampling method for designing diamond and 90° fan filters. The frequency samples are taken at the extrema of analytic functions based on Chebyshev polynomials.

The design of diamond and fan filters with nonuniform frequency samples, as it will be shown in this chapter, is not so difficult as the one of circularly shaped filters. This is

[†] There is a small difference between quincunx sampling and hexagonal sampling.

primarily due to the frequency response contours of these filters which are straight lines. However, some limitations exist in this case, too. In the next section the design of half-band diamond and half-band fan filters with nonuniform frequency sampling will be considered.

4.1 Half-Band Diamond and Fan Zero Phase FIR Filters Symmetries.

It will be useful in the beginning to recall the definition of a 2-D half-band filter. Let's start again with the system function of a 2-D zero phase FIR filter with impulse response region of support of $N \times N$ points, where $N = 2M + 1$.

$$H(z_1, z_2) = \sum_{n_1=-M}^{M} \sum_{n_2=-M}^{M} h(n_1, n_2) z_1^{-n_1} z_2^{-n_2} \tag{4.1.1}$$

The frequency response is a real function, $H(\omega_1 , \omega_2) = H^*(\omega_1 , \omega_2)$, and for real valued impulse response we have symmetry about the origin, $h(n_1 , n_2) = h(-n_1 , -n_2)$.

The definition of a 2-D half-band filter is as follows [27]: A 2-D zero phase FIR filter with fourfold symmetry is said to be a 2-D half-band filter if and only if its frequency response satisfies

$$H(\omega_1, \omega_2) + H(\pi - \omega_1, \pi - \omega_2) = 1 \tag{4.1.2}$$

for arbitrary ω_1 and ω_2 . This means that the frequency response is symmetric about the point $(\omega_{10},\omega_{20}, H(\omega_{10},\omega_{20})) = (\pi/2, \pi/2, 0.5)$ in the region $0 \le \omega_1,\omega_2 \le \pi$. Similarly to the 1-D case of a half-band filter, where $h_{1D}(n) = 0$ for n = even, $n \neq 0$, in 2-D it can be shown that [27]

$$h(n_1, n_2) = 0, \quad \text{for } n_1 + n_2 = \text{even and } n_1, n_2 \neq 0. \tag{4.1.3}$$

This a consequence of the fact that the impulse response contains a factor of the form $\frac{\sin\left[\frac{\pi}{2}(n_1+n_2)\right]}{\pi^2(n_1+n_2)}$. Let's consider the sum

$$H(\omega_1,\omega_2)+H(\omega_1+\pi,\omega_2+\pi)=\sum_{n_1}\sum_{n_2}h(n_1,n_2)e^{-j\omega_1 n_1}e^{-j\omega_2 n_2}+$$

$$\sum_{n_1}\sum_{n_2}h(n_1,n_2)e^{-j(\omega_1+\pi)n_1}e^{-j(\omega_2+\pi)n_2} \qquad (4.1.4)$$

$$=\sum_{n_1}\sum_{n_2}\left\{h(n_1,n_2)e^{-j\omega_1 n_1}e^{-j\omega_2 n_2}\left[1+e^{-j\pi(n_1+n_2)}\right]\right\}$$

For $n_1+ n_2$ = odd the sum in the brackets is zero. For $n_1+ n_2$ = even according to (4.1.3) $h(n_1, n_2) = 0$ except for $n_1 = n_2 = 0$. Therefore,

$$H(\omega_1,\omega_2)+H(\omega_1+\pi,\omega_2+\pi)=2h(0,0)\ . \qquad (4.1.5)$$

Since the impulse response is symmetric about the origin and real valued,

$$H(\omega_1+\pi,\omega_2+\pi)=H(-\omega_1-\pi,-\omega_2-\pi) \qquad (4.1.6)$$

and since $H(\omega_1 , \omega_2)$ is periodic with period 2π in ω_1 and in ω_2,

$$H(-\omega_1-\pi,-\omega_2-\pi)=H(\pi-\omega_1,\pi-\omega_2)\,. \qquad (4.1.7)$$

From this result and from (4.1.2) follows that

$$H(\omega_1,\omega_2)+H(\pi-\omega_1,\pi-\omega_2)=2h(0,0)=1 \qquad (4.1.8)$$

The 90° fan filters have quadrant symmetry, i.e., the frequency response is symmetric with respect to the ω_1 and ω_2 axes. The diamond filters have eightfold symmetry, i.e., in addition to the axes, their frequency response is also symmetric with respect to both diagonals $\omega_1 = \omega_2$ and $\omega_1 = -\omega_2$, Fig. 4.1.

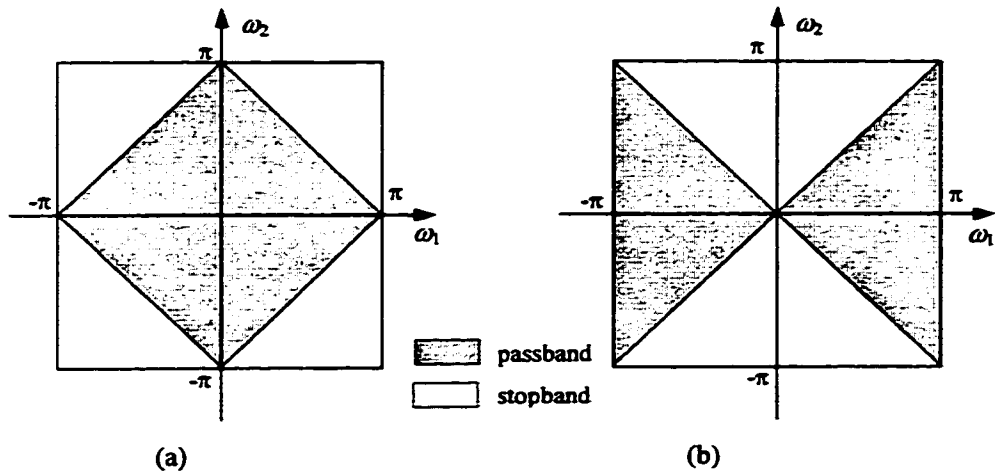


Figure 4.1 (a) Ideal half-band diamond filter; (b) ideal half-band 90° fan filter.

$$H(\omega_1,\omega_2) = H(-\omega_1,\omega_2) = H(\omega_1,-\omega_2) = H(\omega_2,\omega_1) \tag{4.1.9}$$

which implies

$$h(n_1, n_2) = h(-n_1,n_2) = h(n_1,-n_2) = h(n_2,n_1) \tag{4.1.10}$$

As in the previously considered cases, a frequency response with fourfold symmetry can be expressed in the form, cf. Eqs. (2.2.9) and (3.1.5),

$$\begin{aligned} H(\omega_1,\omega_2) = h(0,0) + \sum_{n_1=1}^{M} 2h(n_1,0)\cos(n_1\omega_1) + \sum_{n_2=1}^{M} 2h(0,n_2)\cos(n_2\omega_2) + \\ + \sum_{n_1=1}^{M}\sum_{n_2=1}^{M} 4h(n_1,n_2)\cos(n_1\omega_1)\cos(n_2\omega_2) \end{aligned} \tag{4.1.11}$$

The diamond filter is symmetric with respect both diagonals, $\omega_2 = \omega_1$, and $\omega_2 = -\omega_1$. That implies that $h(n_1, n_2) = h(n_2, n_1)$. Therefore, the independent impulse response points are limited in the shaded triangular region in the first quadrant of the (n_1, n_2) plane, which is shown in Fig. 4.2.

Having in mind Eq. (4.1.8), the frequency response $H(\omega_1, \omega_2)$ takes the following form

$$H(\omega_1,\omega_2)=0.5+\sum_{n_1=1}^{M}\sum_{n_2=0}^{n_1-1}d(n_1,n_2)\left[\cos(n_1\omega_1)\cos(n_2\omega_2)+\cos(n_2\omega_1)\cos(n_1\omega_2)\right] \quad (4.1.12)$$

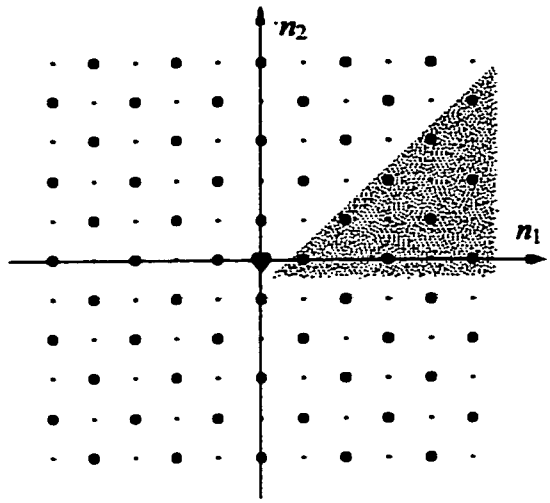


Figure 4.2 Impulse response of a 2-D half-band FIR filter of size 11×11 points.
• $h(n_1, n_2)$ = nonzero; · $h(n_1, n_2)$ = zero ; ♥ $h(0,0) = 0.5$
The shaded region contains the independent filter coefficients.

$$\text{where } d(n_1,0)=\begin{cases}2h(n_1,0), & n_1=\text{odd}\\ 0, & n_1=\text{even}\end{cases} \quad (4.1.12a)$$

$$\text{and} \quad d(n_1,n_2)=\begin{cases}4h(n_1,n_2), & n_1+n_2=\text{odd},\ n_2\neq 0\\ 0, & n_1+n_2=\text{even},\ n_2\neq 0\end{cases} \quad (4.1.12b)$$

With simplifications of notations, the above expression for the frequency response of a half-band diamond FIR filter can be written as

$$H(\omega) = 0.5+\sum_{k=1}^{L} d(k)\phi_k(\omega) \quad (4.1.13)$$

where $\omega = (\omega_1, \omega_2)^T$ is the frequency vector and the basis functions $\phi_k(\omega) = \phi_k(\omega_1, \omega_2)$ are given by

$$\phi_k(\omega_1, \omega_2) = \cos(\omega_1 n_1)\cos(\omega_2 n_2) + \cos(\omega_1 n_2)\cos(\omega_2 n_1) \ . \tag{4.1.14}$$

The coefficients $d(k) = d(n_1, n_2)$. The index k is related to n_1 and n_2 by

$$k = \begin{cases} \dfrac{(n_1 - 1)(n_1 + 1)}{4} + \dfrac{n_2 + 2}{2}, & \text{for } n_1 = \text{odd},\ n_2 = \text{even} \\ \dfrac{n_1^2}{4} + \dfrac{n_2 + 1}{2}, & \text{for } n_1 = \text{even},\ n_2 = \text{odd} \end{cases} \ . \tag{4.1.15}$$

This relations, as well as the number L of independent filter coefficients, the number L of basis functions for interpolation, respectively, were obtained using the arithmetic progression properties of the independent points, see Fig. 4.3 below.

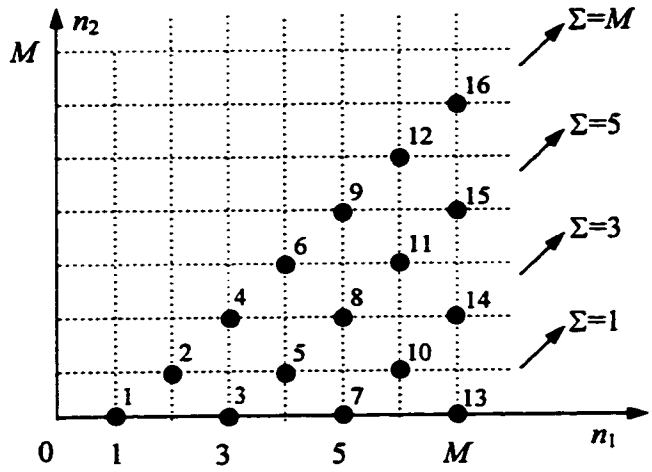


Figure 4.3. The independent points in an impulse response of size $(2M+1)\times(2M+1) = 15\times15$. The index k is marked near each point.

Besides the formula for the number L of independent filter coefficients for a 2-D half-band FIR filter of size $(2M+1)\times(2M+1) = N\times N$,

$$L = \left\lfloor \frac{M+1}{2} \right\rfloor \left\lfloor \frac{M+2}{2} \right\rfloor = \left\lfloor \frac{N+1}{4} \right\rfloor \left\lfloor \frac{N+3}{4} \right\rfloor \tag{4.1.16}$$

which is given in the literature, e. g. [27], from the diagonal sums it follows that

$$L(M)=\begin{cases}\dfrac{M(M+2)}{4}, & \text{for } M=\text{even}\\ \dfrac{(M+1)^2}{4}, & \text{for } M=\text{odd}\end{cases} \tag{4.1.17}$$

Also, there is an recurrence relation which can be used in programming:

$$L(M)=\frac{M(M+1)}{2}-L(M-1), \qquad L(1)=1. \tag{4.1.18}$$

The filter specifications for a diamond half-band filter are shown in Fig. 4.4. The diagonal $\omega_1=\omega_2$ intersects the passband and stopband boundaries at points (ω_p,ω_p) and (ω_s,ω_s), respectively.

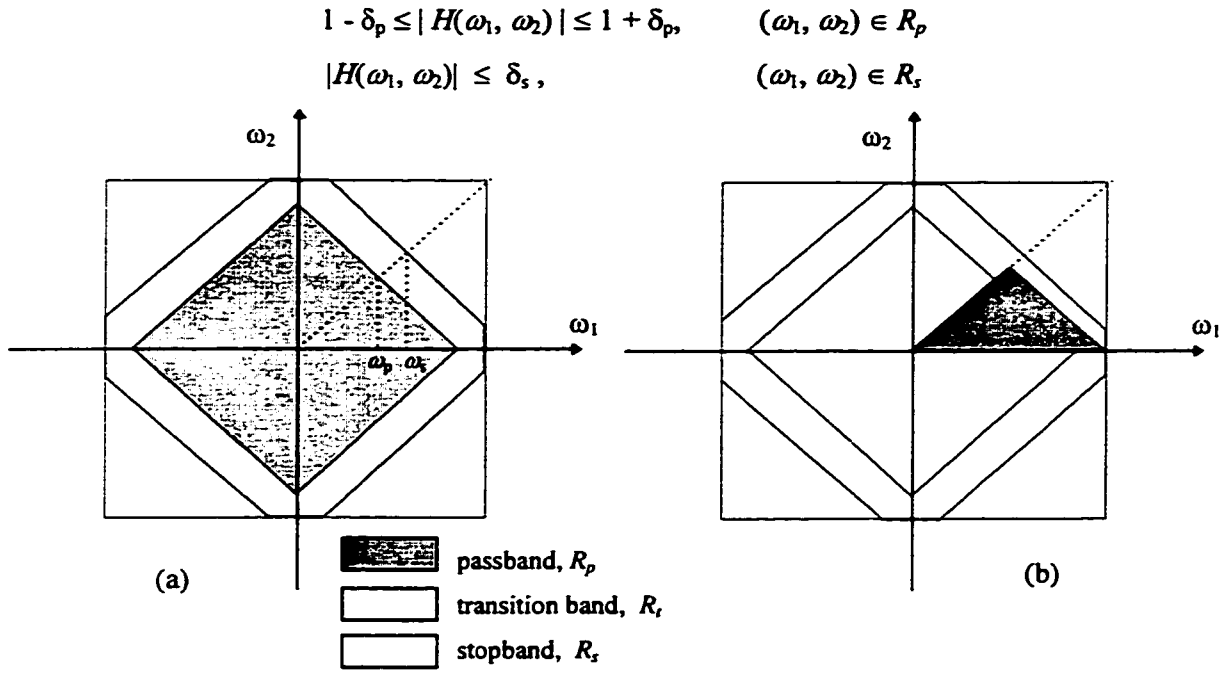


Figure 4.4 (a) Frequency response specifications of a half-band diamond filter; (b) Domain of approximation (the darkly shaded triangle)

L frequency samples taken at distinct locations should be sufficient to determine the coefficients d(k) and, subsequently, the impulse response $h(n_1, n_2)$. Because of the symmetry about the point $(\omega_{10},\omega_{20}, H) = (\pi/2, \pi/2, 0.5)$, the L frequency samples can be

taken only in the darkly shaded triangular region in Fig. 4.4(b). The solution of the interpolation problem depends on the linear independence of the L characteristic vectors

$$\Phi(\omega_k) = \left[\phi_1(\omega_k) \quad \phi_2(\omega_k) \quad \cdots \quad \phi_L(\omega_k)\right]^T, \quad k = 1, \ldots, L \tag{4.1.19}$$

where the functions $\phi_k(\omega) = \phi_k(\omega_1, \omega_2)$ are as defined by (4.1.14). The system of linear equations to be solved can be expressed in matrix-vector form as

$$\mathbf{D}\mathbf{d} = \tilde{\mathbf{H}} \tag{4.1.20}$$

where $\mathbf{D}$ is an $L \times L$ matrix having as rows the characteristic vectors $\Phi(\omega_k)$, $\mathbf{d}$ is an $L \times 1$ vector containing the unknowns $d(k)$, and $\tilde{\mathbf{H}}$ is an $L \times 1$ vector with entries $H(\omega_{1k},\omega_{2k})$ - 0.5, $k = 1, \ldots, L$. In case of nonsingularity of $\mathbf{D}$, the impulse response is found by solving the system

$$\mathbf{d} = \mathbf{D}^{-1}\tilde{\mathbf{H}} \tag{4.1.21}$$

and then using (4.1.12).

The expression for the frequency response of a 90° fan FIR filter can be derived very easily from the expressions for a diamond filter since the fan filter frequency response is obtained by shifting the diamond filter frequency response by π or $-\pi$ in the direction of ω_1 or ω_2, i.e., $H_F(\omega_1,\omega_2) = H_D(\omega_1 - \pi,\omega_2)$, which is equivalent in the space domain to

$$h_F(n_1,n_2) = e^{j\pi n_1} h_D(n_1,n_2) = (-1)^{n_1} h_D(n_1,n_2) \tag{4.1.22}$$

Since the half-band diamond filter is an eightfold symmetric filter, $h_D(n_1, n_2) = h_D(n_2, n_1)$, and since for both filters the nonzero coefficients are at (n_1, n_2) = (even, odd) or (n_1, n_2) = (odd, even) only, from (4.1.22) it follows that

$$h_F(n_1,n_2) = -h_F(n_2,n_1) \tag{4.1.23}$$

Therefore, all the same expressions for a diamond half-band FIR filter can be used for a 90° half-band fan filter. The only difference is in the expression of the basis functions:

$$\phi_k(\omega_1, \omega_2) = \cos(\omega_1 n_1)\cos(\omega_2 n_2) - \cos(\omega_1 n_2)\cos(\omega_2 n_1). \tag{4.1.24}$$

Everything else is the same, of course by taking account of (4.1.23). For completeness,

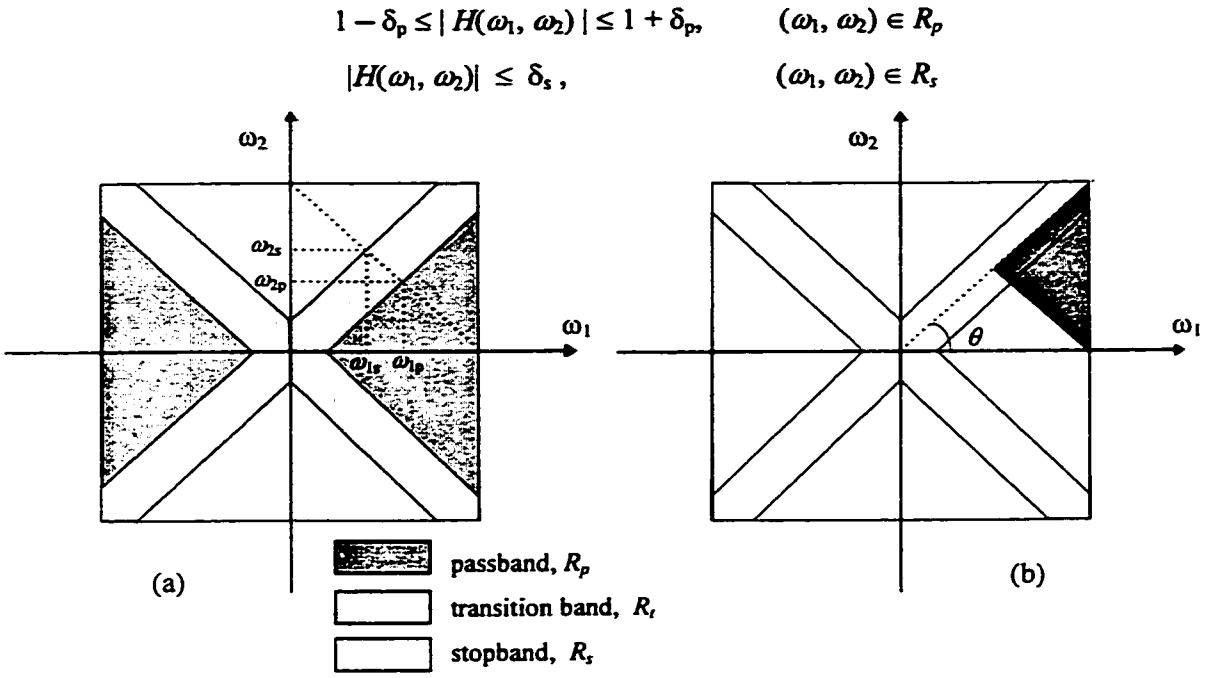


Figure 4.5 (a) Frequency response specifications of a half-band fan filter; (b) Domain of approximation (the darkly shaded triangle)

the filter coefficients are now obtained by solving the system of linear equations

$$\mathbf{Fd} = \tilde{\mathbf{H}} \tag{4.1.25}$$

where $\mathbf{F}$ is an $L \times L$ matrix having as rows the characteristic vectors $\Phi(\omega_k)$, now with entries $\phi_k(\omega_{1k}, \omega_{2k})$ given by (4.1.24). $\mathbf{d}$ is an $L \times 1$ vector containing the unknowns $d(k)$, and $\tilde{\mathbf{H}}$ is an $L \times 1$ vector containing the frequency sample values.

Since in the case of a half-band fan filter the symmetry about the point $(\pi/2, \pi/2, 0.5)$ holds, the region in which the frequency samples are taken can be limited to the triangular

region with vertices $(\pi, 0)$, (π, π), and $(\pi/2, \pi/2)$, shown in Figure 4.5.

Though the specifications of the ideal 90° fan filter are usually given as

$|H(\omega_1, \omega_2)| = 1$ for $-\pi/4 \le \theta \le \pi/4$ (or for $\pi/4 \le \theta \le 3\pi/4$),

$|H(\omega_1, \omega_2)| = 0$ otherwise,

with the tolerance scheme on Fig. 4.5 it is more convenient to use a passband and stopband edge frequencies, defined as shown in the figure: $\omega_p = \omega_{1p} = \omega_{2s}$ and $\omega_s = \omega_{1s} = \omega_{2p}$ and these frequencies are determined from the intersection of the line $\omega_2 = -\omega_1 + \pi$ with the passband edge and stopband edge, respectively.

4.2 Half-band Diamond and Fan FIR Filter Design.

(i) Half-band diamond shaped filters.

The design of these filters with nonuniform frequency sampling employs similar sampling techniques as in the previous two chapters. For good results the samples should be taken in the domain of approximation on contours describing the shape of the filter to be designed. For a diamond shaped filter these contours are parallel lines with slope -1. For a 90° degree fan filter the samples should be taken on parallel lines with slope 1. The sampling parameters are the number of parallel lines, the distribution of these lines in the region of approximation, the number of samples on each line, and the spacing between the samples on a given line. The design simulations showed good results with lines with exponential distribution or lines passing trough frequencies corresponding to extremal frequencies for 1-D optimal filter design. How this mapping is performed will be explained with the next example.

Example 4.1 A half-band diamond FIR filter with frequency edge specifications: $\omega_p = 0.42\pi$, $\omega_s = 0.58\pi$, unity gain in the passband and zero gain in the stopband. Impulse

response region of support size: 13×13 points. Since $M = 6$, from (4.1.17) we have $L = 12$ independent coefficient out of 139. Therefore, 12 frequency samples must be taken in the region of approximation, see Fig. 4.4(b). The lines on which the samples are taken have slope -1 and pass through the intersections of vertical lines with the main diagonal $\omega_2 = \omega_1$. Now, the vertical lines have as coordinates the extremal frequencies up to $\omega_p = 0.42\pi$ obtained with the Remez exchange algorithm and mapped one to one to the ω_1 axis. The algorithm for optimal 1-D linear phase, odd length FIR filter was initiated with frequency edges $\omega'_p = 0.42\pi$, $\omega'_s = 0.58\pi$, and filter order = 16, ($N_{1D} = 17$). which produces $Q = 10$ extremal frequencies. (Remember that $Q = (N_{1D} + 3)/2$ for N_{1D} = odd). From these only the 5 frequencies lying in the passband are taken, from $\omega' = 0$ to $\omega' = \omega'_p$. Therefore, five vertical lines are obtained in the (ω_1, ω_2) plane. Alternatively, this can be seen as mapping the 1-D frequencies [0, π] to the diagonal [(0,0), (π,π)] in the 2-D plane. The samples on each of the slanted line segments are taken approximately proportionally to the segment length and such that their number is $L = 12$, Fig. 4.6(a). The samples on a given line have the values of the corresponding 1-D amplitudes for the given extremal frequency. The deviations in the passband and stopband of the resulting filter are $\delta_p = 0.0575$ and $\delta_s = 0.0754$, respectively. The 1-D optimal prototype filter has ripple $\delta_{1D} = 0.0417$.

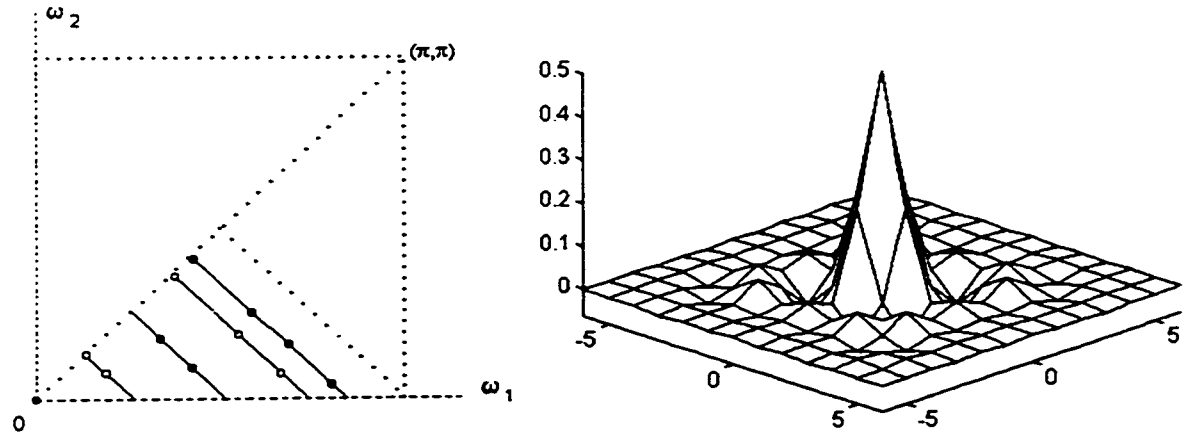


Figure 4.6 Example 4.1: A 13×13 diamond shaped FIR filter with $\omega_p = 0.42\pi$, $\omega_s = 0.58\pi$ and designed from 12 frequency samples.
(a) sample locations, o sample value = $1+\delta_{1D}$, • sample value = $1-\delta_{1D}$;
(b) impulse response of the designed filter; (continues on the next page)

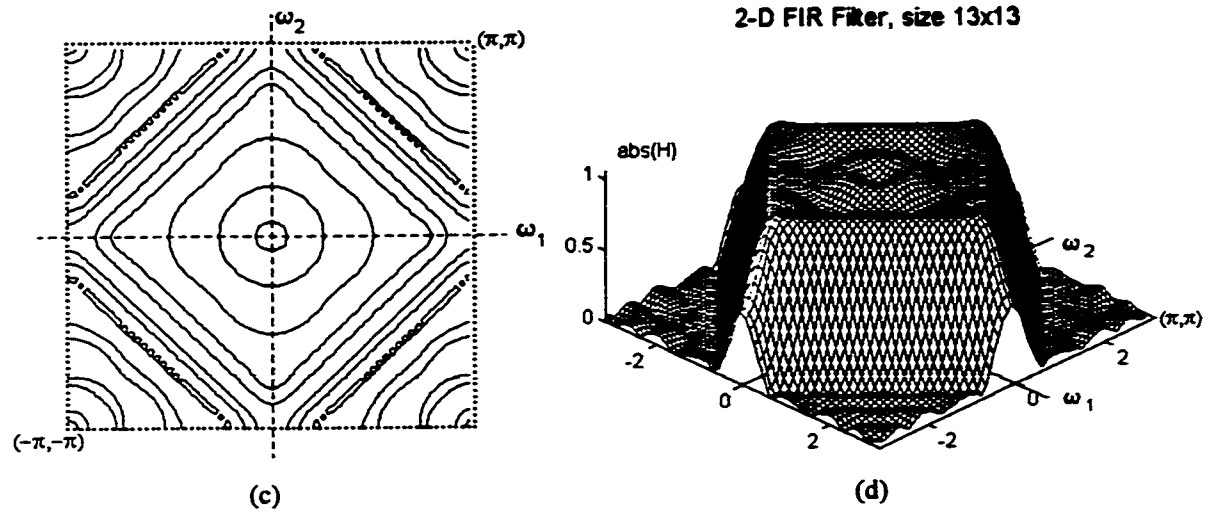


Figure 4.6 (continued) Example 4.1: (c) frequency response contour plot; (d) perspective plot.

A Matlab program `diam2.m` has been written, which takes automatically the frequency samples, given the frequency edges and the 1-D optimal filter order. If the number of samples taken is more than the nearest number satisfying (4.1.17) then the redundant samples are discarded in such a way that the remaining samples cover the region of approximation approximately with uniform density. Using the approach of taking more samples than necessary and then discarding the surplus has a positive effect of improving the condition number of matrix **D**, Eq. (4.1.20). This happens since in this way the conditions of Theorem 1.6 (cf. Sec. 1.2) are receded.

The next example demonstrates the potential of the nonuniform frequency sampling method for *high quality* FIR filter design. A diamond half-band FIR filter with very small deviations is obtained. The frequency samples are taken on parallel lines as in the previous example. However, now all the samples are set to 1, and not the corresponding 1-D values $1\pm\delta_{1D}$.

Example 4.2 A diamond half-band FIR filter with $\omega_p = 0.38\pi$, $\omega_s = 0.62\pi$, filter size 25×25 points. An impulse response with this support size has 42 independent coefficients, Eq. (4.1.16) or (4.1.17). The necessary 42 samples were taken on lines parallel to $\omega_2 = -\omega_1$ and passing trough scaled extremal frequencies along $\omega_2 = \omega_1$. These extremal frequencies were obtained with the Remez exchange algorithm used for an optimal lowpass FIR filter of order 30 and passband and stopband edge frequencies equal to the edges of desired 2-D diamond filter, i.e., $\omega'_p = 0.38\pi$ and $\omega'_s = 0.62\pi$.

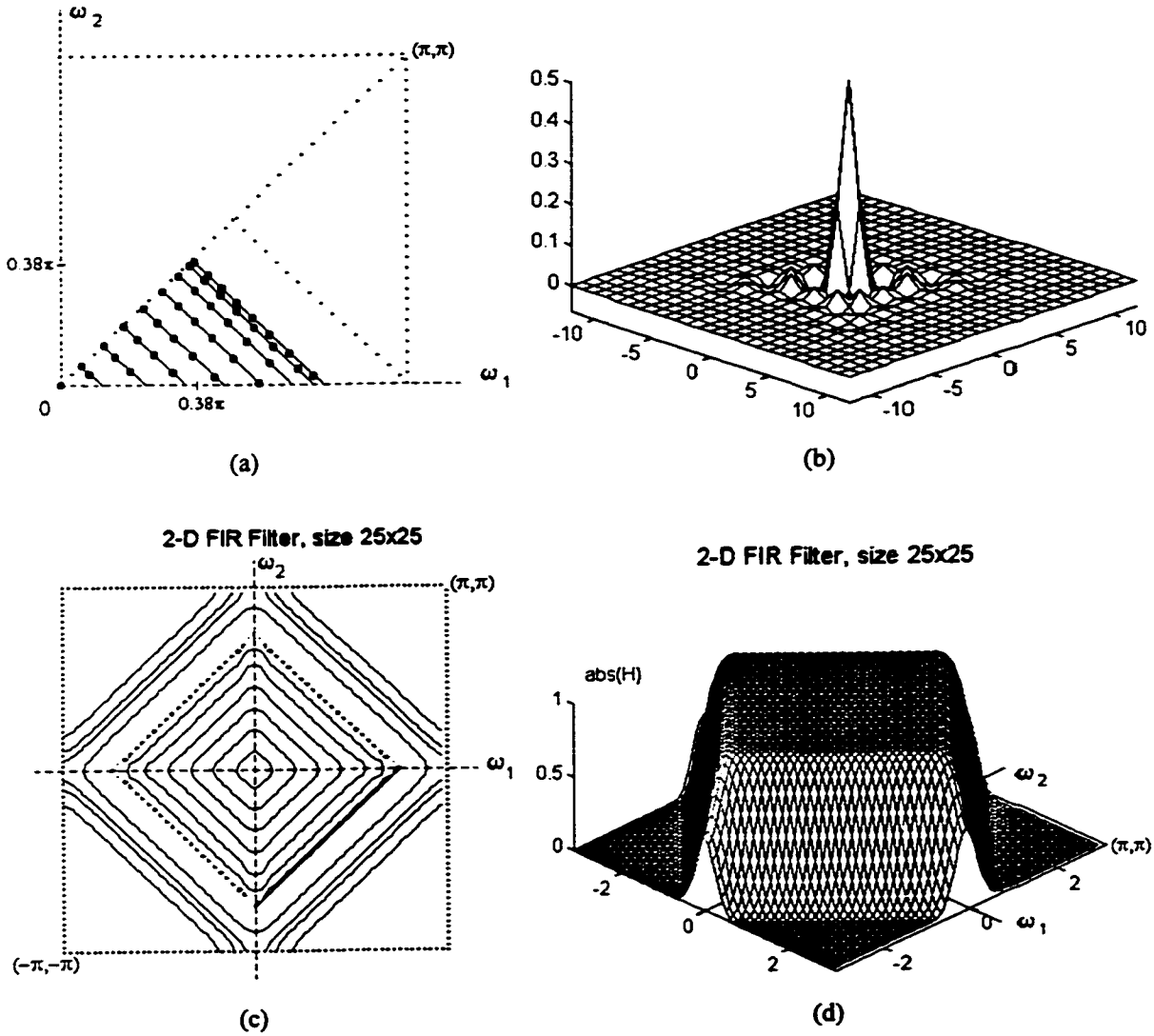


Figure 4.7 Example 4.2: A 25×25 diamond shaped FIR filter with $\omega_p = 0.38\pi$, $\omega_s = 0.62\pi$ and designed from 12 frequency samples. (a) sample locations, • sample value = 1; (b) impulse response of the designed filter; (c) Frequency response contour plot; (d) frequency response magnitude perspective plot.

As in the previous example, the number of samples on each line segment is approximately proportional to the segment length. All samples have value of unity. What is interesting in this example is that the resulting 2-D passband and stopband deviations are *smaller* than the ripple of the 1-D optimal filter which extremal frequencies have been used. The deviations in the passband and stopband of the resulting 2-D diamond shaped filter are $\delta_p = 2.79\times10^{-4}$ and $\delta_s = 1.88\times10^{-4}$, respectively, while the 1-D optimal filter has ripple $\delta_{1D} = 4.48\times10^{-4}$.

It has to be noted that the design of 2-D half-band diamond shaped FIR filters with nonuniform frequency sampling using the technique described with the last two examples is not so easy as it might seem. Because of the basis functions used, $\{\phi_k(\omega_1, \omega_2)\}$, matrix **D** is often badly conditioned, and in some cases may even become singular. For example, in Example 4.1 the condition number of **D** is $\mu = 73.89$, while in the last example of a quality diamond filter design $\mu = 350938.62$.

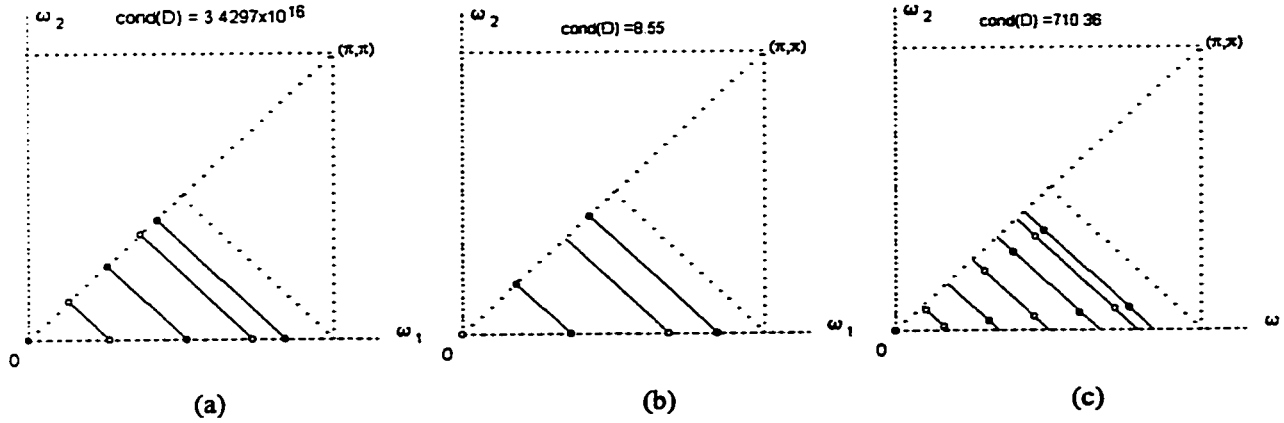


Figure 4.8 Cases when matrix D, Eq. (4.1.20), is (a) nearly singular; (b) and (c) nonsingular.

The interpolation problem runs into singularity when the frequency samples are taken only on the lines $\omega_2 = 0$ and $\omega_2 = \omega_1$ and the degree of Eq. (4.1.11) is ≥ 4 both in ω_2 and ω_1 (9 samples or more). This effect is an indirect consequence of Theorem 1.5, Sec. 1.2, and can be eliminated by simply taking samples in the exterior of approximation region

triangle. This is illustrated in Fig. 4.8.

Several Matlab programs have been written by the author, which take samples in the triangular region in the first frequency plane quadrant using the extremal frequencies of a 1-D optimal filter or on parallel lines distributed exponentially. Some of this programs, `diam2.m`, `diam4.m`, are presented in the appendix. The interpolation is performed with the function `dnint2d.m`, also shown in the appendix. The method with samples on exponentially distributed parallel lines and sample values all set to 1 often gives better results than the use of 1-D extremal frequencies. This sampling technique will be presented with the 90° fan filter design. Both techniques are essentially the same for the two filter types.

(ii) Half-band fan shaped filters.

A 90° fan filter with passband for $-\pi/4 \le \theta \le \pi/4$ is designed by taking samples in the triangular region with vertices $(\pi, 0)$, (π, π), $(\pi/2, \pi/2)$, Fig. 4.5(b). The samples lie on lines parallel to the diagonal $\omega_2 = \omega_1$. These lines pass trough the intersection of $\omega_2 = -\omega_1+\pi$ and horizontal lines having as coordinates the extremal frequencies of 1-D optimal design and mapped one to one to the ω_2 axis. With the technique with samples on exponentially distributed parallel lines the points on the line $\omega_2 = -\omega_1+\pi$ trough which the parallel lines pass are found in the following way. The samples of the function

$$f(x_i) = 1 - e^{-\alpha x_i} \quad \text{at } x_i = i/(P-1) \text{ for } i = 0, 1, ..., P-1 \text{ and fixed } \alpha \tag{4.2.1}$$

are mapped on the ω_2 axis in the interval $[0, \omega_p]$ in the following way

$$\omega_{2i} = \frac{\omega_p}{1 - e^{-\alpha(P-1)/P}} f(x_i). \tag{4.2.2}$$

This is similar to the mapping used in the design of rectangularly shaped FIR filters, see Sec. 2.4 and 2.5. Next, the intersections of the horizontal lines passing trough points ω_{2i} with $\omega_2 = -\omega_1+\pi$ determine the parallel lines with slope 1 on which the frequency points are taken. The number of samples on each line segment is proportional to its length and such that the total number of samples L satisfies (4.1.17). In order to adjust this number and to improve the condition number of matrix **F**, here again is used the approach of taking a bit more samples than necessary and discarding the excess number of them in a "pseudorandom way". The parameter P is proportional to the desired filter size. The parameter α is the most important. It deserves special attention and further investigation since the passband and stopband deviations depend strongly on it and if chosen properly the designed filter even can be optimal in Chebyshev sense. This statement is given without a theoretical proof but some design examples show that this may be the case.

As in the design of diamond shaped filters, both sampling techniques have been used in the design of 90° fan filters. Similarly, in most of the design experiments, the technique with exponentially distributed parallel lines proved to produce superior results in terms of approximation error compared to the technique employing 1-D extremal frequencies. Once a diamond shaped FIR of good quality is obtained from frequency samples, a good quality 90° fan FIR filter is obtained too, by simply employing (4.1.22), with almost the same deviations. The firs example shows the design of an 11×11 fan filter designed with frequency samples taken on exponentially distributed parallel lines in the domain of approximation.

Example 4.3 A 90° fan FIR filter with passband oriented along the ω_1 axis and passband and stopband edges $\omega_p = 0.62\pi$, $\omega_s = 0.38\pi$, respectively. Impulse response region of support size: 11×11 points.

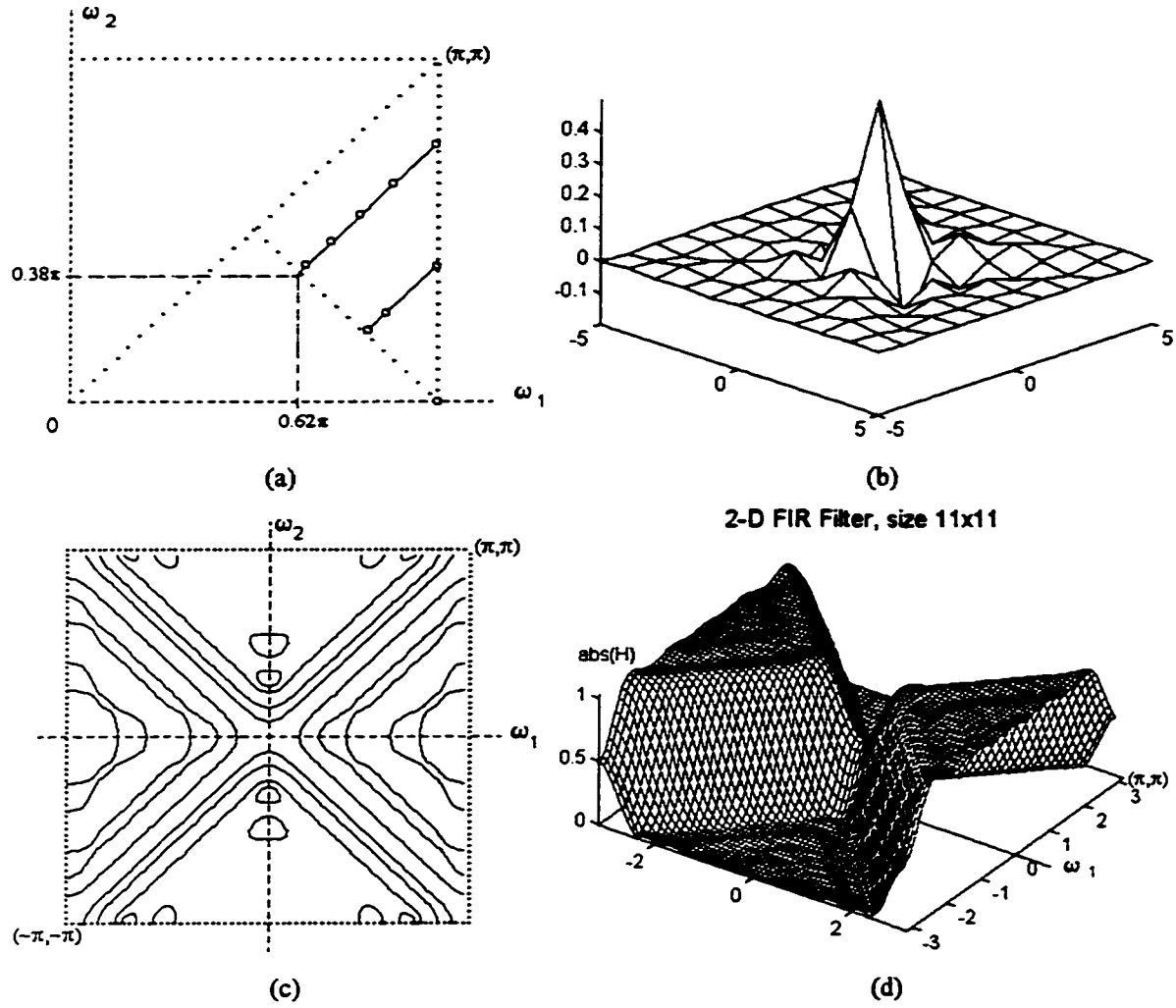


Figure 4.8 Example 4.3: 11×11 90° fan filter designed with samples on exponentially distributed parallel lines in the domain of approximation. (a) Sample locations; (b) impulse response of the designed filter; (c) Frequency response contour plot; (d) frequency response magnitude perspective plot.

Nine frequency samples are sufficient to solve for the 9 independent impulse response points of an 11×11 half-band fan filter. The locations of these samples are shown in Fig. 4.8 (a). All samples have value of unity. For this design the parameter in Eq. (4.2.1) was set to $\alpha = 0.35$. As a practical guidance, good results can be obtained with $\alpha = 0.3$ to 0.8 for low order filters, 7×7 to 13×13, and α should be gradually increased to about $\alpha = 1.6$ to 2.1 for filter sizes 23×23 and larger. The frequency response magnitude is shown in Fig. 4.8(d). The filter exhibits deviations $\delta_p = \delta_s = 0.0338$ both in the passband and stopband. This and the next results show that the proposed method is comparable and in many cases superior to the existing fan FIR filter design methods, both in terms of shape

and peak deviations.

Example 4.4 A half-band fan FIR filter with passband oriented along the ω_1 axis and band edge frequencies $\omega_p = 0.58\pi$, $\omega_s = 0.42\pi$. Impulse response region of support size: 25×25 points.

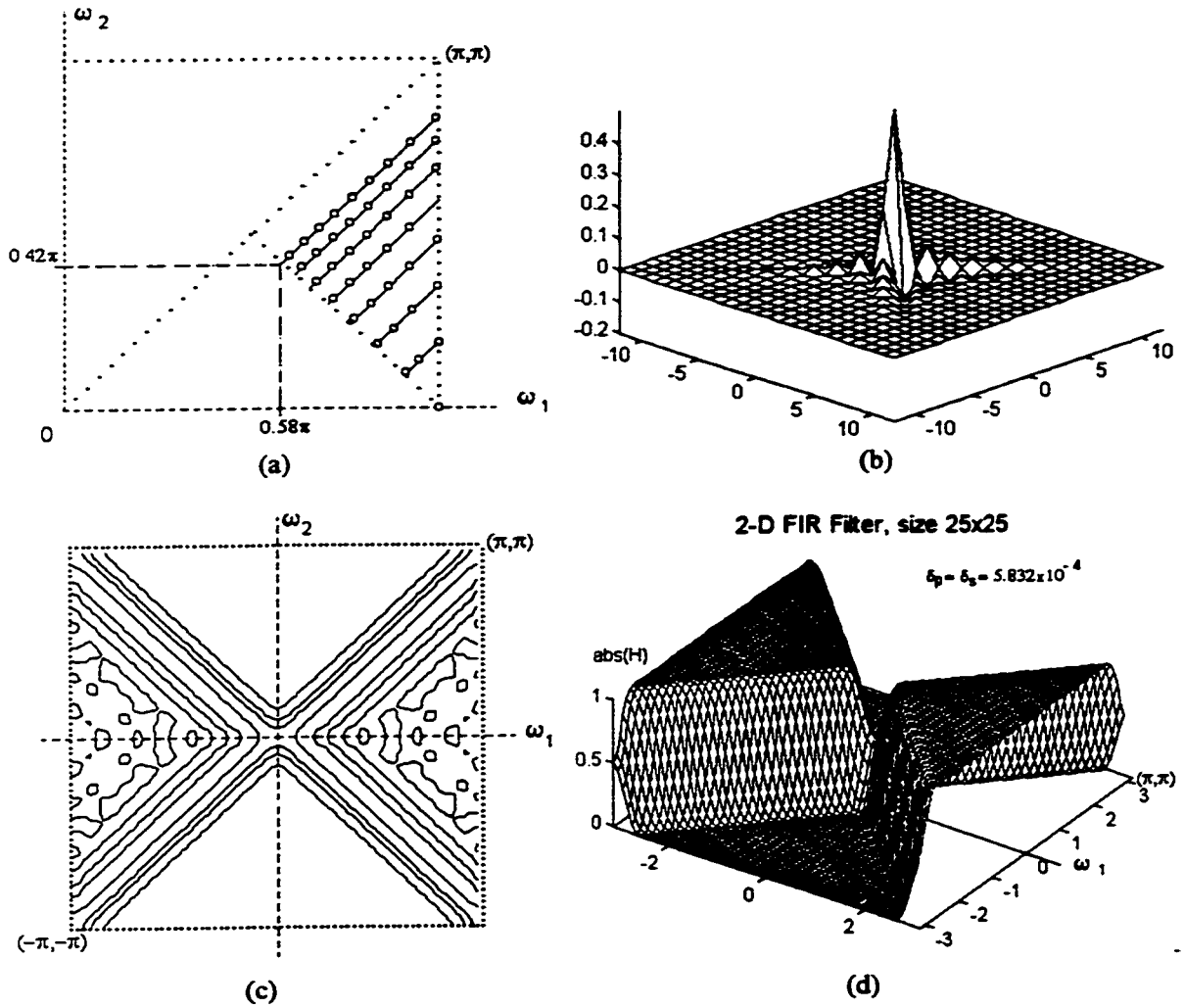


Figure 4.9 Example 4.4: 25×25 90° fan filter designed with samples on exponentially distributed parallel lines in the domain of approximation.
(a) Sample locations; (b) impulse response of the designed filter;
(c) Frequency response contour plot; (d) frequency response magnitude perspective plot.

The number of independent filter coefficients in this case is 42. The 42 frequency samples obtained with $\alpha = 1.41$ in (4.2.1) and having value 1 are shown in Fig. 4.9(a). The shape

regularity can be seen from the frequency response contour plot, Fig. 4.9(c). The maximum deviations for the passband and stopband are equal and small:

$\delta_p = \delta_s = 5.83\times10^{-4}$.

A 90° fan FIR filter with passband oriented along the ω_2 axis, i.e., the ideal filter is defined as

$|H_i(\omega_1, \omega_2)| = 1$ for $-\pi/4 \le \theta \le \pi/4$ (or for $\pi/4 \le \theta \le 3\pi/4$),

$|H_i(\omega_1, \omega_2)| = 0$ otherwise,

can be simply obtained from a fan filter with passband along the ω_2 axis by simply rotating the impulse response matrix by 90° (or –90°, or transposing it), and vice versa. This is the easiest way and it has been done with the filter of the last example, Fig. 4.10.

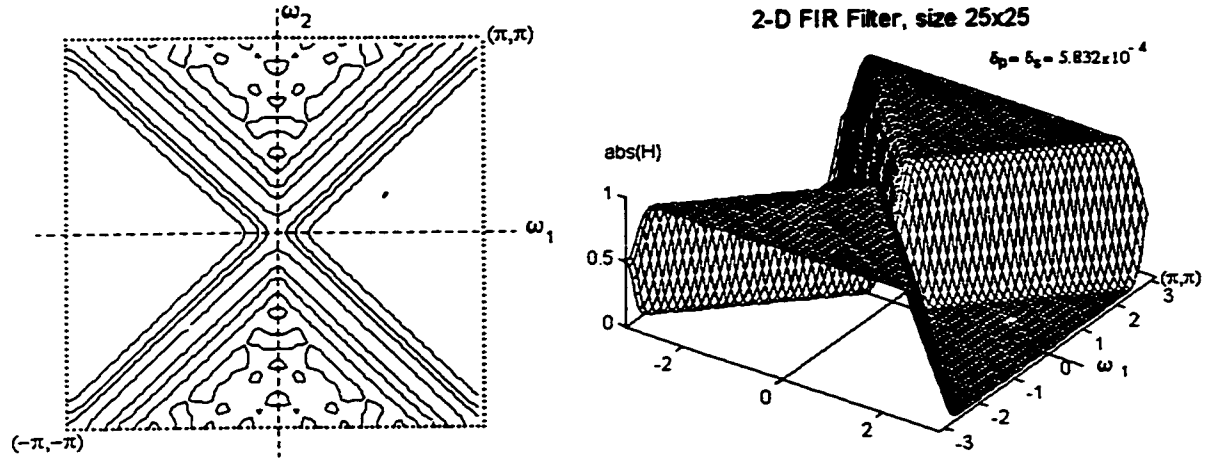


Figure 4.10 A 25×25 90° vertically oriented fan filter obtained from the filter designed in Example 4.4.

Alternatively, this transformation can be done by

$$h_{F\omega_2}(n_1, n_2) = \delta(n_1, n_2) - h_{F\omega_1}(n_1, n_2) \tag{4.2.3}$$

Some of the design results in this section are summarized in Table 4.1. The comparison with other methods for design of 2-D half-band FIR filters is not exhaustive since the

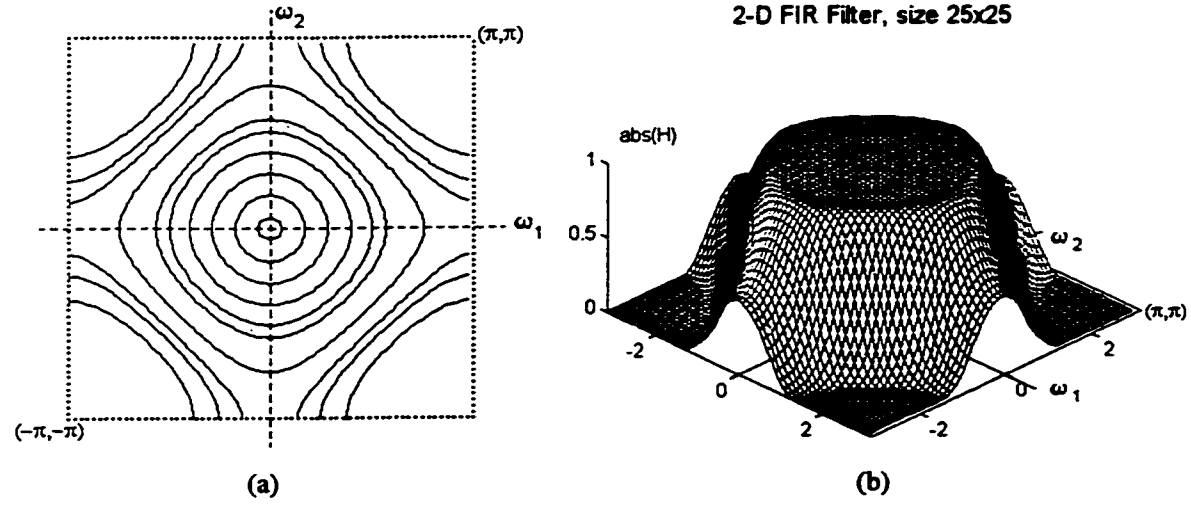
published results are for limited number of filter sizes and transition bands. The proposed here methods are compared mostly with the "standard" frequency transformation method [29] in which a 1-D half-band lowpass optimal FIR filter is transformed to diamond shaped 2-D half-band FIR filter by applying the firs order (3×3) transformation

$$\cos\omega' = 0.5\cos\omega_1 + 0.5\cos\omega_2 \tag{4.2.4}$$

The 3×3 transformation leading to a 2-D 90° fan filter is given by

$$\cos\omega' = 0.5\cos\omega_1 - 0.5\cos\omega_2 \tag{4.2.5}$$

This first order transformations produce curved contours, Figure 4.11, and in result the 2-D frequency edge specifications would fail if they are directly related to 1-D, i.e., $\Delta\omega' = |\omega_s - \omega_p|$. making the 1-D transition band narrower will increase the ripple. In order to obtain more regular shapes with the transformation method, the transformation order must be increased. As a consequence, the 2-D filter size increases considerably. For example if an 11 point 1-D prototype filter is used, the 3×3 transformation produces an 11×11 2-D FIR filter, while a 5×5 transformation will produce a 21×21-point 2-D filter.



(a) (b)

Figure 4.11 A 2-D diamond shaped FIR filter of size 25×25 designed using the frequency transformation given by (4.2.4) and a 25 point 1-D optimal half-band FIR filter with $\omega'_p = 0.38\pi$, $\omega'_s = 0.62\pi$.

Diamond Shaped Filters							
$N \times N$, N_{1D}	Passband edge, ω_p	Stopband edge, ω_s	1-D ripple, δ_{1D}	Passband dev., δ_p	Stopband dev., δ_s	Method (sampling)	Note
9×9, 9	0.36π	0.64π	0.0636	0.0229	0.0218	1-D opt	$H_k = 1$
9×9, –	0.36π	0.64π	-	0.0295	0.0299	Exp	$\alpha = 1.25$
9×9, 9	0.36π	0.64π	0.06362	0.06363	0.27309*	FT	-
9×9, ?	0.36π	0.64π	?	0.0189	0.0184	NDFT [2]	-
11×11, 15	0.4π	0.6π	0.0238	0.0687	0.0677	1-D Opt	$H_k = 1$
11×11, 11	0.4π	0.6π	0.05086	0.05096	0.33513*	FT	-
17×17, 19	0.4π	0.6π	0.0114	0.0020	0.0093	1-D opt	$H_k = 1$
17×17, 17	0.4π	0.6π	0.02376	0.02380	0.30948*	FT	-
25×25, 31	0.38π	0.62π	4.48×10^{-4}	2.8×10^{-4}	1.88×10^{-4}	1-D opt	$H_k = 1$
25×25, –	0.36π	0.64π	-	1.39×10^{-4}	1.04×10^{-4}	Exp	$\alpha = 1.65$
25×25, 25	0.38π	0.62π	0.002395	0.002415	0.17052	FT	–

Table 4.1 (a) Performance comparison for diamond shaped half-band 2-D FIR filters.

90° Fan Filters							
$N \times N$, N_{1D}	Passband edge, ω_p	Stopband edge, ω_s	1-D ripple, δ_{1D}	Passband dev., δ_p	Stopband dev., δ_s	Method (sampling)	α
9×9, –	0.38π	0.62π	-	0.0569	0.0469	Exp	1.9
9×9, 9	0.38π	0.62π	0.08486	0.08486	0.3002*	FT	-
9×9, –	0.35π	0.65π	-	0.0397	0.0397	Z	-
11×11, –	0.38π	0.62π	-	0.03379	0.03379	Exp	0.35
11×11, 11	0.38π	0.62π	0.03319	0.03319	0.25342*	FT	-
17×17, ?	0.43π	0.57π	?	0.0051	0.0051	NDFT [2]	-
19×19, –	0.41π	0.59π	-	0.00843	0.00843	Exp	1.65
19×19, 19	0.41π	0.59π	0.016096	0.016145	0.2761*	FT	-
21×21, –	0.39π	0.61π	-	0.000476	0.001363	Exp	1.85
21×21, 21	0.39π	0.61π	0.00801	0.00803	0.2387*	FT	-
25×25, –	0.42π	0.58π	-	5.83×10^{-4}	5.83×10^{-4}	Exp	1.41
25×25, 25	0.42π	0.58π	0.01269	0.01275	0.309*	FT	-

Table 4.1 (b) Performance comparison for 90° fan half-band 2-D FIR filters (Notations are explained on the next page).

* As in the case of circularly symmetric FIR filters, this deviation is very large because the frequency transformation method was applied with the 2-D frequency edge specifications directly used as frequency band edges for the 1-D optimal prototype. The design should be done by choosing appropriate 1-D pass- and stopband edges, such that the 2-D passband region is completely inside cos $\omega_p' = F(\omega_1,\omega_2)$, and the stopband region is completely outside $\omega_s' = F(\omega_1,\omega_2)$, see Eqs. (4.2.4) and (4.2.5). The deviation in the actual stopband is approximately the same as in the passband.

Exp	Samples on straight lines with exponential distribution in the domain of approximation (proposed sampling technique);
1-D opt	Samples on straight lines passing trough the scaled extremal frequencies obtained from 1-D optimal design (the second Remez exchange algorithm)
FT	Filters designed using the frequency transformations given by Eqs. (4.2.4) and (4.2.5) for diamond and 90° fan filters, respectively;
NDFT [2]	this results are taken from reference [2].
Z	Filters designed with frequency samples taken at the zero locations of the basis function $\{\phi_k(\omega_1, \omega_2)\}$. This method does not always produce good results.

Details of some of the algorithms for determining the frequency sample locations can be seen from the Matlab code listings, presented in appendix A. The program `fan2.m` returns the frequency sample locations and their corresponding values using the extremal frequencies of 1-D optimal filter. The program `fan4.m` uses exponentially distributed parallel lines on which the samples are taken. The filter coefficients are obtained using `fnint2d.m`, also shown in appendix A.

4.3 General diamond and fan FIR filters and other shapes.

In this section the potential of the nonuniform frequency sampling design will be demonstrated. Diamond and fan general (not half-band) FIR filters will be designed, for example a passband diamond shaped FIR filter. 2-D FIR filters with shapes that are usually difficult to be obtained with other design methods are also concerned. Not only the shapes of the designed filters are regular, the performance in terms of passband and stopband error is also good. This designs are performed with one of the two basic sampling techniques proposed in the present work: 1) frequency samples taken on exponentially distributed contours in the approximation domain in the (ω_1, ω_2) plane and 2) frequency samples taken on contours passing trough extremal frequencies translated to 2-D Since these sampling techniques were described and demonstrated in chapters 2, 3, and 4, the designs here are mainly explained with examples.

Example 4.5 In this example a diamond half-band FIR filter is designed, Fig. 4.12, from the 90° fan filter in Example 4.4 by shifting the frequency response by π in the ω_1 direction. This is performed in the spatial domain by

$$h_D(n_1, n_2) = e^{j\pi n_1} h_F(n_1, n_2) = (-1)^{n_1} h_F(n_1, n_2) \tag{4.3.1}$$

where $h_D(n_1, n_2)$ and $h_F(n_1, n_2)$ are the frequency responses of the diamond and fan filters, respectively. After that, a highpass half-band diamond shaped FIR filter is obtained by shifting the frequency response of the fan filter by π in the direction of ω_2:

$$h_{Dhigh}(n_1, n_2) = e^{j\pi n_2} h_F(n_1, n_2) = (-1)^{n_2} h_F(n_1, n_2) \tag{4.3.2}$$

The frequency response magnitude contour and perspective plots of this highpass diamond filter are shown in Fig. 4.13, (a) and (b), respectively. This type of filter can be obtained, of course, by shifting the frequency response of a lowpass diamond filter by π both in the ω_1 and ω_2 directions

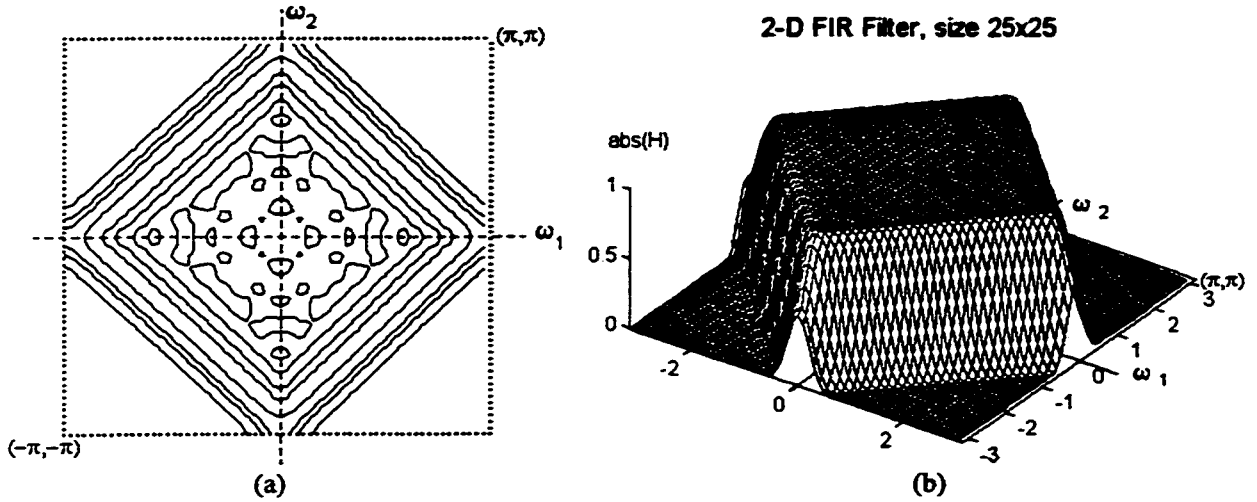


Figure 4.12 A half-band diamond FIR filter obtained from the fan filter in Example 4.4 using (4.3.1).

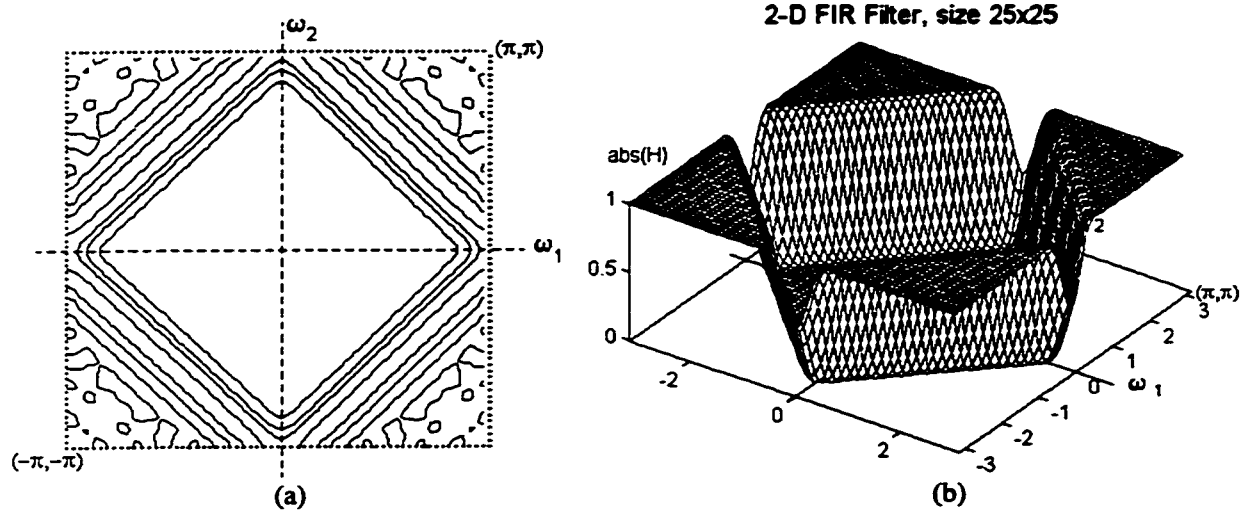


Figure 4.13 A half-band highpass diamond FIR filter obtained from the fan filter in Example 4.4 using Eq. (4.3.2).

A general, not half-band, diamond shaped FIR filter can be designed with nonuniform frequency samples using the relationships from Chapter 3 for circularly shaped filters. Since the frequency response of a general diamond shaped FIR filter possesses eightfold symmetry, the expressions (3.1.6) - (3.1.10) hold and the approximation domain shown in Fig. 3.2 can be used. The interpolation program, `cnint2d.m`, used for circularly shaped filters can also be used without changes. Since the relations (4.1.8) and (4.1.12) do not hold for general diamond and fan filters, now the design cannot be so economical. The number of necessary samples is determined from (3.1.10) instead from (4.1.17).

Example 4.6 A general diamond shaped FIR filter with passband and stopband edge frequencies respectively $\omega_p = 0.25\pi$, $\omega_s = 0.45\pi$. Impulse response region of support size: 17×17 points. The filter is designed from 45 samples taken in the first octant of the (ω_1, ω_2) plane, Fig. 4.14(a). The lines on which the samples are taken pass through the 12 extremal frequencies for a length 21 1-D optimal FIR filter, scaled along the diagonal [(0,0), (π,π)]. The max. deviations are $\delta_p = 0.0166$ and $\delta_s = 0.0329$ for the passband and stopband, respectively.

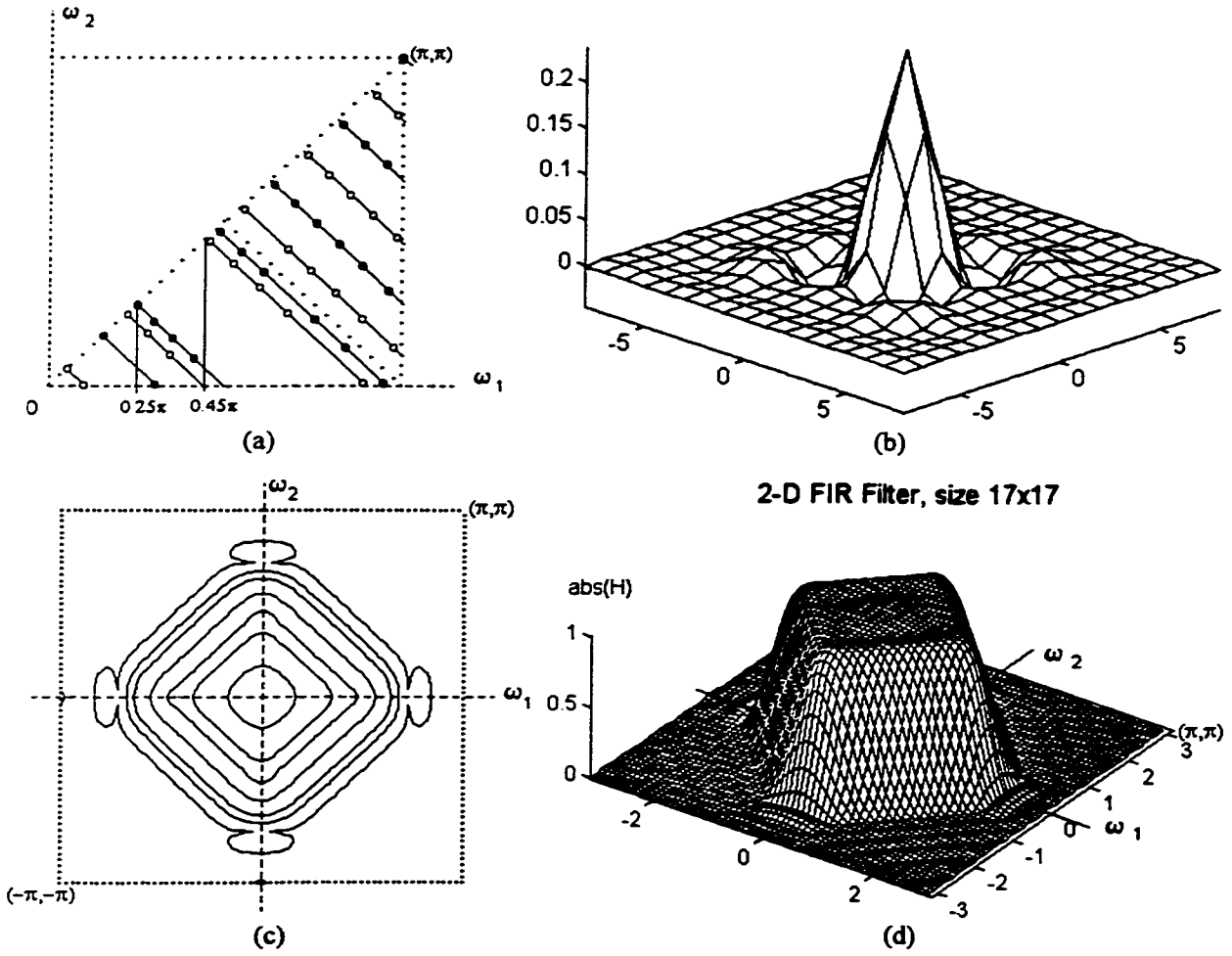


Figure 4.14 Example 4.6: A general diamond shaped FIR filter. (a) Sample locations, o positive sample value: $1+\delta_{1D}$ or $+\delta_{1D}$, • negative sample value: $1-\delta_{1D}$ or $-\delta_{1D}$. $\delta_{1D} = 0.0085$. (b) impulse response; (c) frequency response contour plot; (d) perspective plot of the designed diamond filter frequency response magnitude.

The next example employs the same sampling technique to design a bandpass diamond

shaped zero-phase FIR filter.

Example 4.7 (a) A passband diamond shaped FIR filter with passband and stopband edge frequencies respectively $\omega_{s1} = 0.2\pi$, $\omega_{p1} = 0.4\pi$ $\omega_{p2} = 0.6\pi$, $\omega_{s2} = 0.8\pi$. Impulse response region of support size: 23×23 points. The filter has been designed from 78 frequency samples lying on straight lines with slope −1, Fig. 4.15 (a). These lines pass trough the scaled extremal frequencies obtained with the Remez exchange algorithm (the modified McClellan-Parks-Rabiner program `pmcc.m`) for a 1-D optimal filter of length $N_{1D} = 31$.

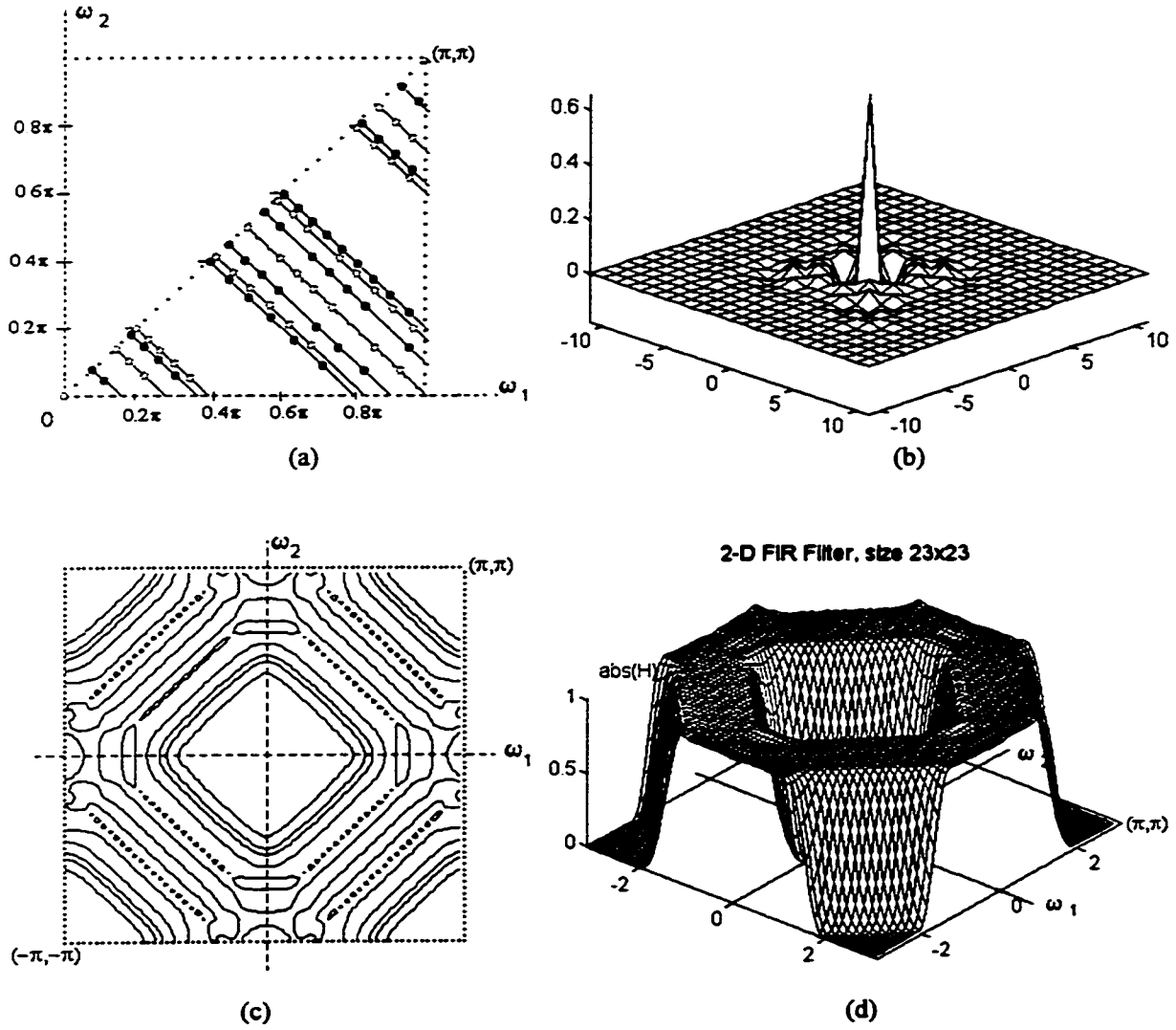


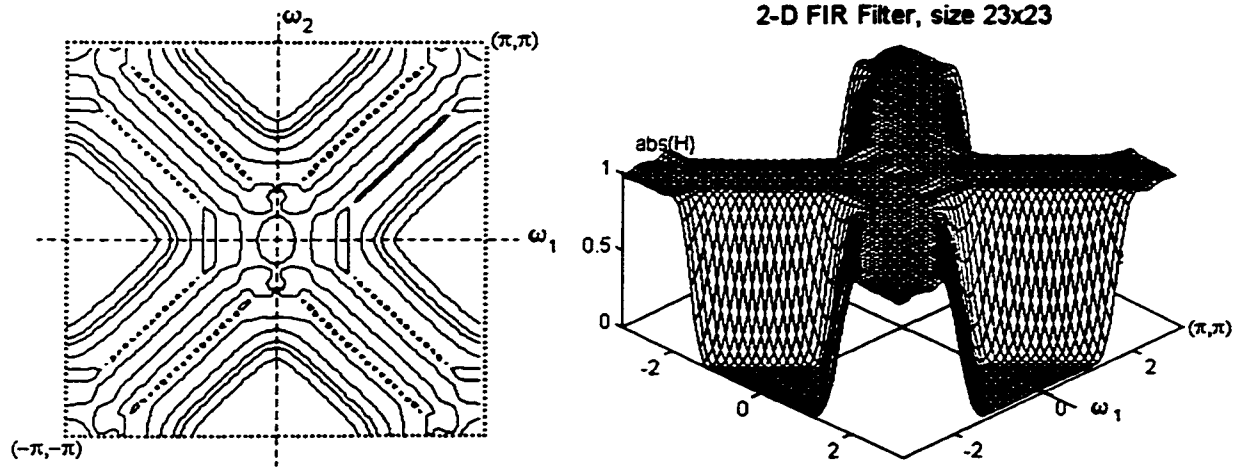
Figure 4.15 Example 4.7(a) A bandpass diamond shaped FIR filter. (a) Sample locations, o positive sample value, $1+\delta_{1D}$ or $+\delta_{1D}$, • negative sample value, $1-\delta_{1D}$ or $-\delta_{1D}$. $\delta_{1D} = 0.0018$. (b) impulse response; (c) contour plot and (d) perspective plot of the designed filter frequency response magnitude.

From this bandpass diamond shaped FIR filter, using shifts of $\pm\pi$ in the direction of ω_1 or ω_2, or by using the passband-stopband conversion

$$h_2(n_1,n_2)=\delta(n_1,n_2)-h_1(n_2,n_1), \tag{4.3.3}$$

the filters shown in Fig. 4.16, 4.17, and 4.18 were obtained. In (4.3.3) $h_1(n_1,n_2)$ is the impulse response of the original filter, $h_2(n_1,n_2)$ is the impulse response of the new filter.

Example 4.7 (b) An "X" shaped zero phase FIR filter is obtained from the bandpass diamond filter by shifting the frequency response by π in the ω_1 direction. (Shifting it in the direction of ω_2 will lead to the same result.) The resulting frequency response contour and perspective plots are shown in Fig. 4.16, (a) and (b), respectively.



(a) (b)

Figure 4.16 An X shaped FIR zero phase filter of size 23×23 points obtained from the filter in Example 4.7 by shifting its frequency response by π in the ω_1 direction.
(a) Frequency response magnitude contour plot; (b) perspective plot.

Example 4.7 (c) An "X" shaped highpass zero phase FIR filter, Fig. 4.17, of size 23×23 points obtained by applying (4.3.3) to the previous filter impulse response.

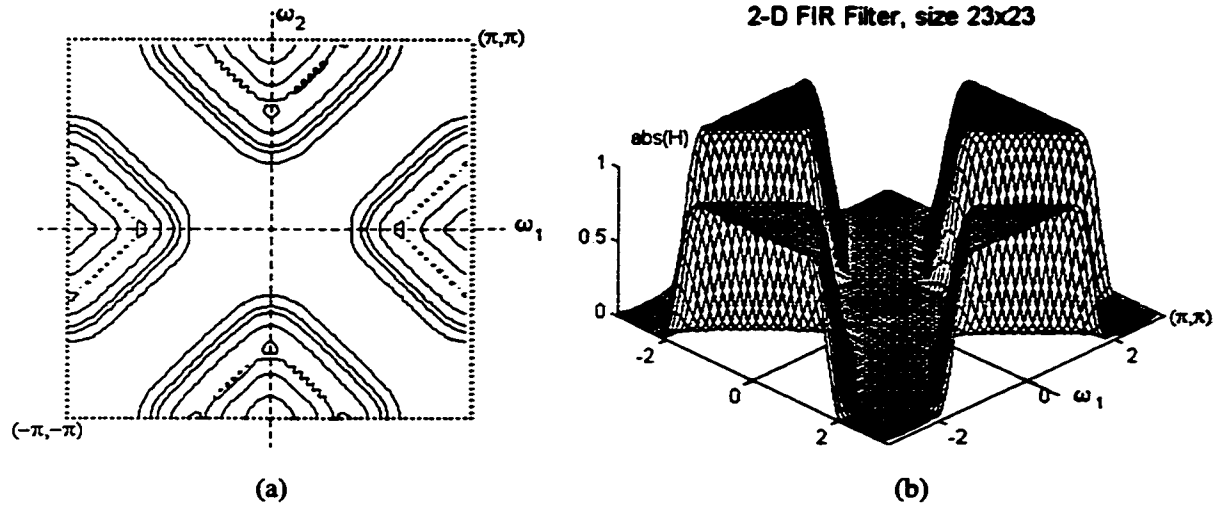


Figure 4.17 An X shaped highpass zero phase FIR filter of size 23×23 points obtained from the filter in Example 4.7 (b) using Eq. (4.3.3). Contour plot (a) and perspective plot (b).

Example 4.7 (d) A 23×23 bandstop diamond shaped zero phase FIR filter obtained by shifting the impulse response of the Example 4.7 (c) by π in the direction of ω_2. (The same result would be obtained if shifting in the direction of ω_1 by π or by applying (4.3.3) to the filter in Example 4.7 (a).)

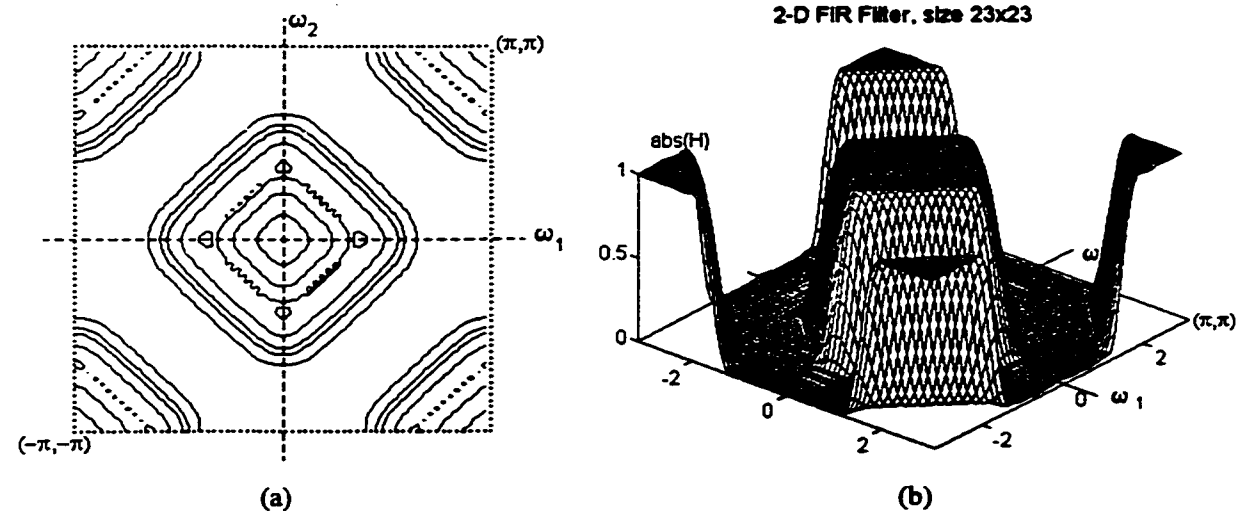
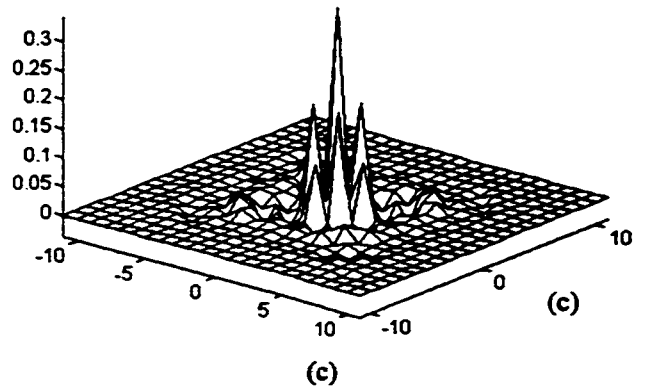


Figure 4.8 Example 4.7 (d). Caption on the next page.



(c)

Figure 4.18 Example 4.7(d) A diamond shaped bandstop zero phase FIR filter of size 23×23 points obtained from the filter in Example 4.7(a) using Eq. (4.3.3).
(a) frequency response contour plot; (b) perspective plot; (c) impulse response.

The last example demonstrates the capabilities of the nonuniform frequency sampling to design a cross shaped 2-D zero phase FIR filters. The frequency edge specifications of such a filter are shown in Fig. 4.19.

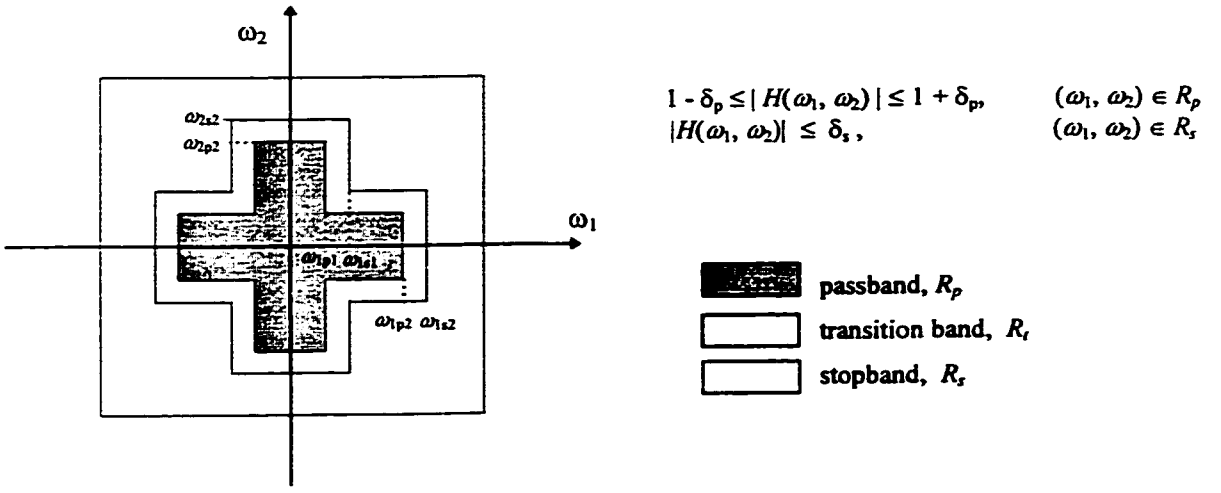


Figure 4.19 Tolerance scheme for a cross shaped filter.

Example 4.8 (a) A cross shaped zero phase FIR filter of size 37×37 points. This filter has been designed using the sampling technique introduced in Subsection. 2.5.2. This is possible because of the filter shape. The frequency edge specifications of this filter are, see the figure above: $\omega_{1p1} = \omega_{2p1} = 0.2\pi$, $\omega_{1s1} = \omega_{2s1} = 0.35\pi$, $\omega_{1p2} = \omega_{2p2} = 0.6\pi$, $\omega_{1s2} = \omega_{2s2} = 0.8\pi$. The frequency samples are taken at the vertices of a nonuniform rectangular grid only in the first frequency plane quadrant. The line coordinates of this grid are obtained by sampling functions of the type of (2.5.10) and (2.5.11). The samples of these functions are mapped using (2.5.12) and (2.5.13) into three bands along the ω_1 and ω_2 axes. The samples of a function like (2.5.12) are mapped to $[0, \omega_{1p1}]$, $[0, \omega_{2p1}]$, $[\omega_{1s1}, \omega_{1p2}]$, and $[\omega_{2s1}, \omega_{2p2}]$. The samples of a functions like (2.5.11) are mapped to the intervals $[\omega_{1s2}, \pi]$ and $[\omega_{2s2}, \pi]$. The sample values in the passband are set to unity and in the stopband to zero. While in the case of a rectangularly shaped FIR filter it is not always necessary to take samples in the transition band, in the case of a cross shaped filter this is imperative in order to obtain good performance. The values of the transition band samples are obtained with linear interpolation of neighboring samples. In the present example there are just one additional row or column of samples in the transition band and, therefore, their values are 0.5. Together with the transition region samples the total number of samples taken is 361. The total number of points in the impulse response is 1369. The frequency sample locations are shown in Fig. 4.20 (a). The impulse response of the designed filter is shown in Fig. 4.20 (b). The filter coefficients were obtained from Eqs. (2.5.4) and (2.2.10) with the Matlab program developed for rectangularly shaped FIR filters, `grd2d.m`, see Appendix A. The frequency response magnitude contour and perspective plots of the designed filter are shown in Fig. 4.20 (c) and (d), respectively. The maximum deviations are $\delta_p = 0.0052$ and $\delta_s = 0.0472$ in the passband and stopband, respectively.

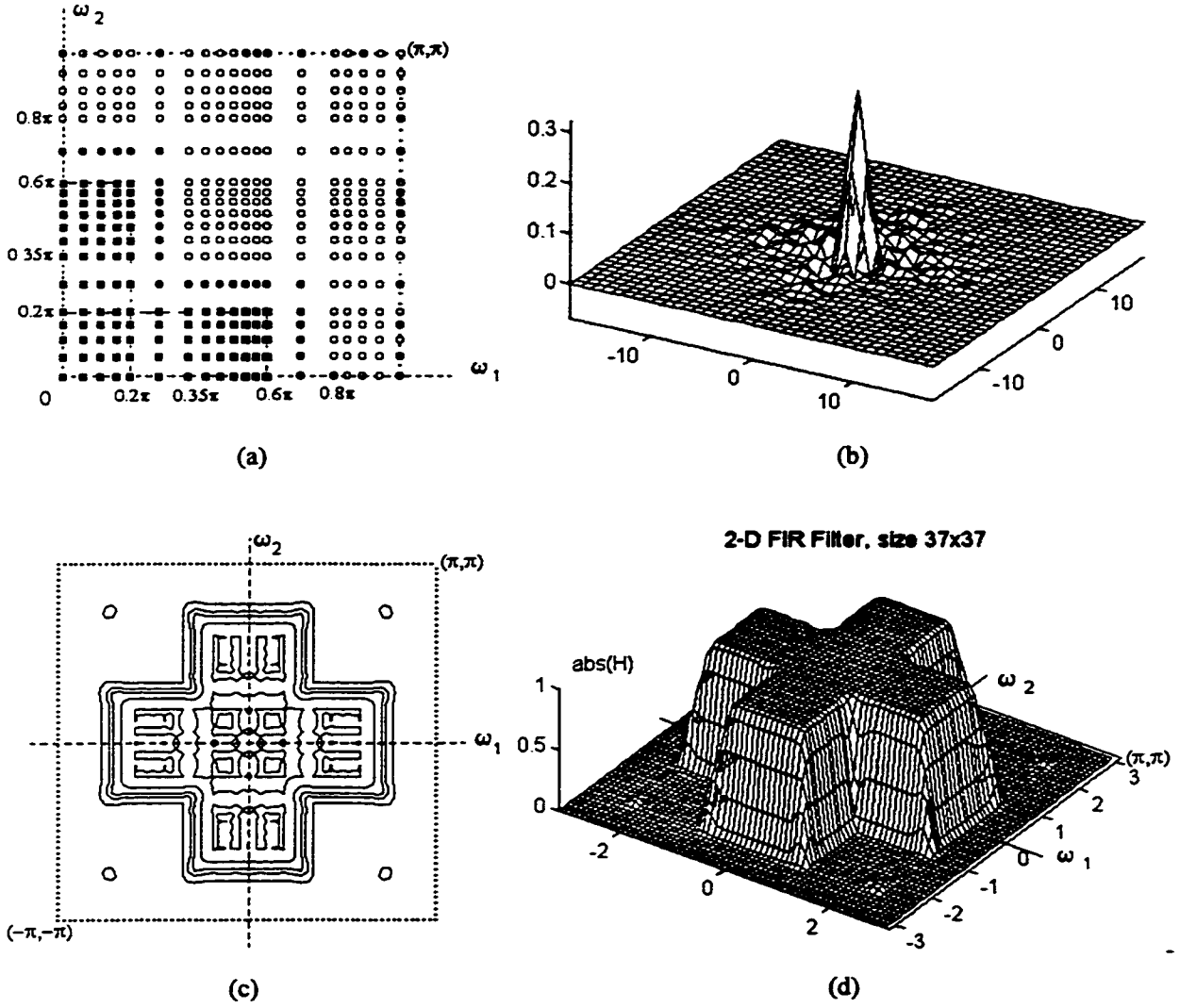


Figure 4.20 A cross shaped zero phase FIR filter of size 37×37 points from Example 4.8(a). (a) Frequency sample locations; (b) impulse response of the designed filter; (c) Frequency response contour plot; (d) frequency response perspective plot.

Example 4.8 (b) Another cross shaped FIR filter of size 53×53 points and frequency edge specifications $\omega_{1p1} = \omega_{2p1} = 0.15\pi$, $\omega_{1s1} = \omega_{2s1} = 0.22\pi$, $\omega_{1p2} = \omega_{2p2} = 0.7\pi$, $\omega_{1s2} = \omega_{2s2} = 0.85\pi$. The cross shaped FIR filters require increased filter order in comparison with the other shapes in order to obtain the same error-level performance. This filter has maximum error $\delta_p = 0.0140$ in the passband and $\delta_s = 0.0076$ in the stopband. The frequency sample locations are shown in Fig. 4.21(a). The impulse response of the designed filter is shown in Fig. 4.21(b), while its frequency response contour and perspective plots are presented in Fig. 4.21(c) and (d), respectively.

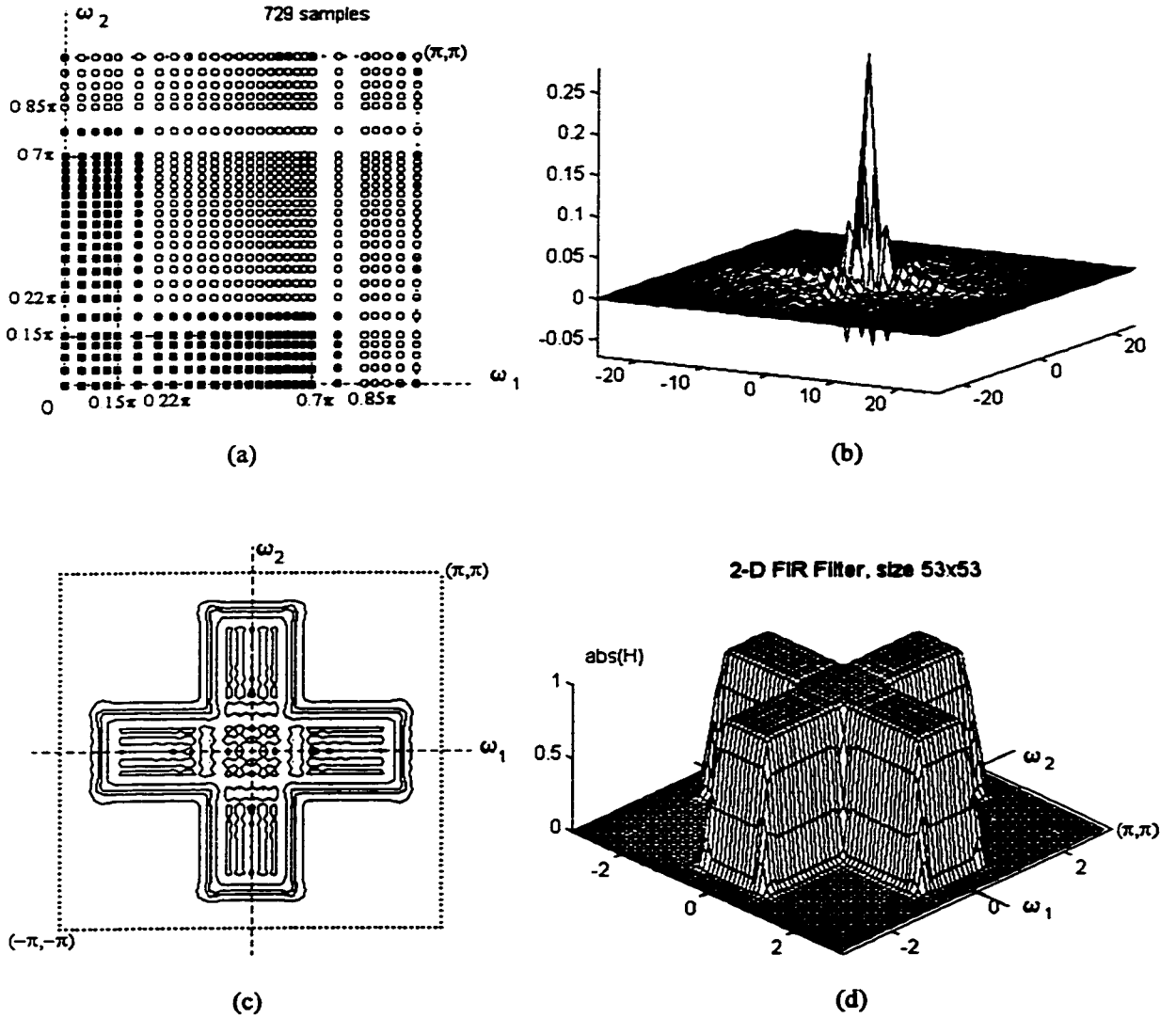


Figure 4.21 A cross shaped zero phase FIR filter of size 53×53 points from Example 4.8(b). (a) Frequency sample locations; (b) impulse response of the designed filter; (c) Frequency response contour plot; (d) frequency response perspective plot.

Some trials have been made to design elliptically shaped zero-phase FIR filters. Since these filters have fourfold symmetric frequency response and impulse response, the frequency samples should be taken on elliptic contours in the first quadrant of the frequency plane. This makes the inverse problem much more difficult than the circular shaped filter case. The results were not satisfactory, partly because of the increased ill-conditioning, partly because of time shortage to improve the sample taking program. That is why no results are presented here.

4.4 Summary and Conclusions.

In this chapter techniques for designing 2-D half-band as diamond, half-band 90° fan, and other shapes FIR filters have been proposed. The techniques are based on the great freedom of the nonuniform frequency sampling and are conceptually simple. Using the existing symmetry properties of 2-D half-band FIR filters, the number of necessary frequency samples is decreased significantly. The frequency samples are taken on parallel lines in the approximation region. These lines are parallel to the shape contours of the ideal (desired) filters. Two main techniques have been considered for determining the line coordinates: 1) scaled extremal frequencies obtained from 1-D optimal filter design and 2) exponential distribution of parallel lines. Better results are obtained when all frequency samples have value of unity than using the amplitudes corresponding to the 1-D extremal frequencies.

A drawback of these techniques is the occurring in some cases ill-conditioning. Also, singularities are not excluded and due measures should be taken to avoid them. Nevertheless, it has been show with several design examples that the proposed techniques can produce half-band diamond and half-band 90° fan zero-phase FIR filters of high

performance in terms of approximation error and shape regularity. Special attention deserves the optimization of parameter α, Eq. (4.2.1) since the maximum pass- and stopband error depends strongly on it. The design simulations showed that there is an optimal value of α for which the maximum error is minimized for this sampling technique and it is the same in the passband and stopband, see the results in Table 4.1 (a) and (b). Very possibly in this case the design is equiripple but that is still to be proved.

At the end of this chapter some of the capabilities of nonuniform frequency sampling with the proposed sampling techniques in chapters 2 and 3 have been demonstrated. Zero phase FIR filters with eightfold and fourfold symmetries of different shapes has been designed. These filters exhibit regular shape and low approximation error. Not all possibilities have been examined.

Chapter 5
Conclusions and Directions for Further Research

5.1 Conclusions

In this thesis several frequency sampling techniques have been proposed that can be used for the design of 2-D zero-phase FIR filters. These techniques are simple conceptually and produce good performance 2-D zero-phase FIR filters with piecewise constant specifications: lowpass, highpass, bandpass, and bandstop. The range of transition bandwidths and filter sizes covers most of the practical application needs. Despite that the design method, including the sampling techniques plus the solution to the 2-D interpolation problem, does not minimize the approximation error, and more precisely any error norm, it makes steps forward to a good approximation of the optimal 2-D FIR filter design. A common feature of all proposed sampling techniques is that the frequency samples are taken on contours that match the contours of the desired filter shape, leading to a high degree of shape regularity in the resulting filters. Also, the filter frequency response, i.e., the bivariate function being approximated using 2-D polynomial interpolation, has been presented as a generalized polynomial -- a linear combination of basis functions. The design is performed by solving for the coefficients of the bivariate polynomial, which are all real numbers. All the proposed approaches use only real arithmetic operations.

In Chapter 2, the design of rectangular shape zero-phase FIR filters with nonuniform samples has been considered and several sampling techniques have been proposed. Using the existing fourfold symmetries in both the frequency response and impulse response,

the number of independent filter coefficients, the number of necessary frequency samples, respectively, is reduced and the frequency samples are taken only in the first quadrant of the frequency plane. The most promising among the proposed techniques appear to be the one taking frequency samples on the vertices of a nonuniform rectangular grid, where the grid line coordinates have exponential or harmonic distribution in the passband(s) and stopband(s). A single parameter controls the grid line distribution and, consequently, the maximum approximation error in the region of approximation. The proposed techniques ensure computational simplicity since the bivariate interpolation problem is divided into two univariate problems which are guaranteed to have a unique solution. The rectangular shape FIR filters are the easiest to design with nonuniform frequency sampling due to the fact that the sampling contours are straight vertical and horizontal lines.

The nonuniform frequency sampling design of zero-phase FIR filters with circular shape has been considered in Chapter 3. The circular shape FIR filters appear to be the most challenging for design with nonuniform frequency sampling. Several sampling techniques have been proposed and investigated. Common for all of them is that the frequency samples lie on circles centered at the origin of the frequency plane with radius changing from zero to π. The rest of approximation domain has been covered using different approaches. The technique giving the best results among the proposed is the one placing the samples on hyperbolic contours in the corner of approximation domain (near the point (π, π)). Using the existing eightfold symmetries in the frequency response of a circularly symmetric FIR filter and in its impulse response respectively, the number of frequency samples is decreased approximately by factor of 4 to 7 compared to the uniform sampling method. The approximation domain in which the frequency samples are taken is the first octant of the frequency plane. The region and arrangement of frequency samples adopted do not allow the problem to be divided into several smaller 1-D problems, as in the rectangularly shaped filter case. As a consequence, the interpolation matrix becomes ill-conditioned for filter sizes larger than approximately 25×25 points and transition bands narrower than approximately 0.1π. In such cases inaccurate results may be obtained. Nevertheless, the proposed techniques produce filters with good

performance: low approximation error and very regular shape for medium and low filter orders and not so narrow transition bands. With the technique, using the extremal frequencies and amplitudes of 1-D optimal design, equiripple behavior is observed around the origin of the frequency plane, for some cases extending to radius of π. It was shown at the end of the third chapter that circular shape filters with narrow transition band (about 0.05π) and increased impulse response support size can be designed with nonuniform frequency sampling using the proposed techniques with the number of samples greater than the number of independent filter coefficients, i.e., by solving an overdetermined system of linear equations. A drawback of this approach is the increased number of arithmetic operations.

In chapter 4, the design of 2-D half-band FIR filters with nonuniform frequency sampling has been considered. Sampling techniques for the design of diamond and fan shaped filters have been proposed. The symmetries and properties of the frequency response and impulse response of 2-D half-band filters allow even smaller number of frequency samples to be taken than the rectangular and circular shape case. The samples are now taken in a small triangular region in the first octant of the frequency plane. The samples lie on straight lines parallel to the isocontours of desired filter magnitude. Two sampling techniques have been proposed. The first of them calculates the locations of the parallel lines using the extremal frequencies of a 1-D optimal filter design. The sample values on each line can have the corresponding values of the 1-D extremal points or are simply set to unity. The other technique uses exponential distribution of parallel lines. All samples have value of unity. The changing density of parallel lines is controlled by supplying a properly chosen parameter. In this way the approximation error can be controlled in a certain degree. It has been show with several design examples that the proposed techniques can produce half-band diamond and half-band 90° fan zero-phase FIR filters of high performance in terms of approximation error and shape regularity. As drawbacks can be pointed the ill-conditioning of the interpolation matrix for many cases, especially for increased filter sizes, and the possible occurrence of singular cases, for whose circumvention adequate measures must be taken. At the end of this chapter, the

capabilities of the nonuniform sampling techniques proposed in this thesis have been demonstrated with the design of zero-phase FIR filters with different shapes.

5.2 Possible directions for further research

For the rectangular shape FIR filters the sampling techniques taking samples at the vertices of a nonuniform rectangular grid are to be preferred since the computational complexity is greatly reduced and the results are comparable with these obtained with samples on parallel lines. The most immediate step that can be taken is the investigation and optimization of the parameters controlling the line density in the cases of exponential and harmonic distribution of the grid lines, [see Eqs. [2.5.5), (2.5.6), (2.5.12), and (2.5.13)]. This can be done by using an iterative procedure, in which the maximum error is minimized with respect to the parameter α, similar to the steepest gradient method. However, the iterative procedures will make the design unnecessary computationally expensive. One of the main ideas of the present work was that the design should avoid complex iterative algorithms. A possible solution is the use of iterative procedures only for tabulation of the controlling parameter α for variety of filter sizes and transition bandwidths. Also, it is believed that separate functions for the passband(s) and stopband(s) will improve further the designed filter performance.

As far as the circular shape FIR filters are concerned, the first further step should be the improvement of the algorithm for the contour shape and locations in the corner region R_C, see Sec. 3.1. The maximum error is almost always into this region and better contour locations can improve significantly the filter performance. The algorithm for the number of samples on each contour should be also improved. Actually, this is not so much of an issue. Much more important is the number of contours and their relative and absolute

locations. A possible way to determine better contour shape in the corner region R_C is to use circular contours in the region R_π and to apply one of the optimal filter techniques, described in [21], for example, on a dense grid only in the corner region R_C. In this way, the critical points in this region will probably outline the contour shape.

Although in many cases where the circular contours pass trough the extremal frequencies of a 1-D optimal filter we have almost equiripple behavior in the region R_π, when the 1-D filter order is greater than about 26, the interpolation matrix condition number increases drastically, since the contours happen to be very "close" to each other. Therefore, another technique should be used to calculate the radii of the concentric contours if a high-order 2-D filter is required with low deviations. Recent experiments have shown that circular contours in R_π exponentially distributed in the passband(s) and stopband(s) combined with hyperbolic contours in R_C give sometimes better result than the proposed technique in Subsection 3.3.3. This technique with exponentially distributed contours in R_π can be further improved and all recommendations made above for the similar technique for rectangular filters are valid here too.

An empirical formula can be found for estimating the size of the circular filter from the given passband and stopband deviations. Also, the order of the 1-D optimal prototype filter should be selected automatically. This won't be a difficult problem. The design simulations have shown that the maximum deviation, usually obtained in the corner region R_C, is 2 to 5 times greater than the 1-D ripple. Therefore, the Matlab signal processing routine `remezord.m` could be used in an appropriate way.

Ways should be found for better solution of the 2-D interpolation problem. A more numerically stable method should be used as, for example, dividing the 2-D problem to several smaller 1-D problems. Interpolating in polar coordinates is also an option, see Marvasty [31].

Most of the directions given for rectangular and circular shape FIR filters hold for the sampling techniques used for 2-D half-band diamond and fan shaped filters. It became clear that the sampling technique using 1-D extremal frequencies is less efficient than the other one using exponentially distributed parallel lines. Therefore, the efforts should be put on the improvement of the second technique. Some iterative procedures can be employed in order to optimize the parameter α controlling the sampling line distributions. Thus α can be tabulated for different filter sizes and for different passband and stopband edge specifications. Further work is needed to determine if this sampling can really produce equiripple filters, as it appears from some design examples. Probably in the case of half-band FIR filters a way can also be found to break the 2-D interpolation problem to several smaller problems and thus improve the numerical stability of the interpolation method.

An efficient sampling technique and a reliable interpolation method for 2-D zero-phase elliptic shape filters are yet to be proposed.

The proposed sampling techniques can be used as an initial step in iterative algorithms for minimax 2-D FIR filter design. Prior to start the multiple exchange procedure, the conditions implied by the characterization theorem for the optimal approximation, Theorem 1.8, Sec. 1.2, must be checked. Using frequency samples obtained with the proposed sampling techniques in this thesis as initial critical point set can significantly reduce the number of iterations and the design time.

References:

[1] S K. Mitra, S. Chakrabarti, and E. Abreu, "Nonuniform Discrete Fourier Transform and its Applications in Signal Processing," in *Proc. EUSIPCO '92, Sixth European Signal Processing Conf.*, vol. 2, Brussels, Belgium, pp. 909-912, Aug. 1992.

[2] Bagchi, S., and S. K. Mitra, "The Nonuniform Discrete Fourier Transform and its Applications in Filter Design," *IEEE Trans. Circuits Syst. II,* vol. 43, pp. 422-444, June 1996.

[3] E. W. Cheney, *Introduction to Approximation Theory*, McGraw-Hill, Inc., 1966.

[4] J. R. Rice, *The Approximation of Functions,* vol. 2, *Nonlinear and Multivariate Theory*, Reading, Mass.: Addison-Wesley, 1969.

[5] J. H. Mathews, *Numerical Methods for Mathematics, Science, and Engineering*, Englewood Cliffs, NJ: Prentice Hall, Inc., 1992.

[6] W. J. Rozwod, C. W. Therrien, and J. S. Lim, "Design of 2-D FIR Filters by Nonuniform Frequency Sampling", *IEEE Trans. Sig. Proc.*, vol. 39, no. 11, pp.2508-2513, Nov. 1991.

[7] E. Angelidis, "A Recursive Frequency-Sampling Method for Designing Zero-Phase FIR Filters by Nonuniform Samples," *IEEE Trans. Sig. Proc.*, vol. 43, no. 6, pp. 1461-1467, June 1995.

[8] E. Angelidis, "A Novel Method for Modeling 2-D FIR Digital Filters in Frequency Domain with Nonuniform Samples," *IEEE Trans. Circuits Syst.* II, vol. 41, no. 7, pp. 482-486, July 1994.

[9] A. Zakhor and G, Alvstad, "Two Dimensional Polynomial Interpolation from Nonuniform Samples," *IEEE Trans. Sig. Proc.*, vol. 40, no.1, pp. 169-180, Jan. 1992.

[10] J. S. Lim, *Two-Dimensional Signal and Image Processing* , Prentice Hall Inc., 1990.

[11] D. E. Dudgeon and R. M. Mersereau, *Multidimensional Digital Signal Processing,* Englewood Cliffs, NJ: Prentice Hall, Inc., 1984

[12] L. R. Rabiner, R. W. Schafer, and C. M. Rader, "The Chirp z-Transform Algorithm," *IEEE Trans. Audio Electroacoust.*, vol. AU-17, pp. 86-92, June 1969.

[13] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, Englewood Cliffs, NJ: Prentice Hall, Inc., 1989

[14] E. Angelidis and J. E. Diamessis, " A Novel Method for Designing FIR Digital Filters with Nonuniform Frequency Samples," *IEEE trans. Sig. Proc.*, vol. 42, no. 2, pp. 259-267, Feb. 1994.

[15] P. A. Regalia and S. K. Mitra, "Kronecker products, unitary matrices, and signal processing applications," *SIAM Review*, vol. 31, pp.586-613, Dec. 1989.

[16] T. W. Parks and C. S. Burrus, *Digital Filter Design*. John Wiley & Sons, Inc., 1987.

[17] J. H. McClellan and T. W. Parks, "A Unified Approach to the Design of Optimum FIR Linear-Phase Digital Filters," *IEEE Trans. Circuit Theory,* vol. CT-20, No. 6, pp. 697-701, Nov. 1973.

[18] J. H. McClellan, T. W. Parks, and L. W. Rabiner, "A computer Program for Designing Optimum FIR Linear Phase Digital Filters," *IEEE Trans. Audio Electroacoust.*, vol. AU-21, No. 6, pp. 506-526, Dec. 1973.

[19] J. G. Fiasconaro, "Two-Dimensional Nonrecursive Filters," in *Picture Processing and Digital Filtering*, T. S. Huang, ed., Springer-Verlag, pp. 43-46, 1975.

[20] J. V. Hu and L. R. Rabiner, "Design Techniques for Two-Dimensional Digital Filters," *IEEE Trans. Audio Electroacoust.*, vol. AU-20, No. 4, pp.249-256, 1972.

[21] D. B. Harris and R. M. Mersereau, "A comparison of algorithms for minimax design of two-dimensional linear phase FIR digital filters," *IEEE Trans. Acoust., Speech and Sig. Proc.*, vol. ASSP-25, pp. 492-500, Dec. 1977.

[22] Y. Kamp and J. P. Thiran, "Chebyshev Approximation for Two-Dimensional Nonrecursive Digital Filters," *IEEE Trans. Circuits Syst.,* vol. CAS-22, No.3, pp. 208-217, March 1975.

[23] M. Gasca and J. I. Maeztu, "On Lagrange and Hermite Interpolation in R^k," *Numerische Mathematik*, 39, 1-14, Springer-Verlag, 1982.

[24] J. H. McClellan, "The Design of Two-Dimensional Digital Filters by Transformations," in *Proc. 7th Annu. Princeton Conf. Inform. Sci. and Syst.*, pp. 247-251, 1973.

[25] T. S. Huang, "Two-Dimensional Windows," *IEEE Trans. Audio Electroacoust.*, vol. AU- 20, No.1, pp. 88-89, March 1972.

[26] F. Mintzer, "On Half-Band, Third-Band, and Nth-Band FIR Filters and Their Design," *IEEE Trans. Acoust., Speech and Sig. Proc.*, vol. ASSP-30, pp. 734-738, Oct. 1982.

[27] T. Yoshida, A. Nishihara, and N. Fujii, "A Design Method of 2-D Maximally Flat Diamond-Shaped Half-Band FIR Filters," *Trans. IEICE*, vol. E 73, pp.901-907, June 1990.

[28] S. A. H. Aly and M. M. Fahny, "Symmetry in Two-Dimensional Rectangularly Sampled Digital Filters," *IEEE Trans. Acoust., Speech and Sig. Proc.*, vol. ASSP-29, pp. 794-805, Aug. 1981.

[29] E. Isaacson and H. Keller, *Analysis of Numerical Methods*, John Wiley & Sons, Inc., 1966.

[30] K. C. Chung and T. H. Yao, "On Lattices Admitting Unique Lagrange Interpolations," *SIAM J. Numer. Analysis*, vol. 14, No. 4, pp. 735-743, Sept. 1977.

[31] F. A. Marvasti, "Extension of Lagrange Interpolation to 2-D Nonuniform Samples in Polar Coordinates," *IEEE Trans. Circuits Syst.*, vol. 37, pp. 567-568, April 1990.

[32] G. H. Golub and C. F. Van Loan, Matrix Computations, *Matrix Computations*, Baltimore, MD: John Hopkins University press. 1989.

[33] G. H. L. M. Heideman and R. N. J. Veldhuis, "A Sampling Model for Two-Dimensional Functions of Finite Circular Extent: an Orthogonal Transform," in *Signal Processing II: Theories and Applications*, Elsevier Sci. Pub., EURASIP, 1983

Additional bibliography used

[A1] S. N. Hazra and M. S. Reddy, "Design of Circularly Symmetric Low-Pass Two-Dimensional FIR Digital Filters Using Transformation," *IEEE Trans. Circuits Syst.*, vol. CAS-33, pp. 1022-1026, Oct. 1986.

[A2] A. Papoulis, *Signal Analysis*, NY: McGraw-Hill, Inc., 1977.

[A3] A. Antoniou, *Digital Filters: Analysis, Design, and Applications*, McGraw-Hill, Inc., 1993.

[A4] P. A. Rattey and A. G. Lindgren, "Sampling the 2-D Radon Transform," *IEEE Trans. Acoust., Speech and Sig. Proc.*, vol. ASSP-29, pp. 994-1002, Oct. 1981.

[A5] J. Little and C. Moler, *MATLAB*, Sherborn, MA: The MathWorks.

[A6] D. S. Chen and J. P. Allebach, "Analysis of Error in Reconstruction of Two-Dimensional Signals from Irregularly Spaced Samples," *IEEE Trans. Acoust., Speech and Sig. Proc.*, vol. ASSP-35, pp. 173-179, Feb. 1987.

[A7] J. J. Clark, M. R. Palmer, and P. d. Lawrence, "A Transformation Method for the Reconstruction of Functions from Nonuniformly Spaced Samples," *IEEE Trans. Acoust., Speech and Sig. Proc.*, vol. ASSP-33, pp. 1151-1165, Oct. 1985.

Appendix A
Program Listings

A1. Programs for Chapter 2: Rectangular Shape FIR Filters

```
function [w1, w12, f12] = tsam2(fp1, fs1, fp2, fs2, M1, M2)

%     [w1, w12, f12] = TSAM2(fp1, fs1, fp2, fs2, M1, M2)
%     Takes frequency samples in the first
%     quadrant of (w1,w2) plane placed on M1 vertical lines.
%     There are M2 samples on each line.
%     The sample coordinates are calculated using EXPONENTIAL
%     functions. The sampling points are stored in
%     vector w1 and matrix w12, and the corresponding
%     sample values -- in matrix f12. In this way
%     TSAM can be used with pnint2d.m to design a rectangular
%     shape 2-D FIR filter.
%
%          See also PNINT1D, PNINT2D, TSAM

%          © Val Ninov, Aug. 1998
%          Last revision: Oct. 1998

alpha = 1.25;

if nargin <5
 M1 = 5; M2 = 5;
end
wp1 = fp1*pi; ws1 = fs1*pi;
if nargin < 3
 wp2 = wp1; ws2 = ws1;
else
 wp2 = fp2*pi; ws2 = fs2*pi;
end

N1 = 2*M1-1; N2 = 2*M2-1;
PBwidth1 = wp1; PBwidth2 = wp2;
SBwidth1 = pi-ws1; SBwidth2 = pi-ws2;
P1 = PBwidth1/(PBwidth1+SBwidth1)*M1;
P1 = round(P1); P1 = max(1,P1);
P2 = PBwidth2/(PBwidth2+SBwidth2)*M2;
P2 = round(P2); P2 = max(1,P2);
S1 = round(M1 - P1); S1 = max(1,S1);
N1 = 2*(P1+S1)-1;
S2 = round(M2 - P2); S2 = max(1,S2);
N2 = 2*(P2+S2)-1;
fprintf('   Filter size %d x %d,  \n', N1, N2);
fprintf(' P1 = %d, S1 = %d, P2 = %d,  S2 = %d\n', P1, S1, P2, S2);
```

```
% Vertical lines coordinates w1
x1 = 0:1/P1:1-1/P1;
g1 = 1 - exp(-alpha*x1);
 if max(g1)==0
  w1p = wp1;
 else
  w1p = wp1*g1/max(g1);
 end
x1 = 0:1/S1:1-1/S1;
g1 = exp(alpha*x1)-1;
 if max(g1)==0 w1s = (ws1+pi)/2;
 else
  w1s = (pi-ws1)*g1/max(g1) + ws1;
 end
w1 =[w1p w1s];

% sample coordinates w12 on each line w1
w12 = zeros(M2,M1); f12 = zeros(M2,M1);
x2 = 0:1/P2:1-1/P2;
g2 = 1 - exp(-alpha*x2);
 if max(g2)==0
  w2p = wp2;
 else
  w2p = wp2*g2/max(g2);
 end
x2 = 0:1/S2:1-1/S2;
g2 = exp(alpha*x2)-1;
 if max(g2)==0 w2s = (ws2+pi)/2;
 else
  w2s = (pi-ws2)*g2/max(g2) + ws2;
 end
w2ss = pi/(2*M2):pi/M2:pi;
for i=1:M1
 if i <= P1
  w12(:,i) = [w2p w2s]';
 else
  w12(:,i) = w2ss';
 end
end

 for i = 1:M1;  for j = 1:M2
    if i<=P1 & j<=P2  f12(j,i) = 1; end
 end; end

% PLOTTING
figure(2); clf reset
hold on
x1=0:0.2:3.6; y1=zeros(size(x1));
x2 = 0:0.1:pi; y2=pi*ones(size(x2));
plot(x1,y1,':w', y1, x1,':w',x2,y2,'.w');
plot(y2,x2,'.w');
ha=text(3.8,0.1,'w');
set(ha,'FontName','Symbol')  %,'FontAngle','italic')
ha=text(4,0,'1'); set(ha,'FontSize',8)
```

```
ha=text(0.1,3.6,'w');
set(ha,'FontName','Symbol')  %,'FontAngle','italic')
ha=text(0.3,3.5,'2'); set(ha,'FontSize',8);
ha=text(3.2,3.2,'(p,p)');
set(ha,'FontName','Symbol','FontSize',10);
ha=text(-.2,-.2,'0');
set(ha,'FontName','Symbol','FontSize',11);
axis equal; axis off;
axis([-0.2 pi+0.5 -0.2 pi+0.5]);
plot([0 wp1],[wp2 wp2],'w-.',[0 ws1],[ws2,ws2],'w-.');
plot([wp1 wp1],[0 wp2],'w-.',[ws1 ws1],[0 ws2],'w-.');

for i=1:M1
  plot([w1(i) w1(i)],[0 pi],'g');
end

xx=[-.1 -.1  .1 .1 ]*0.4;  %
yy=[ .1 -.1 -.1 .1 ]*0.4;  %  Patch size
for j=1:M1
  for i=1:M2
      u=xx+w1(j);
      v=yy+w12(i,j);
         if  f12(i,j)>0.5;
          fill(u,v,'g');
         else     plot(w1(j),w12(i,j),'go','markersize',4);
         end
  end
end

%--------  END of tsam2  --------
```

```
function h = pnint2d(w1,w12,f12)

%     h = PNINT2D(w1,w12,f12) returns 2-D FIR filter coefficients.
%     using nonuniform sampling ponts. w1 is a column vector
%     containing the sampling frequencies along w1 axis. w12 is a
%     matrix each column containing sampling points in w2 direction
%     corresponding to a vertical line at w1(i). f12 is the matrix
%     holding the magnitude samples at points {w1(i), w12(i,j)}.
%
%     See also PNINT1D, TSAM2, ANFIR1

%     Val Ninov, Last Revision Sept. 1998

[M2, M1] = size(f12);
a2d = zeros(M2, M1);
g = zeros(M2,M1);

  for k=1:M1
     V2 = ones(M2,M2);
     for i = 1:M2
       for j = 2:M2
          V2(i,j) = cos((j-1)*w12(i,k));
       end
     end
     g(:,k) = flipud(V2\f12(:,k));
   end

  V1 = ones(M1,M1);
  for i = 1:M1
    for j = 2:M1
     V1(i,j) = cos((j-1)*w1(i));
    end
  end
  for m=1:M2
     a2d(m,:) = (V1\g(m,:)')';
  end
%~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
   a2d = rot90(a2d,-1);
   [n1,n2]=size(a2d);

   a2d(1,2:n2)=0.5*a2d(1,2:n2);
   a2d(2:n1,1)=0.5*a2d(2:n1,1);
   a2d(2:n1,2:n2)=0.25*a2d(2:n1,2:n2);
   h11 = flipud(fliplr(a2d(2:n1,2:n2)));
   h12 = flipud(a2d(2:n1,1:n2));
   h21 = fliplr(a2d(1:n1,2:n2));
   h = [ h11 h12 ;
         h21 a2d];

if ~nargout
   fiplot(h);              % Plot the frequency response
end

%--------  END of pnint2d  --------
```

```
function [w1, w2, Hk] = nexp2g(fp1, fs1, fp2, fs2, M1, M2)

%     [w1, w2, Hk] = NEXP2G(fp1, fs1, fp2, fs2, M1, M2)
%     Takes frequency samples in the first
%     quadrant of (w1,w2) plane placed on a nonuniform grid
%     vertices with M1 x M2 grid lines.
%     The sample coordinates are calculated using EXPONENTIAL %
      functions.
%     The sampling points are stored in
%     matrices w1 and w2, and the corresponding
%     sample values -- in matrix Hk. In this way
%     NEXP2G can be used with GRD2D.m to design a rectangular
%     shape 2-D FIR filter.
%
%          See also GRD2D, SHARM3, SCHEBY3

%          © Val Ninov, Sept. 1998
%          Last revision: Dec. 1998

alpha = 1.25;

if nargin <5  M1 = 5; M2 = 5;  end
wp1 = fp1*pi; ws1 = fs1*pi;
if nargin < 3
 wp2 = wp1; ws2 = ws1;
else
 wp2 = fp2*pi; ws2 = fs2*pi;
end

 w1 = zeros(1,M1);
 w2 = zeros(1,M2);
 Hk = zeros(M1,M2);

N1 = 2*M1-1; N2 = 2*M2-1;
PBwidth1 = wp1; PBwidth2 = wp2;
SBwidth1 = pi-ws1; SBwidth2 = pi-ws2;
P1 = PBwidth1/(PBwidth1+SBwidth1)*M1;
P1 = round(P1); P1 = max(1,P1);
P2 = PBwidth2/(PBwidth2+SBwidth2)*M2;
P2 = round(P2); P2 = max(1,P2);
S1 = round(M1 - P1); S1 = max(1,S1);
N1 = 2*(P1+S1)-1;
S2 = round(M2 - P2); S2 = max(1,S2);
N2 = 2*(P2+S2)-1;
fprintf('   Filter size %d x %d,  \n', N1, N2);
fprintf('   P1 = %d, S1 = %d,     P2 = %d,  S2 = %d\n', P1, S1,
P2, S2);

% Coordinates along w1 axis
x1 = 0:1/P1:1-1/P1;
g1 = 1 - exp(-alpha*x1);
 if max(g1)==0
  w1p = wp1;
 else
```

```
  w1p = wp1*g1/max(g1);
 end
x1 = 0:1/S1:1-1/S1;
g1 = exp(alpha*x1)-1;
 if max(g1)==0 w1s = (ws1+pi)/2;
 else
  w1s = (pi-ws1)*g1/max(g1) + ws1;
 end
w1 =[w1p w1s];

% sample coordinates alon w2 axis
x2 = 0:1/P2:1-1/P2;
g2 = 1 - exp(-alpha*x2);
 if max(g2)==0
  w2p = wp2;
 else
  w2p = wp2*g2/max(g2);
 end
x2 = 0:1/S2:1-1/S2;
g2 = exp(alpha*x2)-1;
  if max(g2)==0 w2s = (ws2+pi)/2;
  else
   w2s = (pi-ws2)*g2/max(g2) + ws2;
  end
w2 = [w2p w2s];

[ww2,ww1]=meshgrid(w2,w1);

Hk(1:P1,1:P2)=ones(P1,P2);

plotsam(ww1,ww2,Hk,fp1,fs1,fp2,fs2);

%--------  END of nexp2g  --------
```

```
function [w1, w2, Hk] = sharm3(fp1,fs1,fp2,fs2,M1,M2);


%     [w1, w2, Hk] = SHARM3(fp1,fs2,fp2,fs2,M1,M2)
%     returns M1*M2 frequency samples
%     taken in the first quadrant of the (w1,w2) plane
%     using harmonic series.
%     The passband edges are fp1*pi and fp2*pi, the stopband
%     edges are ps1*pi and fs2*pi, respectively.
%     The function returns the grid line coordinates in vectors w1
%     and w2, and the frequency samle values in matrix Hk.
%     The desired 2-D LP filter has squared shape.
%     Use GRD2D to design the 2-D zero phase FIR filter.
%
%     See also: GRD2D, NEXP2G

%          © Val Ninov, Sept. 1998
%          Last revision: Dec. 1998



alpha = -0.8;
if nargin <5
 M1 = 5; M2 = 5;
end
wp1 = fp1*pi; ws1 = fs1*pi;
if nargin < 3
 wp2 = wp1; ws2 = ws1;
else
 wp2 = fp2*pi; ws2 = fs2*pi;
end
 w1 = zeros(1,M1);
 w2 = zeros(1,M2);
 Hk = zeros(M1,M2);

N1 = 2*M1-1; N2 = 2*M2-1;
PBwidth1 = wp1; PBwidth2 = wp2;
SBwidth1 = pi-ws1; SBwidth2 = pi-ws2;
P1 = PBwidth1/(PBwidth1+SBwidth1)*M1;
P1 = round(P1); P1 = max(2,P1);
P2 = PBwidth2/(PBwidth2+SBwidth2)*M2;
P2 = round(P2); P2 = max(2,P2);
S1 = round(M1 - P1); S1 = max(2,S1);
S2 = round(M2 - P2); S2 = max(2,S2);
if P1+S1>M1
   if P1>S1 P1=P1-1;
   else S1=S1-1; end
end
if P2+S2>M2
   if P2>S2 P2=P2-1;
   else S2=S2-1; end
end
N1 = 2*(P1+S1)-1;
N2 = 2*(P2+S2)-1;
```

```
fprintf('   Filter size %d x %d,   \n', N1, N2);
fprintf('   P1 = %d, S1 = %d,      P2 = %d,  S2 = %d\n', P1, S1,
P2, S2);

% Coordinates along w1 axis
%~~~~~~~~~~~~~~~~~~~~~~~~~~~
x = zeros(P1,1);
for k=1:P1;
  x(k) = sum((1:k).^alpha)-1;
end
x1 = x*wp1/max(x);
x = zeros(S1,1);
for k=1:S1;
  x(k)= 1 - sum((1:k).^alpha);
end
x2 =   x*(ws1-pi)/min(x) + pi;
w1 = [x1; flipud(x2)];

% Coordinates along w2 axis
%~~~~~~~~~~~~~~~~~~~~~~~~~~~
y = zeros(P2,1);
for k=1:P2;
  y(k) = sum((1:k).^alpha)-1;
end
y1 = y*wp2/max(y);
y = zeros(S2,1);
for k=1:S2;
  y(k)= 1 - sum((1:k).^alpha);
end
y2 =   y*(ws2-pi)/min(y) + pi;
w2 = [y1; flipud(y2)];

Hk(1:P1,1:P2)=ones(P1,P2);

[ww2,ww1]=meshgrid(w2,w1);
plotsam(ww1,ww2,Hk,fp1,fs1,fp2,fs2);


%--------  END of sharm3  --------
```

```
function h = grd2d(w1,w2,H,fp,fs)

%     h = GRD2D(w1,w2,H) returns 2-D FIR filter coefficients.
%     using nonuniform sampling ponts on the vertices of a %
      nonuniform rectangular grid.
%     w1 is a vector containing the sampling frequencies along w1
%     axis. w2 is a vector containing sampling points in w2
%     direction. H is a matrix holding the sample values H(i,j)
%     at points {w1(i), w2(j)}.
%
%     See also  PNINT2D, PNINT1D, NEXP2G

%     Val Ninov, July. 1998
%     Last revision: 20 Dec. 1988

[M1, M2] = size(H);
a = zeros(M1, M2);
g = zeros(M1,M2);

V1 = ones(M1,M1);
V2 = ones(M2,M2);

  for i = 1:M1
      for j = 2:M1
      V1(i,j) = cos((j-1)*w1(i));
      end
  end
  for i = 1:M2
      for j = 2:M2
      V2(i,j) = cos((j-1)*w2(i));
      end
  end
c1= cond(V1); c2= cond(V2);
a = V1\(H/(V2'));

   a(1,2:M2)=0.5*a(1,2:M2);
   a(2:M1,1)=0.5*a(2:M1,1);
   a(2:M1,2:M2)=0.25*a(2:M1,2:M2);
   h11 = flipud(fliplr(a(2:M1,2:M2)));
   h12 = flipud(a(2:M1,1:M2));
   h21 = fliplr(a(1:M1,2:M2));
   h = [ h11 h12 ;
         h21 a];

if ~nargout
   fiplot(h);
end

if nargin == 5
 [rp,rs]=rippler(h,fp,fs,128)
end

%--------  END of grd2d  --------
```

A2. Programs for Chapter 3: Circular Shape FIR Filters

```
function [w1,w2, H] = vcirc3(fp,fs,N1)

%     [w1,w2,Hk]=VCRC3(fp,fs,N1) takes samples placed on circular
%     contours in (w1,w2) plane. These contours
%     pass through the extremal frequencies
%     obtained using the Remez exchange algorithm
%     for optimal 1-D filter of order N1.
%     The obtained samples can be used with CNINT2D
%     to design a circularly symmetric 2-D lowpass FIR filter
%     by nonuniform sampling.
%     The edge frequencies fp and fs are specified in rad/pi units
%     between [0 1], 1 corresponding to pi rad/s (Nyquist).
%
%     See also: VCIRC4, VCIRC7, CNINT2D,

%     © Val Ninov, Aug. 1998
%     Last Revision: Dec.98

W1, H1] = pmcc(N1,[0 fp fs 1],[ 1 1 0 0]);
   W1 = W1*2*pi;
   W1 = W1(:); H1 = H1(:);
end
L = max(size(W1));
if abs(W1(L)-pi)>10*eps
  W1 = [W1;pi]; H1 = [H1; H1(L-1)]; L = L+1;
end

w1 = zeros(10,1);
w2 = zeros(10,1);
H  = zeros(10,1);
 if W1(1)==0
  w1(1) = W1(1);
  w2(1) = W1(1);
  H(1)  = H1(1);
  k=2; o=2;
 else
  k=1; o=1;
 end
  for j = o:L
    phi(:) = pi/4:pi/4/j:pi/2;
     for i=1:j
      w1(k) = W1(j)*sin(phi(i));
      w2(k) = W1(j)*cos(phi(i));
      H(k) = H1(j);
      k =k+1;
     end
     clear phi
  end
N = max(size(w1));
```

```
%PLOTTING
%~~~~~~~~~~~~~~~~
t = 0:pi/100:pi/2;
figure(1); clf reset
hold on
for i=1:L
   plot(W1(i)*sin(t),W1(i)*cos(t),'g');
end

x1=0:0.2:3.6; y1=zeros(size(x1));
x2 = 0:0.1:pi; y2=pi*ones(size(x2));
plot(x1,y1,':w', y1, x1,':w',x2,y2,'.w');
plot(y2,x2,'.w');
ha=text(3.8,0.1,'w');
set(ha,'FontName','Symbol')  %,'FontAngle','italic')
ha=text(4,0,'1'); set(ha,'FontSize',8)
ha=text(0.1,3.6,'w');
set(ha,'FontName','Symbol')  %,'FontAngle','italic')
ha=text(0.3,3.5,'2'); set(ha,'FontSize',8);
ha=text(3.2,3.2,'(p,p)');
set(ha,'FontName','Symbol','FontSize',10);
ha=text(-.2,-.2,'0');
set(ha,'FontName','Symbol','FontSize',11);
axis equal; axis off;
axis([-0.2 pi+0.5 -0.2 pi+0.5]);

for k = 1:N
  if H(k) > 0.5
    if H(k)>1
     plot(w1(k),w2(k),'go','markersize',4)
        plot(w1(k),w2(k),'.','markersize',8)
    else plot(w1(k),w2(k),'m.','markersize',12)
    end
  else
    if H(k)>0
     plot(w1(k),w2(k),'go','markersize',4)
        plot(w1(k),w2(k),'.','markersize',8)
    elseif H(k)==0 plot(w1(k),w2(k),'.','markersize',12)
    else plot(w1(k),w2(k),'m.','markersize',12)
    end
  end
end
title('Sample Locations')
fos
hold off

%--------  END of vcirc3  --------
```

```
function [w1,w2, H] = vcirc4(fp,fs,N1)

%     [w1,w2,Hk]=VCRC4(fp,fs,N1) takes samples placed on circular
%     contours in (w1,w2) plane. These contours
%     pass through the extremal frequencies in
%     obtained using the Remez exchange algorithm
%     for optimal 1-D filter of order N1.
%     The obtained samples can be used with VNINT2D
%     to design a circularly symmetric 2-D lowpass FIR filter
%     by nonuniform sampling.
%     In contrast with VCIRC3, VCIRC4 takes samples in the region
%     (w1^2+w2^2 > pi^2)U(0<=w1<=pi & 0<=w2<=w1)
%     This is obtaimed by scaling (expanding) the frequency axis
%     of the 1-D filter till sqrt(2)*pi
%
%     See also: VCIRC3, VCIRC7, CNINT2D,

%     © Val Ninov, Sept. 1998
%     Last Revision: Dec.  98

  sc = 1.41;
  fp = fp/sc; fs = fs/sc;
  [W1, H1] = pmcc(N1,[0 fp fs 1],[ 1 1 0 0]);
  W1 = W1*2*pi;
  W1 = W1(:); H1 = H1(:);

L = max(size(W1));
W1(2:L) = sc*W1(2:L);
w1 = zeros(10,1);  w2 = zeros(10,1);
H  = zeros(10,1);
if W1(1)==0
 w1(1) = W1(1); w2(1) = W1(1); H(1)  = H1(1);
 k=2; o=2;
else
 k=1; o=1;
end
  for j = o:L
    phi(:) = pi/4:-pi/4/j:0;
     for i=1:j+1
      u = W1(j)*cos(phi(i)); v = W1(j)*sin(phi(i));
      if u < pi
     w1(k)=u; w2(k)=v;  H(k) = H1(j);
        k =k+1;
      end
     end
    clear phi
  end
N = max(size(w1));
for kk=1:30  nn(kk)=sum(1:kk); end
MM = max(find(nn<=N));
if nn(MM) ~=N
 Nd = N-nn(MM);
 fprintf('\n %d samples will be discarded\n', Nd);
 M = length(w1);
```

```
 if fs>0.5  r = fix(M/6);
   else     r = fix(M/3);
 end
 for k=r:3:r+3*Nd-1
   w1 = [w1(1:k-1,1); w1(k+1:M,1)];
   w2 = [w2(1:k-1,1); w2(k+1:M,1)];
   H  = [ H(1:k-1,1);  H(k+1:M,1)];
   M = length(w1);
 end
end
N = max(size(w1));
fprintf(' %d samples\n', N)

%PLOTTING
figure(1); clf; hold on
x1=0:0.2:3.6; y1=zeros(size(x1));
x2 = 0:0.1:pi; y2=pi*ones(size(x2));
plot(x1,y1,':w', y1, x1,':w',x2,y2,'.w');
plot(y2,x2,'.w'); ha=text(3.8,0.1,'w');
set(ha,'FontName','Symbol')  %,'FontAngle','italic')
ha=text(4,0,'1'); set(ha,'FontSize',8)
ha=text(0.1,3.6,'w');
set(ha,'FontName','Symbol')  %,'FontAngle','italic')
ha=text(0.3,3.5,'2'); set(ha,'FontSize',8);
ha=text(3.2,3.2,'(p,p)');
set(ha,'FontName','Symbol','FontSize',10);
ha=text(-.2,-.2,'0');
set(ha,'FontName','Symbol','FontSize',11);
axis equal; axis off;
axis([-0.2 pi+0.5 -0.2 pi+0.5]);
t = 0:pi/100:pi/2;
for i=1:L
   plot(W1(i)*sin(t),W1(i)*cos(t),'g');
end

for k = 1:N
  if H(k) > 0.5
    if H(k)>1
     plot(w1(k),w2(k),'go','markersize',4)
        plot(w1(k),w2(k),'.','markersize',8)
    else plot(w1(k),w2(k),'m.','markersize',12)
    end
 else
    if H(k)>0
     plot(w1(k),w2(k),'go','markersize',4)
        plot(w1(k),w2(k),'.','markersize',8)
    elseif H(k)==0 plot(w1(k),w2(k),'.','markersize',12)
    else plot(w1(k),w2(k),'m.','markersize',12)
    end
  end
end
title('Sample Locations'); fos
hold off
%--------  END of  vcirc4 --------
```

```
function [w1,w2, H] = vcirc7(fp,fs,N1,wts)

%     [w1,w2,Hk]=VCRC7(fp,fs,N1)  takes samples placed on circular
%     contours in (w1,w2) plane. These contours
%     pass through the extremal frequencies in
%     obtained using the Remez exchange algorithm
%     for optimal 1-D filter of order N1.
%     The obtained samples can be used with CNINT2D
%     to design a circularly symmetric 2-D lowpass FIR filter
%     by nonuniform sampling.
%     The edge frequencies fp and fs are specified in rad/pi units
%     between [0 1], 1 corresponding to pi rad/s (Nyquist).
%     VCIRC7 takes samples in the region
%     (w1^2+w2^2 > pi^2)U(0<=w1<=pi & 0<=w2<=w1)
%     These samples lie on hyperbolic contours.
%
%     See also:  VCIRC6, CNINT2D

%     © Val Ninov, Oct. 1998
%     Last Revision: Jan. 1999

 if length(fp)>1 W1=fp; H1=fs; N1=2*(length(fp)-2);
 else
  if nargin ==4
  [W1, H1] = pmcc(N1,[0 fp fs 1],[1 1 0 0],wts);
  else  [W1, H1] = pmcc(N1,[0 fp fs 1],[1 1 0 0]);
  end
  W1 = W1*2*pi;
  W1 = W1(:); H1 = H1(:);
 end
  L = max(size(W1));
if abs(W1(L)-pi)>10*eps
  W1 = [W1;pi]; H1 = [H1; H1(L-1)]; L = L+1;
end
  Sq = sqrt(2);
  d = sum(diff(W1))/length(diff(W1));
  W2 = Sq*pi-0.1:-d:W1(L);      %:(d):Sq*pi;
  W2 = fliplr(W2);
  L2 = max(size(W2));

 w1 = zeros(300,1);
 w2 = zeros(300,1);
 H  = zeros(300,1);
 if W1(1) <= eps
  w1(1) = 0;
  w2(1) = 0;
  H(1)  = H1(1);
  k=2; o=2;
 else
  k=1; o=1;
 end

  for j = o:L
    phi(:) = pi/4:-pi/(pi*(j-1)+5):0;
```

```
      for i=1:1:length(phi)
       w1(k) = W1(j)*cos(phi(i));
       w2(k) = W1(j)*sin(phi(i));
       H(k) = H1(j);
       k =k+1;
      end
     clear phi
   end
   Nbase = max(find(w1));

%  Now take the additional samples
%~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
 for j = 1:L2
   st = 8/(2*pi*(L+j-1)+7);   %if N1>=16 st = 0.8*st; end
   phi(:) = 0:-st:-1;
       wx = W2(j)*cosh(phi)/Sq; wy = W2(j)*sinh(phi)/Sq;
       wxx = wx - wy;
       wyy = wx + wy;
    if j==1&N1<=10 o=2;
    elseif j==1 & N1>10 o=2;
    else o=1;end
     for i=o:length(wxx)
        if wxx(i)<=pi
          w1(k) = wxx(i); w2(k) = wyy(i);
          if rem(j,2) H(k) = H1(L);
           else  H(k) = H1(L-1);
          end
          k =k+1;
        end
     end
    clear phi wx wy
 end
%w1(k) = pi; w2(k) = 1.3; H(k) = H1(L); k=k+1;
%w1(k)=3.092505;  w2(k)=1.423534; H(k) = H1(L); k=k+1;
if N1>=20
 w1(k) = 3.12; w2(k) = 1.3; H(k) = H1(L);
end
N = max(find(w1~=0));
w1 = w1(1:N); w2 = w2(1:N); H = H(1:N);

  for kk=1:30  nn(kk)=sum(1:kk); end
MM = max(find(nn<=N));
remove=0;
if abs(N-nn(MM))<= fix(N1/6)
  Nd = N-nn(MM-1); remove=1;
end
if (nn(MM) ~=N  & abs(N-nn(MM))> fix(N1/6)) | (nn(MM)~=N & N1<12)
  Nd = N-nn(MM); remove=1;
end
if N==nn(MM)&N1<12 | N<10 remove=0;
end
if remove
 fprintf('\n %d samples. %d samples will be discarded\n', N, Nd);
 M = length(w1);
```

```
 p = 3; count = 0;
 r = max(4,fix((M-p*Nd)/3));
 dr = 10*eps;
 for k=r:p:r+p*Nd-1
   w1 = [w1(1:k-1,1); w1(k+1:M,1)];
   w2 = [w2(1:k-1,1); w2(k+1:M,1)];
   H  = [ H(1:k-1,1);  H(k+1:M,1)];
   M = length(w1);
   count = count+1;
 end
end
N = max(size(w1));
 fprintf(' %d samples\n', N);


% PLOTTING
%~~~~~~~~~~~~~~~~~~~~~~
t = 0:pi/100:pi/2;
theta(:) = pi/4:-pi/32:0;
t2 = 0.5:-0.01:-0.5;
c2 = pi*ones(size(t2));
x=[0 4];
y = zeros(size(x));

figure(1); clf
hold on
for i=1:L
   plot(W1(i)*sin(t),W1(i)*cos(t),'g');
end
 for i=1:max(size(theta))
    y = tan(pi/2-theta(i))*x;
    %plot(x,y,'w')
 end

x1=0:0.2:3.6; y1=zeros(size(x1));
x2 = 0:0.1:pi; y2=pi*ones(size(x2));
plot(x1,y1,':w', y1, x1,':w',x2,y2,'.w');
plot(y2,x2,'.w');
ha=text(3.8,0.1,'w');
set(ha,'FontName','Symbol')  %,'FontAngle','italic')
ha=text(4,0,'1'); set(ha,'FontSize',8)
ha=text(0.1,3.6,'w');
set(ha,'FontName','Symbol')  %,'FontAngle','italic')
ha=text(0.3,3.5,'2'); set(ha,'FontSize',8);
ha=text(3.2,3.2,'(p,p)');
set(ha,'FontName','Symbol','FontSize',10);
ha=text(-.2,-.2,'0');
set(ha,'FontName','Symbol','FontSize',11);
axis equal; axis off;
axis([-0.2 pi+0.5 -0.2 pi+0.5]);

for i=1:L
   plot(W1(i)*cos(t),W1(i)*sin(t),'g');
  if i<=L2
```

```
    X = W2(i)*cosh(t2)/Sq; Y = W2(i)*sinh(t2)/Sq;
    plot(X-Y,X+Y,'c');
  end
end
plot([0 pi],[0 pi],'w-')
 %for i=1:max(size(theta))
    % y = tan(theta(i))*x;
    % plot(x,y,'w')
 %end
%plot([0 pi],[pi pi],'r'); plot([pi pi],[0 pi],'r');
%axis('equal'); grid on;
%axis([0 pi+0.2 0 pi+.2]);

for k = 1:N
  if H(k) > 0.5
    if H(k)>1
     plot(w1(k),w2(k),'go','markersize',4)
       plot(w1(k),w2(k),'.','markersize',8)
    else plot(w1(k),w2(k),'m.','markersize',12)
    end
  else
    if H(k)>0
     plot(w1(k),w2(k),'go','markersize',4)
       plot(w1(k),w2(k),'.','markersize',8)
    elseif H(k)==0 plot(w1(k),w2(k),'.','markersize',12)
    else plot(w1(k),w2(k),'m.','markersize',12)
    end
  end
end
title('Sample Locations')
fos
hold off

%----------  END of vcirc7   ---------
```

```
function [h,V,a] = cnint2d(w1, w2, H, fp, fs)

%     h = CNINT2D(w1,w2,H) solves for the coefficients
%     of an eigthfold symmetrical FIR filter (e.g. circular)
%     with nonuniform frequency samples as input arguments.
%     The algorithm is based on the properties
%     of an eightfold symmetric zero-phase FIR
%     filter of size (2*M-1)X(2*M-1). The input arguments are
%     the sample point coordinates stored in vectors w1 and w2,
%     and the sample values in vector H.
%     The pass- and stopband edge frequencies fp and fs
%     in rad/pi) are  optional and if supplied the program
%     returns the maximum pass- and stopband deviations.

%     © Val Ninov, Aug. 1998
%     Last Revision: Nov. 98


w1 = w1(:);
w2 = w2(:);
H = H(:);

for k=1:30  n(k)=sum(1:k); end
len = max(size(w1));
M = max(find(n<=len));
N = n(M);
if N ~= len
  fprintf('With %d samples a %d X %d filter wil be designed\n',...
  len, 2*M-1, 2*M-1);
  fprintf(' Only the first %d samples will be used\n', N);
  fprintf(' If you want %d X %d size take %d more samples\n',...
  2*M+1,2*M+1,n(M+1)-len);
end

a1 = zeros(N,1);
a = zeros(M,M);
V = zeros(N,N);
flops(0);

for k=1:N;
  l = 1;
  for n1 = 0:M-1
    for n2 = 0:n1
        if n1==n2
      V(k,l) = cos(n1*w1(k))*cos(n2*w2(k));
        else
         V(k,l) = cos(n1*w1(k))*cos(n2*w2(k))...
               + cos(n2*w1(k))*cos(n1*w2(k));
        end
        l = l+1;
    end
  end
end
```

```
condV = cond(V)
a1 = V\H(1:N);

 j=1;
 for i=1:M
   j=j+i-1;
   a(i,1:i)=a1(j:j+i-1,1)';
 end

  for i=1:M
   for j=1:M
     if j>i  a(i,j)=a(j,i); end
   end
  end
aa = a;
   a(1,2:M)=0.5*a(1,2:M);
   a(2:M,1)=0.5*a(2:M,1);
   a(2:M,2:M)=0.25*a(2:M,2:M);
   h11 = flipud(fliplr(a(2:M,2:M)));
   h12 = flipud(a(2:M,1:M));
   h21 = fliplr(a(1:M,2:M));
   h = [ h11 h12 ;
         h21 a];

if ~nargout
   fiplot(h);        % Plot the frequency response
end

if nargin == 5
  [dp,ds]=ripplec(h,fp,fs,128)      %  Determine max. deviations
end


%--------------  END of cnint2d   ----------------
```

```
function [rp,rs] = ripplec(h, fp, fs, N)

%     [rp,rs] = RIPPLEC(h, fp, fs) returns the
%     passband and stopband max ripples of a 2-D
%     LP or HP FIR  filter with circular symmetry.
%     h is the filter impulse response, fp and fs are the
%     frequency edges (in radians/pi), fp=wp/pi, fs=ws/pi.
%     N is the 2-D DFT size: N-by-N points.

%     © Val Ninov, Sept. 1998
%     Last Rrevision: Oct. 98


if nargin < 4
   N = 128;
end

H = fft2(h,N,N);
H = abs(H(1:N/2,1:N/2));
high=0;
if H(1,1)<0.5  high =1;  end

Rp = round(N/2*fp + 1);
Rs = round(N/2*fs + 1);

[f1,f2] = meshgrid(1:N/2,1:N/2);
ind = sqrt(f1.^2 + f2.^2);
passb = find(ind <= Rp);
stopb = find(ind > Rs);

if ~high    % Lowpass
 rp = max(max(abs(H(passb))-1));
 rs = max(max(H(stopb)));
else        % Highpass
 rp = max(max(abs(H(stopb))-1));
 rs = max(max(H(passb)));
end

iMaxs = find(H==max(max(H(stopb))));
xMaxs = fix(iMaxs/(N/2)) + 1;
yMaxs = rem(iMaxs, N/2);
if ~yMaxs  yMaxs = N/2; end
yMaxs = yMaxs+1;
w1sMax = ((xMaxs-1)*2*pi/N);
w2sMax = ((yMaxs-1)*2*pi/N);
fprintf('\n Maximum error in the stopband at frequencies:\n');
fprintf('      ( %f ,  %f ) \n', w1sMax, w2sMax);

%---------------   END of ripplec   -------------------
```

A3. Programs for Chapter 4: Diamond and Fan FIR Filters

```
function [w1,w2, H] = diam2(fp,fs,N1,wts)

%     [w1,w2,Hk]=diam1(fp,fs,N1)  takes  frequaency samples
%     in 1/4 of the first quadrant of (w1,w2) plane.
%     The samples lie on parallel lines obtained from 1-D optimal
%     filter design. The obtained samples can be used with DNINT2D
%     to design a 2-D half-band diamond shaped FIR filter
%     by nonuniform sampling. The edge frequencies fp and fs are
%     specified in normalized by pi frequencies between [0 1],
%     1 corresponding to pi rad/s (Nyquist).
%
%     See also:  DNINT2D

%     © Val Ninov, Dec. 1998
%     Last Revision: Feb. 99

if fp+fs~=1
  error('Incorect edge frequencyes for a halfband filter');
end

 if length(fp)>1 W1=fp; H1=fs; N1=2*(length(fp)-2);
 else
  if nargin ==4
  [W1, H1] = pmcc(N1,[0 fp fs 1],[1 1 0 0],wts);
  else  [W1, H1] = pmcc(N1,[0 fp fs 1],[1 1 0 0]);
  end
  W1 = W1*2*pi;
  W1 = W1(:); H1 = H1(:);
 end
  L1 = max(size(W1));
if abs(W1(L1)-pi)>10*eps
  W1 = [W1;pi]; H1 = [H1; H1(L1-1)]; L1 = L1+1;
end
  Sq = sqrt(2);
  L1 = find(W1 == fp*pi);

 w1 = zeros(300,1);
 w2 = zeros(300,1);
 H  = zeros(300,1);
 if W1(1) <= eps
  w1(1) = 0;  w2(1) = 0;  H(1)  = H1(1);
  k=2; o=2;
 else
  k=1; o=1;
 end
  for j = o:L1
      m = tan(pi/4-0.02:-pi/(pi*(j-1)+5):0);
      for i=1:length(m)
        w1(k) = 2*W1(j)/(m(i)+1);
        w2(k) = m(i)*w1(k);
        H(k) = H1(j);
```

```
        k =k+1;
      end
        clear m;
  end

for kk=1:30  n(k)=fix((kk+1)/2)*fix((kk+2)/2); end
Ns = max(find(w1));
M = max(find(n<=Ns));
L = n(M);
if L ~= Ns
  %for j = L1+1:length(W1)
   %      w1(k) = W1(j);  w2(k) = W1(j);
   %      H(k) = H1(j);   k =k+1;
  %end
end

Ns = max(find(w1));
w1 = w1(1:Ns); w2 = w2(1:Ns); H = H(1:Ns);

for kk=1:30  nn(kk)=fix((kk+1)/2)*fix((kk+2)/2); end
 MM = max(find(nn<=Ns));
remove=0;
 if Ns ~= nn(MM)
   Nd = Ns-nn(MM); remove=1;
 end
if remove
  fprintf('\n %d samples. %d samples will be discarded\n', Ns,
Nd);
  q = length(w1);
  p = 3; %count = 0;
  r = max(4,fix((q-p*Nd)/3));
  for k=r:p:r+p*Nd-1
    w1 = [w1(1:k-1,1); w1(k+1:q,1)];
    w2 = [w2(1:k-1,1); w2(k+1:q,1)];
    H  = [ H(1:k-1,1);  H(k+1:q,1)];
    q = length(w1);
    %count = count+1;
  end
end

Ns = max(size(w1));
 fprintf(' %d samples\n', Ns);


%-----------------   END of diam2   -----------------
```

```
function [w1,w2, H] = diam4(fp,fs,N,alpha)

%     [w1,w2,Hk]=diam4(fp,fs,N1)  takes  frequaency samples
%     in 1/4 of the first quadrant of (w1,w2) plane
%     on parallel lines with exponential distribution.
%     The obtained samples can be used with DNINT2D
%     to design a 2-D half-band diamond shaped FIR filter
%     by nonuniform sampling.
%     The edge frequencies fp and fs are specified in normalized
%     by pi frequencies between [0 1], 1 corresponding to pi rad/s
%
%     See also:  DIAM2, DNINT2D

%     © Val Ninov, Jan. 1999
%     Last Revision: Feb.99

if fp+fs~=1
  error('Incorect edge frequencyes for a halfband filter');
end

if nargin<4  alpha = 1.7; end

P = max(2,fix(N/4));
 x = 0:1/P:1-1/P;
 g = 1 - exp(-alpha*x);
W1 = fp*pi*g/max(g);
Sq = sqrt(2);
L1 = length(W1);

 w1 = zeros(300,1); w2 = zeros(300,1);
 if W1(1) <= eps
   w1(1) = 0;  w2(1) = 0;
   k=2; o=2;
   else  k=1; o=1;
 end
 for j = o:L1
     m = tan(pi/4-0.05:-(pi/4-0.06)/(j):0.01);
     for i=1:length(m)
       w1(k) = 2*W1(j)/(m(i)+1);
       w2(k) = m(i)*w1(k);
       k =k+1;
     end
       clear m;
  end
Ns = max(find(w1));
w1 = w1(1:Ns); w2 = w2(1:Ns);
for kk=1:30  nn(kk)=fix((kk+1)/2)*fix((kk+2)/2); end
 MM = max(find(nn<=Ns));
remove=0;
 if Ns ~= nn(MM)
   Nd = Ns-nn(MM); remove=1;
 end
if remove
 fprintf('\n %d samples. %d samples will be discarded\n', Ns, Nd);
```

```
  q = length(w1);  p = 3;
  r = max(3,fix((q-p*Nd)/3));
  for k=r:p:r+p*Nd-1
    w1 = [w1(1:k-1,1); w1(k+1:q,1)];
    w2 = [w2(1:k-1,1); w2(k+1:q,1)];
    q = length(w1);
  end
end
Ns = max(size(w1));
H = ones(Ns,1);
fprintf(' %d samples\n', Ns);

% PLOTTING
%~~~~~~~~~~~~~~~~~~~~~~~
figure(1); clf reset; hold on
x1=0:0.2:3.6; y1=zeros(size(x1));
x2 = 0:0.1:pi; y2=pi*ones(size(x2));
x3 = pi/2:0.1:pi; y3 = -x3+pi;
plot(x1,y1,':w', y1, x1,':w',x2,y2,'.w');
plot(y2,x2,'.w', x2,x2,'.w', x3,y3,'.w');
ha=text(3.8,0.1,'w');
set(ha,'FontName','Symbol')  %,'FontAngle','italic')
ha=text(4,0,'1'); set(ha,'FontSize',8)
ha=text(0.1,3.6,'w');
set(ha,'FontName','Symbol')  %,'FontAngle','italic')
ha=text(0.3,3.5,'2'); set(ha,'FontSize',8);
ha=text(3.2,3.2,'(p,p)');
set(ha,'FontName','Symbol','FontSize',10);
ha=text(-.2,-.2,'0');
set(ha,'FontName','Symbol','FontSize',11);
axis equal; axis off;
axis([-0.2 pi+0.5 -0.2 pi+0.5]);
for i=1:L1
   plot([W1(i) 2*W1(i)],[W1(i) 0],'g');
end
for k = 1:Ns
  if H(k) > 0.5
    if H(k)>1
     plot(w1(k),w2(k),'go','markersize',4)
        plot(w1(k),w2(k),'.','markersize',8)
    else plot(w1(k),w2(k),'m.','markersize',12)
    end
  else
    if H(k)>0
     plot(w1(k),w2(k),'gc','markersize',4)
        plot(w1(k),w2(k),'.','markersize',8)
    elseif H(k)==0 plot(w1(k),w2(k),'.','markersize',12)
    else plot(w1(k),w2(k),'m.','markersize',12)
    end
  end
end
title('Sample Locations'); hold off

%--------------   END of diam4   ----------------
```

```
function [w1,w2, H] = fan2(fp,fs,N1,wts)

%     [w1,w2,Hk]=FAN(fp,fs,N1)  takes  frequaency samples
%     in 1/4 of the first quadrant of (w1,w2) plane.
%     The samples lie on parallel lines passing trough
%     extremal frequencies obtained from a 1-D optimal filter
%     The obtained samples can be used with FNINT2D
%     to design a  2-D fan FIR filter
%     by nonuniform sampling.
%
%     See also:  FNINT2D, DIAM, DNINT2D

%     Val Ninov, Dec. 1988


 if length(fp)>1 W1=fp; H1=fs; N1=2*(length(fp)-2);
 else
  if nargin ==4
  [W1, H1] = pmcc(N1,[0 fp fs 1],[1 1 0 0],wts);
  else  [W1, H1] = pmcc(N1,[0 fp fs 1],[1 1 0 0]);
  end
  W1 = W1*2*pi;
  W1 = W1(:); H1 = H1(:);
 end
  L1 = max(size(W1));
if abs(W1(L1)-pi)>10*eps
  W1 = [W1;pi]; H1 = [H1; H1(L1-1)]; L1 = L1+1;
end
  Sq = sqrt(2);
  L1 = find(W1 == fp*pi);

 w1 = zeros(300,1);
 w2 = zeros(300,1);
 H  = zeros(300,1);
 if W1(1) <= eps
  w1(1) = pi;  w2(1) = 0;
  H(1)  = H1(1);
  k=2; o=2;
 else
  k=1; o=1;
 end
  for j = o:L1
      m = pi-W1(j)+0.02:(W1(j)-0.03)/j:pi-0.01;
      for i=1:length(m)
        w1(k) = m(i);
        w2(k) = w1(k)+2*W1(j)-pi;
        H(k) = H1(j);
        k =k+1;
      end
        clear m;
  end

for kk=1:30  n(k)=fix((kk+1)/2)*fix((kk+2)/2); end
Ns = max(find(w1));
```

```
M = max(find(n<=Ns));
L = n(M);
if L ~= Ns
  %for j = L1+1:length(W1)
   %      w1(k) = W1(j);  w2(k) = W1(j);
   %      H(k) = H1(j);    k =k+1;
  %end
end

Ns = max(find(w1));
w1 = w1(1:Ns); w2 = w2(1:Ns); H = H(1:Ns);


 for kk=1:30  nn(kk)=fix((kk+1)/2)*fix((kk+2)/2); end
 MM = max(find(nn<=Ns));

remove=0;
 if Ns ~= nn(MM)
   Nd = Ns-nn(MM); remove=1;
 end
if remove
  fprintf('\n %d samples. %d samples will be discarded\n', Ns,
Nd);
  q = length(w1);
  p = 3; %count = 0;
  r = max(4,fix((q-p*Nd)/3));
  for k=r:p:r+p*Nd-1
    w1 = [w1(1:k-1,1); w1(k+1:q,1)];
    w2 = [w2(1:k-1,1); w2(k+1:q,1)];
    H  = [ H(1:k-1,1);  H(k+1:q,1)];
    q = length(w1);
    %count = count+1;
  end
end

Ns = max(size(w1));
 fprintf(' %d samples\n', Ns);


%-----------------  END of fan2   ---------------------
```

```
function [w1,w2, H] = fan4(fp,fs,N,alpha)

%     [w1,w2,Hk]=FAN(fp,fs,N1)  takes  frequaency samples
%     in 1/4 of the first quadrant of (w1,w2) plane
%     on parallel lines with exponential distribution.
%     The obtained samples can be used with FNINT2D
%     to design a  2-D 90° fan FIR filter
%     by nonuniform sampling.
%
%     See also:  FNINT2D, DIAM4, DNINT2D

%     © Val Ninov, Dec. 1998
%     Last Rrevision: Feb. 1999

if nargin ~=4
  alpha = 1.65;
end
P = max(2,fix(N/4));
 x = 0:1/P:1-1/P;
 g = 1 - exp(-alpha*x);
 W1 = fp*pi*g/max(g);
 Sq = sqrt(2);
 L1 = length(W1);

 w1 = zeros(210,1);   % MAX filter size 57x57
 w2 = zeros(210,1);
 if W1(1) <= eps
  w1(1) = pi;  w2(1) = 0;
  k=2; o=2;
 else
  k=1; o=1;
 end
  for j = o:L1
      %m = tan(3*pi/4-0.01:-pi/(pi*(j-1)+5):pi/2+0.01);
      m = tan(3*pi/4-0.05:-(pi/4-0.06)/(j):pi/2+0.01);
      for i=1:length(m)
        w1(k) = 2*W1(j)/(m(i)-1)+pi;
        w2(k) = w1(k)+2*W1(j)-pi;
        k =k+1;
      end
        clear m;
  end

Ns = max(find(w1));
w1 = w1(1:Ns); w2 = w2(1:Ns);
 for kk=1:30  nn(kk)=fix((kk+1)/2)*fix((kk+2)/2); end
 MM = max(find(nn<=Ns));
remove=0;
 if Ns ~= nn(MM)
   Nd = Ns-nn(MM); remove=1;
 end
if remove
 fprintf('\n %d samples. %d samples will be discarded\n', Ns, Nd);
 q = length(w1);
```

```
  p = 4;
  r = max(4,fix((q-p*Nd)/2));
  for k=r:p:r+p*Nd-1
    w1 = [w1(1:k-1,1); w1(k+1:q,1)];
    w2 = [w2(1:k-1,1); w2(k+1:q,1)];
    q = length(w1);
  end
end

Ns = max(size(w1)); H = ones(Ns,1);
 fprintf(' %d samples\n', Ns);


% PLOTTING
%~~~~~~~~~~~~~~~~~~~~~~

figure(1); clf reset
hold on
x1=0:0.2:3.6; y1=zeros(size(x1));
x2 = 0:0.1:pi; y2=pi*ones(size(x2));
x3 = pi/2:0.1:pi; y3 = -x3+pi;
plot(x1,y1,':w', y1, x1,':w',x2,y2,'.w');
plot(y2,x2,'.w', x2,x2,'.w', x3,y3,'.w');
ha=text(3.8,0.1,'w');
set(ha,'FontName','Symbol')  %,'FontAngle','italic')
ha=text(4,0,'1'); set(ha,'FontSize',8)
ha=text(0.1,3.6,'w');
set(ha,'FontName','Symbol')  %,'FontAngle','italic')
ha=text(0.3,3.5,'2'); set(ha,'FontSize',8);
ha=text(3.2,3.2,'(p,p)');
set(ha,'FontName','Symbol','FontSize',10);
ha=text(-.2,-.2,'0');
set(ha,'FontName','Symbol','FontSize',11);
axis equal; axis off;
axis([-0.2 pi+0.5 -0.2 pi+0.5]);

for i=1:L1
   plot([pi-W1(i) pi],[W1(i) 2*W1(i)],'g');
end


for k = 1:Ns
     plot(w1(k),w2(k),'go','markersize',4)
       plot(w1(k),w2(k),'.','markersize',8)
end
title('Sample Locations')
fos
hold off

%-----------------   END of fan4  ----------------------
```

```
function h = dnint2d(w1, w2, H, fp, fs)

%     h= DNINT2D(w1,w2,H) solves for the coefficients
%     of an eigthfold symmetrical, DIAMOND shaped FIR filter
%     with nonuniform frequency samples as input arguments.
%     The algorithm is based on the properties
%     of a half-band diamond (eightfold symmetric) zero-phase
%     FIR filter.

%     Val Ninov, Dec. 1998
%     Last revision: Jan. 1999

w1 = w1(:);
w2 = w2(:);
H = H(:);

for k=1:30  n(k)=fix((k+1)/2)*fix((k+2)/2); end
len = max(size(w1));
M = max(find(n<=len));
N = 2*M+1;
L = n(M);
if L ~= len
 fprintf(' With %d samples a %d X %d filter wil be designed\n' ,
len, N, N);
 fprintf(' Only the first %d samples will be used\n', L);
 fprintf(' If you want %d X %d size take %d more samples\n',
N+2,N+2,n(M+1)-len);
end


a1 = zeros(L,1);
a = zeros(M+1,M+1);
V = zeros(L,L);

for  k = 1:L;
  l = 1;
  for n1 = 1:M
    for n2 = 0:n1-1
        if  rem(n1+n2,2)
           V(k,l) = cos(n1*w1(k))*cos(n2*w2(k))...
                  + cos(n2*w1(k))*cos(n1*w2(k)); l=l+1;
        end
    end
  end
end

if size(V)~=[L,L]
  error('Matrix V is calculated in a wrog way !!')
end
cond_V = cond(V)
a1 = V\(H(1:L)-0.5);

a(1,1) = 0.25;   % Guess why?
k=1;
```

```
 for i=2:M+1
  o = rem(i,2)+1;
    for j = o:2:i-1
       a(i,j)=a1(k); k=k+1;
    end
 end

a = a + a';      % answer of the above question

   a(1,2:M+1)=0.5*a(1,2:M+1);
   a(2:M+1,1)=0.5*a(2:M+1,1);
   a(2:M+1,2:M+1)=0.25*a(2:M+1,2:M+1);
   h11 = flipud(fliplr(a(2:M+1,2:M+1)));
   h12 = flipud(a(2:M+1,1:M+1));
   h21 = fliplr(a(1:M+1,2:M+1));
   h = [ h11 h12 ;
         h21 a];

if ~nargout
   fiplot(h);
end

if nargin == 5
 [rp,rs]=rippled(h,fp,fs,128)
end


%------------------------   END of dnint2d  ------------------
```

```
function h = fnint2d(w1, w2, H, fp, fs)

%     h= FNINT2D(w1,w2,H) solves for the coefficients
%     of a FAN shaped 2-D half-band FIR filter
%     with nonuniform frequency samples as input arguments.
%     The algorithm is based on the properties
%     of a half-band FAN (90°) zero-phase FIR filter.

%     Val Ninov, Dec. 1998
%     Last revision: Jan. 1999


w1 = w1(:);
w2 = w2(:);
H = H(:);

for k=1:30  n(k)=fix((k+1)/2)*fix((k+2)/2); end
len = max(size(w1));
M = max(find(n<=len));
N = 2*M+1;
L = n(M);
if L ~= len
 fprintf(' With %d samples a %d X %d filter wil be designed\n' ,
len, N, N);
 fprintf(' Only the first %d samples will be used\n', N);
 fprintf(' If you want %d X %d size take %d more samples\n',
N+2,N+2,n(M+1)-len);
end

a1 = zeros(L,1);
a = zeros(M+1,M+1);
V = zeros(L,L);

for  k = 1:L;
  l = 1;
  for n1 = 1:M
    for n2 = 0:n1-1
        if  rem(n1+n2,2)
           V(k,l) =  cos(n1*w1(k))*cos(n2*w2(k))...
                   - cos(n2*w1(k))*cos(n1*w2(k));
           l=l+1;
        end
    end
  end
end

if size(V)~=[L,L]
  error('Matrix V is calculated in a wrog way !!')
end
cond_V = cond(V)
a1 = V\(H(1:L)-0.5);

k=1;
 for i=2:M+1
```

```
  o = rem(i,2)+1;
    for j = o:2:i-1
        a(i,j)=a1(k); k=k+1;
    end
 end

a = a - a';
a(1,1)=0.5;
   a(1,2:M+1)=0.5*a(1,2:M+1);
   a(2:M+1,1)=0.5*a(2:M+1,1);
   a(2:M+1,2:M+1)=0.25*a(2:M+1,2:M+1);
   h11 = flipud(fliplr(a(2:M+1,2:M+1)));
   h12 = flipud(a(2:M+1,1:M+1));
   h21 = fliplr(a(1:M+1,2:M+1));
   h = [ h11 h12 ;
         h21 a];

if ~nargout
   fiplot(h);
end

if nargin == 5
 [rp,rs]=rippled(h,fp,fs,128)
end

%----------------------   END of fnint2d   ------------
```

```
function [rp,rs] = rippled(h, fp, fs, N)

%     [rp,rs] = RIPPLED(h, fp, fs) returns the
%     passband and stopband max deviations of a 2-D
%     DIAMOND FIR  filter.
%     h is the imp. response, fp and fs are the
%     frequency edges (in radians/pi), fp=wp/pi, fs=ws/pi.
%     N is the 2-D DFT size: N-by-N points.

%     © Val Ninov, Jan. 1999
%     Last Rrevision: Feb. 99


if nargin < 4
   N = 128;
end

H = fft2(h,N,N);
H = abs(H(1:N/2,1:N/2));
high=0;
if H(1,1)<0.5  high =1;  end

Rp = round(N/2*fp + 1);
Rs = round(N/2*fs + 1);
[f1,f2] = meshgrid(1:N/2,1:N/2);
passb = find(f1+f2<=2*Rp);
stopb = find(f1+f2>=2*Rs);

if ~high    % Lowpass
 rp = max(max(abs(H(passb))-1));
 rs = max(max(H(stopb)));
else        % Highpass
 rp = max(max(abs(H(stopb))-1));
 rs = max(max(H(passb)));
end

iMaxs = find(H==max(max(H(stopb))));
xMaxs = fix(iMaxs/(N/2)) + 1;
yMaxs = rem(iMaxs, N/2);
if ~yMaxs  yMaxs = N/2; end
yMaxs = yMaxs+1;
w1sMax = ((xMaxs-1)*2*pi/N);
w2sMax = ((yMaxs-1)*2*pi/N);
fprintf('\n Maximum error in the stopband at frequencies:\n');
fprintf('      ( %f ,  %f ) \n', w1sMax, w2sMax);


%------------------  END of rippled   -------------------
```

Appendix B
Approximation Error Plots

Some plots of the approximation error for the circularly shaped FIR filter designed in Example 3.6, pp. 115-116, are shown. The contour and perspective plots of the designed filter frequency response magnitude are repeated here in Figs. B1 and B2.

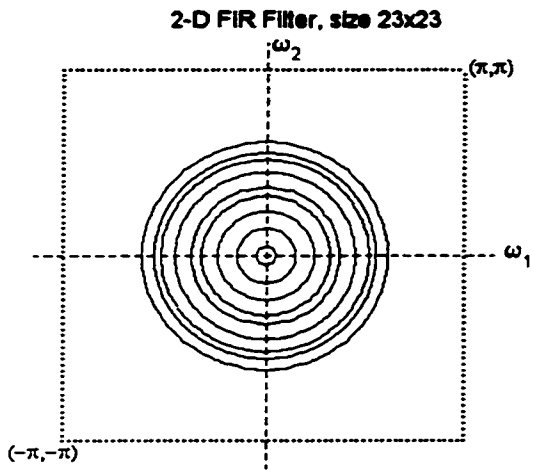


Figure B1

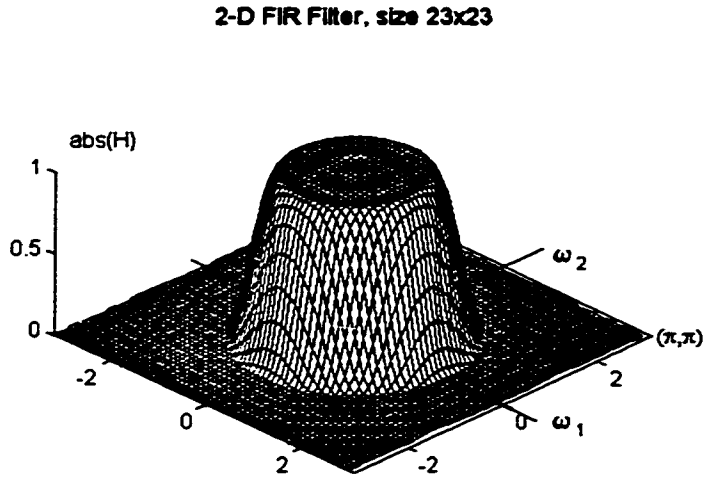


Figure B2

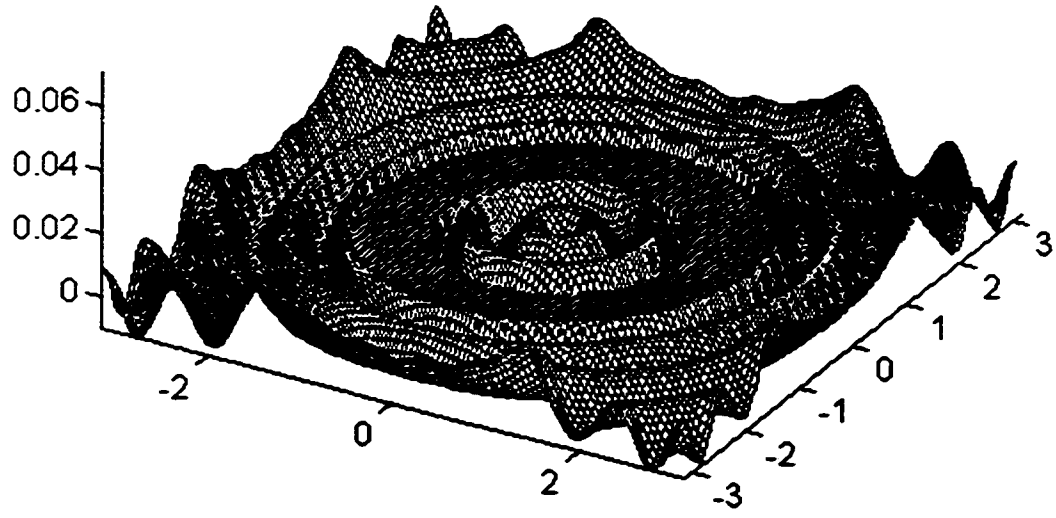


Figure B3

The nearly equiripple behavior and the low 1-D error are maintained in the passband of the designed filter; $\delta_{1D} = 0.0056492$, $\delta_p = 0.0056607$. The perspective error plot is shown in Fig. B3. Details in the passband can be seen from the magnified perspective plot in Fig. B4. Figs. B5 and B6 represent the approximation error in a cross-section along the diagonal $\omega_1 = \omega_2$ and along the ω_1 axis, respectively.

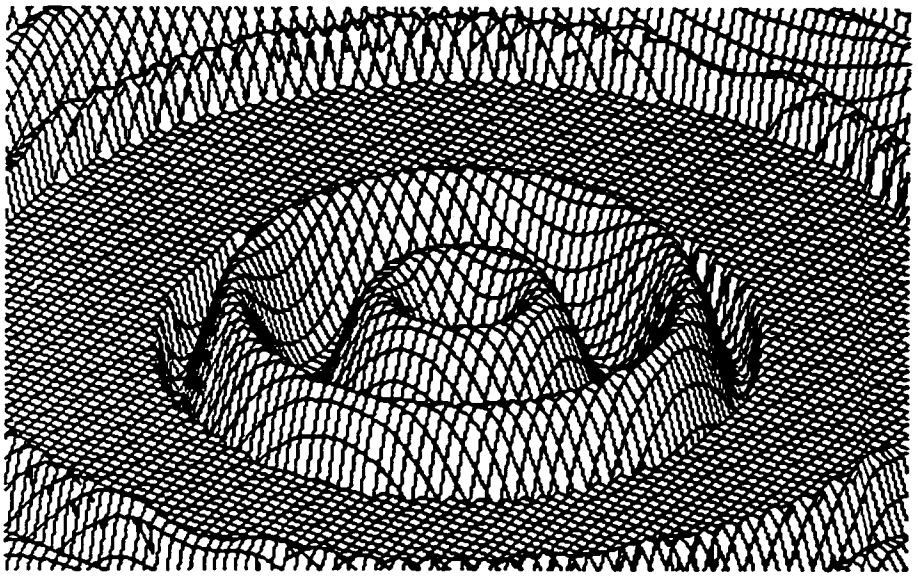


Figure B4

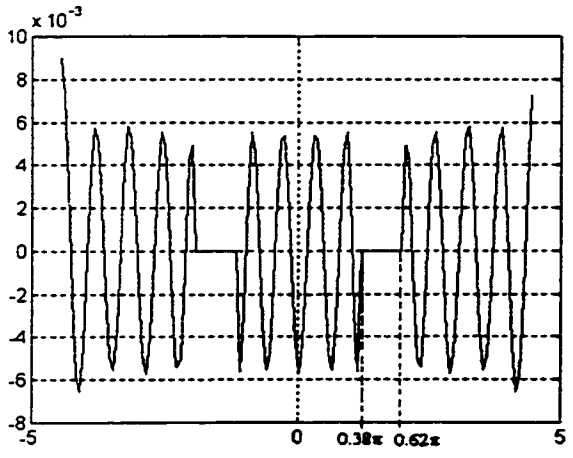


Figure B5

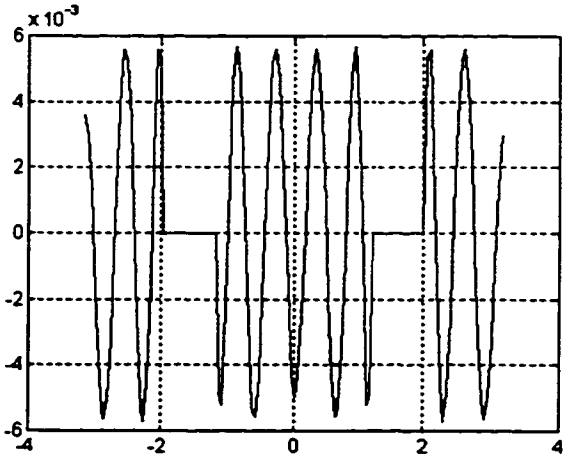


Figure B6

Appendix C

Number of independent impulse response points for circular (eightfold symmetric) and half-band FIR filters

Eightfold Symmetric FIR Filters		2-D Half-band FIR Filters	
Filter size $N \times N$	# independent points $L = \frac{(N+1)(N+3)}{8}$	Filter size $N \times N$	# independent points $L = \left\lfloor \frac{N+1}{4} \right\rfloor \left\lfloor \frac{N+1}{4} \right\rfloor$
3 × 3	3	3 × 3	1
5 × 5	6	5 × 5	2
7 × 7	10	7 × 7	4
9 × 9	15	9 × 9	6
11 × 11	21	11 × 11	9
13 × 13	28	13 × 13	12
15 × 15	36	15 × 15	16
17 × 17	45	17 × 17	20
19 × 19	55	19 × 19	25
21 × 21	66	21 × 21	30
23 × 23	78	23 × 23	36
25 × 25	91	25 × 25	42
27 × 27	105	27 × 27	49
29 × 29	120	29 × 29	56
31 × 31	136	31 × 31	64
33 × 33	153	33 × 33	72
35 × 35	171	35 × 35	81
37 × 37	190	37 × 37	90
39 × 39	210	39 × 39	100
41 × 41	231	41 × 41	110
43 × 43	253	43 × 43	121
45 × 45	276	45 × 45	132
47 × 47	300	47 × 47	144
49 × 49	325	49 × 49	156
51 × 51	351	51 × 51	169
53 × 53	378	53 × 53	182
55 × 55	406	55 × 55	196
57 × 57	435	57 × 57	210
59 × 59	465	59 × 59	225