

Multidimensional Signals and Systems for Image and Video Processing

Eric Dubois

Winter 2012

Chapter 1

Introduction

These notes present the theory of multidimensional (multiD) signals and systems, primarily in the context of image processing. Images are considered to be functions defined on some domain \mathcal{D} with values belonging to a set \mathcal{R} , the range. These values represent the brightness or color of the image at each point in the domain. With this interpretation, an image f is represented as

$$f : \mathcal{D} \rightarrow \mathcal{R} : \mathbf{x} \mapsto f(\mathbf{x}). \quad (1.1)$$

In conventional continuous-time one-dimensional (1D) Signals and Systems theory [Oppenheim 97], \mathcal{D} and \mathcal{R} are both the set of real numbers \mathbb{R} , and so a real 1D signal would be written $f(t)$, $t \in \mathbb{R}$. MultiD signals arise when the domain \mathcal{D} is a space with two or more dimensions. The domain can be continuous, as in the case of real-world still and time-varying images, or discrete, as in the case of sampled images defined on a lattice. In addition, the range \mathcal{R} can also be a higher-dimensional space, for example the three-dimensional color space of human vision.

In these notes, we are mainly concerned with conventional still and time-varying images. A conventional planar image is written $f(x, y)$, where $\mathbf{x} = (x, y)$ lies in a planar region associated with \mathbb{R}^2 . Here, x denotes horizontal spatial position and y is vertical spatial position. The domain \mathcal{D} can be \mathbb{R}^2 itself, or a discrete subset (e.g., a lattice) in the case of sampled images. Similarly, a conventional time-varying image is written $f(x, y, t)$, where $\mathbf{x} = (x, y, t)$ lies in a subset (possibly discrete) of $\mathbb{R}^2 \times \mathbb{R}$,

which may also be written \mathbb{R}^3 . Here, x and y are as above, and t represents time. Higher-dimensional cases also exist, for example time-varying volumetric images with $\mathcal{D} = \mathbb{R}^3 \times \mathbb{R}$. The domain \mathcal{D} can also be a more complicated manifold such as a cylinder or a sphere, as in panoramic imaging. We will examine a few such examples as well.

We will present in detail the relevant theory of multiD signals and systems for both continuous-domain and discrete-domain cases. The relationship between them (sampling and reconstruction) will also be presented. The presentation will rely heavily on the theory of Fourier transforms. The theory will be presented for scalar signals initially, and will then be extended to higher-dimensional signal values once color theory has been introduced.

Chapter 2

Continuous-Domain Signals and Systems

2.1 Introduction

This chapter presents the relevant theory of continuous-domain signals and systems as it applies to still and time-varying images. This is a classical topic, well covered in many texts such as Papoulis's treatise on *Systems and Transforms with Applications in Optics* [Papo 68] and the encyclopedic *Foundations of Image Science* by Barrett and Myers [Barr 04]. The goal of this chapter is to present only the necessary material to understand image acquisition and reconstruction systems, and the relation to discrete-domain signals and systems. Fine points of the theory and vastly more material can be found in the above references.

A continuous-domain planar time-varying image $f(x, y, t)$ is a function of two spatial dimensions x and y and time t , usually observed in a rectangular spatial window \mathcal{W} over some time interval \mathcal{T} . In the case of a still image, $f(x, y, t)$ has a constant value for each (x, y) , independently of t . In this case, we usually suppress the time variable, and write $f(x, y)$. We use a vector notation $f(\mathbf{x})$ to simplify notation and handle two and three-dimensional (and higher-dimensional) cases simultaneously. Thus \mathbf{x} is understood to mean (x, y) in the two-dimensional case and

(x, y, t) in the three-dimensional case. We will denote $\mathbb{R}^2 = \{(x, y) \mid x, y \in \mathbb{R}\}$ and $\mathbb{R}^3 = \{(x, y, t) \mid x, y, t \in \mathbb{R}\}$, where \mathbb{R} is the set of real numbers. To cover both cases, we write \mathbb{R}^D , where normally $D = 2$ or $D = 3$; also the one-dimensional case is covered with $D = 1$ and most results apply for dimensions higher than 3 as well. It is often convenient to express the independent variables as a column matrix, i.e.,

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix} \quad \text{or} \quad \mathbf{x} = \begin{bmatrix} x \\ y \\ t \end{bmatrix}.$$

Since there is no essential difference between \mathbb{R}^D and the space of $D \times 1$ column matrices, we do not distinguish between these different representations. We will often abbreviate two-dimensional as 2D and three-dimensional as 3D.

The spatial window is of dimension $\text{ph} \times \text{pw}$ where pw is the picture width and ph is the picture height. Since the absolute physical size of an image depends on the sensor or display device used, we often choose to adopt the picture height (ph) as the basic unit of spatial distance, as is common in the broadcast video industry. However, we are free to choose any convenient unit of length in a given application, for example, the size of a sensor or display element, or an absolute measure of distance such as the meter or micron. The ratio $ar = \text{pw}/\text{ph}$ is called the aspect ratio, the most common values being 4/3 for standard TV and 16/9 for HDTV. With this notation, $1\text{pw} = ar \text{ ph}$ (see Fig. 2.1). Time is measured in seconds, denoted s. Examples of continuous space-time images include the illumination on the sensor of a video camera, or the luminance of the light reflected by a cinema screen or emitted by a television display.

Since the image is undefined outside the spatial window \mathcal{W} , we are free to extend it outside the window as we see fit to include all of \mathbb{R}^D as the domain. Some possibilities are to set the image to zero outside \mathcal{W} , to periodically repeat the image, or to extrapolate it in some way. Which of these is chosen depends on the application.

There are two common ways to attach an xy coordinate system to the image window, involving the location of the origin and the orientation of the x and y axes, as shown in Fig. 2.2. The standard orientation used in mathematics would place

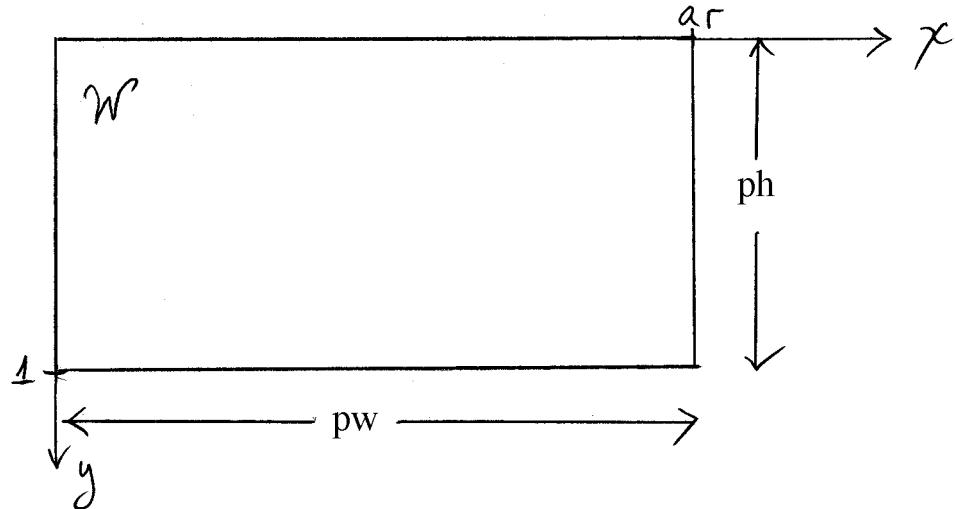


Figure 2.1: Illustration of image window \mathcal{W} with aspect ratio ar : $1 \text{ pw} = ar \text{ ph}$.

the origin at the lower left corner of the image with the y -axis pointing upward. However, because traditionally images have been scanned from top to bottom, most image file formats store the image line-by-line, with the top line first, and line numbers increasing from top to bottom of the image. This makes the orientation shown in Fig. 2.2(b) more convenient, with the origin in the upper left corner of the image and the y -axis pointing downward. For this reason, we will use the orientation of Fig. 2.2(b) here.

2.2 Special multidimensional signals

In addition to naturally occurring continuous space-time images, many analytically defined two and three-dimensional functions are useful in image processing theory. A few of these are introduced here.

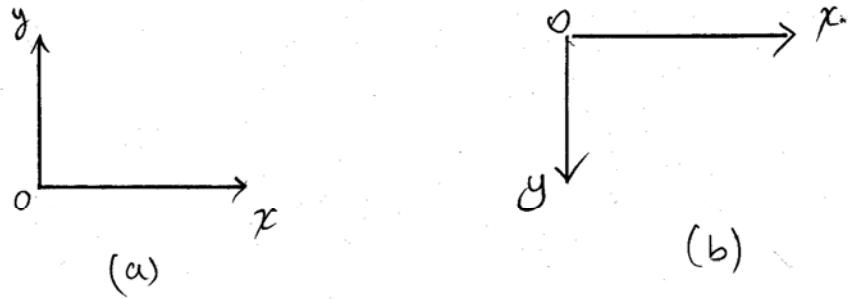


Figure 2.2: Orientation of xy axes. (a) Common bottom-to-top orientation in mathematics. (b) Scanning-based top-to-bottom orientation.

2.2.1 Zero-one functions

Let \mathcal{A} be a region in the two or three-dimensional space, $\mathcal{A} \subset \mathbb{R}^D$. We define the *zero-one function* $p_{\mathcal{A}}(\mathbf{x})$ as illustrated in Fig. 2.3(a) by

$$p_{\mathcal{A}}(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} \in \mathcal{A}; \\ 0, & \text{otherwise.} \end{cases} \quad (2.1)$$

Sometimes, $p_{\mathcal{A}}(\mathbf{x})$ is called the indicator function of the region \mathcal{A} . Different functions are obtained with different choices of the region \mathcal{A} . The most commonly used ones in image processing are the rect and the circ functions in two dimensions. Specifically, for a unit-square region \mathcal{A} we obtain (Fig. 2.3(b))

$$\text{rect}(x, y) = \begin{cases} 1, & \text{if } |x| \leq 0.5 \text{ and } |y| \leq 0.5; \\ 0, & \text{otherwise.} \end{cases} \quad (2.2)$$

For a circular region of unit radius we have (Fig. 2.3(c))

$$\text{circ}(x, y) = \begin{cases} 1, & \text{if } x^2 + y^2 \leq 1; \\ 0, & \text{otherwise.} \end{cases} \quad (2.3)$$

These signals can be extended to the three-dimensional case in a straightforward fashion, and the single notation $\text{rect}(\mathbf{x})$ or $\text{circ}(\mathbf{x})$ can be used to cover all cases.

We will see later how these basic signals can be shifted, scaled, rotated or otherwise transformed it generate a much richer set of zero-one functions. Other zero-one functions that we will encounter correspond to other polygonal regions such as triangles, hexagons, octagons, etc.

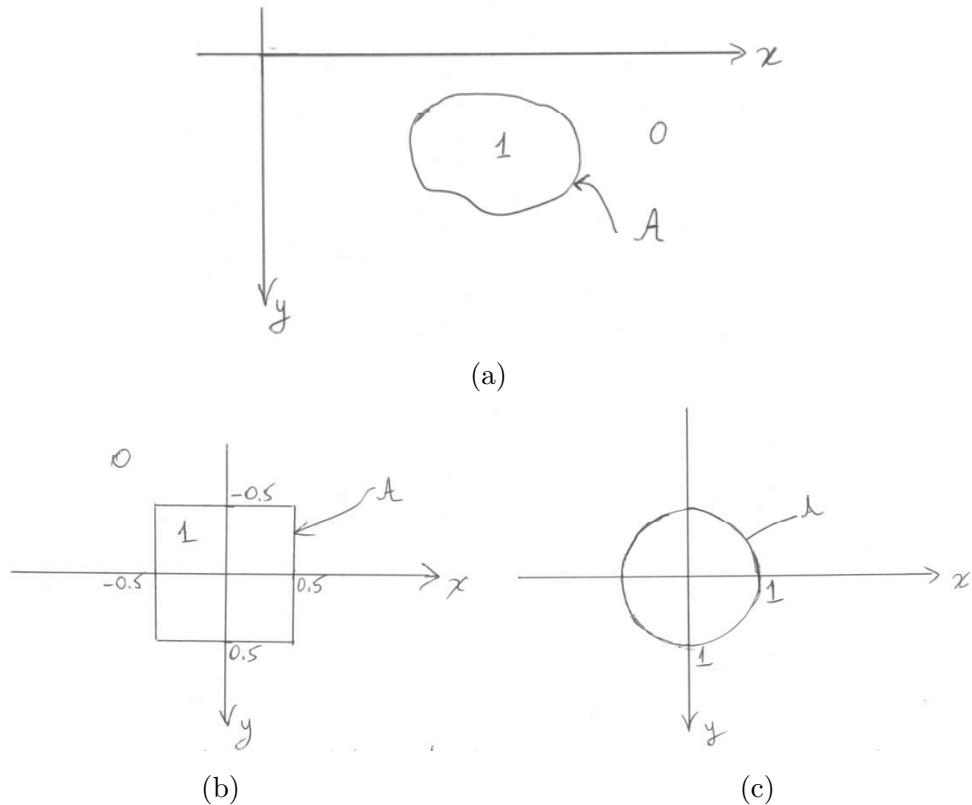


Figure 2.3: Zero-one functions. (a) General zero-one function. (b) Rect function. (c) Circ function.

2.2.2 Sinusoidal signals

As in one-dimensional signals and systems, sinusoidal signals play an important role in the analysis of image processing systems. A general, real two-dimensional signal

with spatial frequency (u, v) is given by

$$f(x, y) = A \cos(2\pi(ux + vy) + \phi) \quad (2.4)$$

where x and y are spatial coordinates in ph, u and v are *fixed* spatial frequencies in units of c/ph (cycles per picture height) and ϕ is the phase. Fig. 2.4 illustrates a typical sinusoidal signal with horizontal frequency 1.5 c/ph and vertical frequency 2.5 c/ph. From this figure, we can identify a number of features of the spatial sinusoidal signal. The sinusoidal signal is periodic in both the horizontal and vertical directions, with horizontal period $1/|u|$ and vertical period $1/|v|$. The signal $f(x, y)$ is constant if $ux + vy$ is constant, i.e., along lines parallel to the line $ux + vy = 0$.

The one-dimensional signal along any line through the origin is a sinusoidal function of distance along the line. The maximum frequency is $\sqrt{u^2 + v^2}$, along the line $y = (v/u)x$, as illustrated in Fig. 2.4. The proof is left as an exercise.

As in one dimension, the complex exponential sinusoidal signals play an important role, e.g., in Fourier analysis. The corresponding complex exponential is

$$f(x, y) = C \exp(j2\pi(ux + vy)) \quad (2.5)$$

where $C = Ae^{j\phi}$ can be complex. We often will adopt the vector notation

$$f(\mathbf{x}) = C \exp(j2\pi \mathbf{u} \cdot \mathbf{x}) \quad (2.6)$$

where \mathbf{x} denotes (x, y) as before, \mathbf{u} denotes (u, v) and $\mathbf{u} \cdot \mathbf{x}$ denotes $ux + vy$. When using the column matrix notation, then $\mathbf{u} = [u \ v]^T$ and we can write $f(\mathbf{x}) = C \exp(j2\pi \mathbf{u}^T \mathbf{x})$. Using Euler's formula, the complex exponential can be written in terms of real sinusoidal signals

$$f(x, y) = A \cos(2\pi(ux + vy) + \phi) + jA \sin(2\pi(ux + vy) + \phi). \quad (2.7)$$

2.2.3 Real exponential functions

First order exponential signals are given by

$$f(x, y) = A \exp(-(a|x| + b|y|)) \quad (2.8)$$

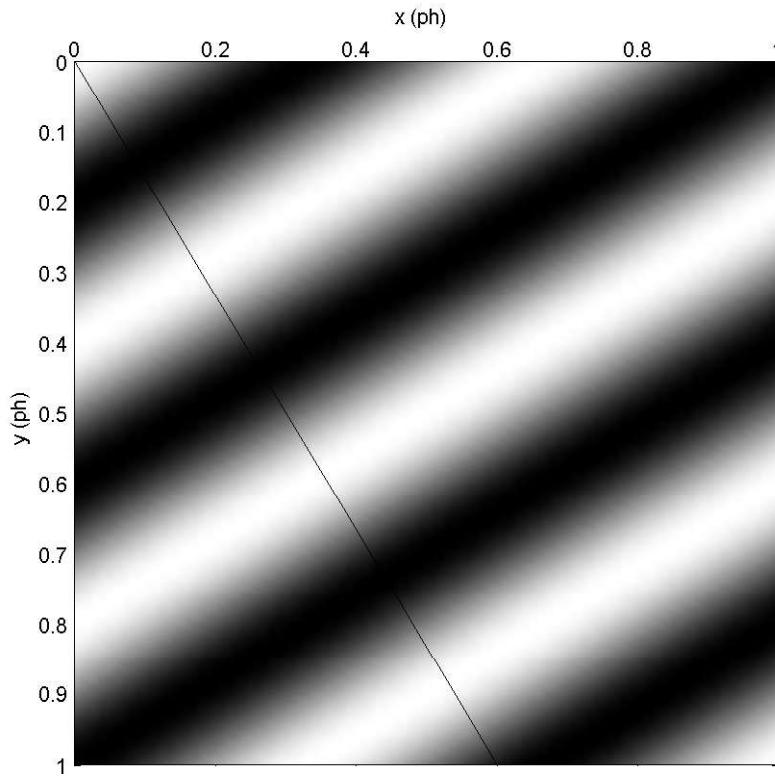


Figure 2.4: Sinusoidal signal with $u = 1.5$ c/ph and $v = 2.5$ c/ph. The horizontal period is $(2/3)$ ph and the vertical period is .4 ph. The frequency along the line $y = \frac{5}{3}x$ is 2.9 c/ph, corresponding to a period of 0.34 ph.

and second-order, or Gaussian, signals by

$$f(x, y) = A \exp(-(x^2 + y^2)/2r^2). \quad (2.9)$$

Gaussian signals, similar in form to the Gaussian probability density, are widely used in image processing.

2.2.4 Singularities

As in one-dimensional signals and systems, singularities play an important role in multidimensional signal and system analysis. These singularities are not functions

in the conventional sense; they can be described as functionals and are often referred to as *generalized functions*, rigorously treated using distribution theory. However, following common practice, we will nevertheless sometimes refer to them as functions (e.g., delta functions). We present here some basic properties of singularities suitable for our purposes. More details can be found in [Papo 68] and [Barr 04], and a more rigorous but relatively accessible development is given in [Rich 90].

The one-dimensional Dirac delta, denoted $\delta(t)$, is characterized by the property that

$$\int_{-\infty}^{\infty} \delta(t) f(t) dt = f(0) \quad (2.10)$$

for any function $f(t)$ that is continuous at $t = 0$. In particular, taking $f(t) = 1$, we have

$$\int_{-\infty}^{\infty} \delta(t) dt = 1.$$

The Dirac delta can be considered to be the limit of narrow pulses of unit area, for example, $\delta_{\Delta}(t) = (1/\Delta) \text{rect}(t/\Delta)$ or $\delta_{\Delta}(t) = (1/\Delta) \exp(-\pi t^2/\Delta^2)$, as $\Delta \rightarrow 0$.

The two-dimensional Dirac delta is defined in a similar fashion by the requirement

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(x, y) f(x, y) dx dy = f(0, 0) \quad (2.11)$$

for any function $f(x, y)$ that is continuous at $(x, y) = (0, 0)$. Again, the 2D Dirac delta can be considered to be the limiting case of narrow two-dimensional pulses of unit volume, e.g.,

$$\begin{aligned} \delta_{\Delta}(x, y) &= \frac{1}{\Delta^2} \text{rect}(x/\Delta, y/\Delta), \\ \delta_{\Delta}(x, y) &= \frac{1}{\Delta^2} \exp(-\pi(x^2 + y^2)/\Delta^2), \\ \delta_{\Delta}(x, y) &= \frac{1}{\pi\Delta^2} \text{circ}(x/\Delta, y/\Delta). \end{aligned}$$

The 2D Dirac delta satisfies the scaling property

$$\delta(ax, by) = \frac{1}{|ab|} \delta(x, y). \quad (2.12)$$

Other important properties of Dirac deltas will emerge as we investigate their role in multidimensional system analysis.

The Dirac delta can be extended to the three-dimensional case in an obvious fashion, with the notation $\delta(\mathbf{x})$ covering all cases. The conditions (2.10) and (2.11) are written

$$\int_{\mathbb{R}^D} \delta(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = f(\mathbf{0}) \quad (2.13)$$

in the general case, where $d\mathbf{x}$ is understood to mean dt , $dx dy$ or $dx dy dt$ according to context. As a consequence of (2.12) in the general case,

$$\delta(-\mathbf{x}) = \delta(\mathbf{x}). \quad (2.14)$$

In addition to the point singularities defined above, we can have singularities on lines or curves in two or three dimensions, or on surfaces in three dimensions. See Papoulis [Papo 68] for a discussion of singularities on a curve in two dimensions.

2.2.5 Zone-plate

A very useful two-dimensional function that is often used as a test pattern in imaging systems is the *zone-plate*, defined by

$$f(x, y) = \cos(\pi(x^2 + y^2)/r^2) \quad (2.15)$$

and illustrated in figure 2.5. It is often convenient to consider the zone-plate as the real part of the complex exponential function $\exp(j\pi(x^2 + y^2)/r^2)$.

Examining this figure, it can be seen that locally (say, within a small square window) the function is a sinusoidal signal with a horizontal frequency that increases with horizontal distance from the origin, and similarly a vertical frequency that increases with vertical distance from the origin. To make this concept of local frequency more precise, consider a conventional two-dimensional sinusoidal signal

$$\cos(2\pi(ux + vy)) = \cos(\phi(x, y)) \quad (2.16)$$

where $\phi(x, y) = 2\pi(ux + vy)$. In this case we see that the horizontal and vertical frequency are given by

$$u = \frac{1}{2\pi} \frac{\partial \phi}{\partial x};$$

$$v = \frac{1}{2\pi} \frac{\partial \phi}{\partial y}.$$

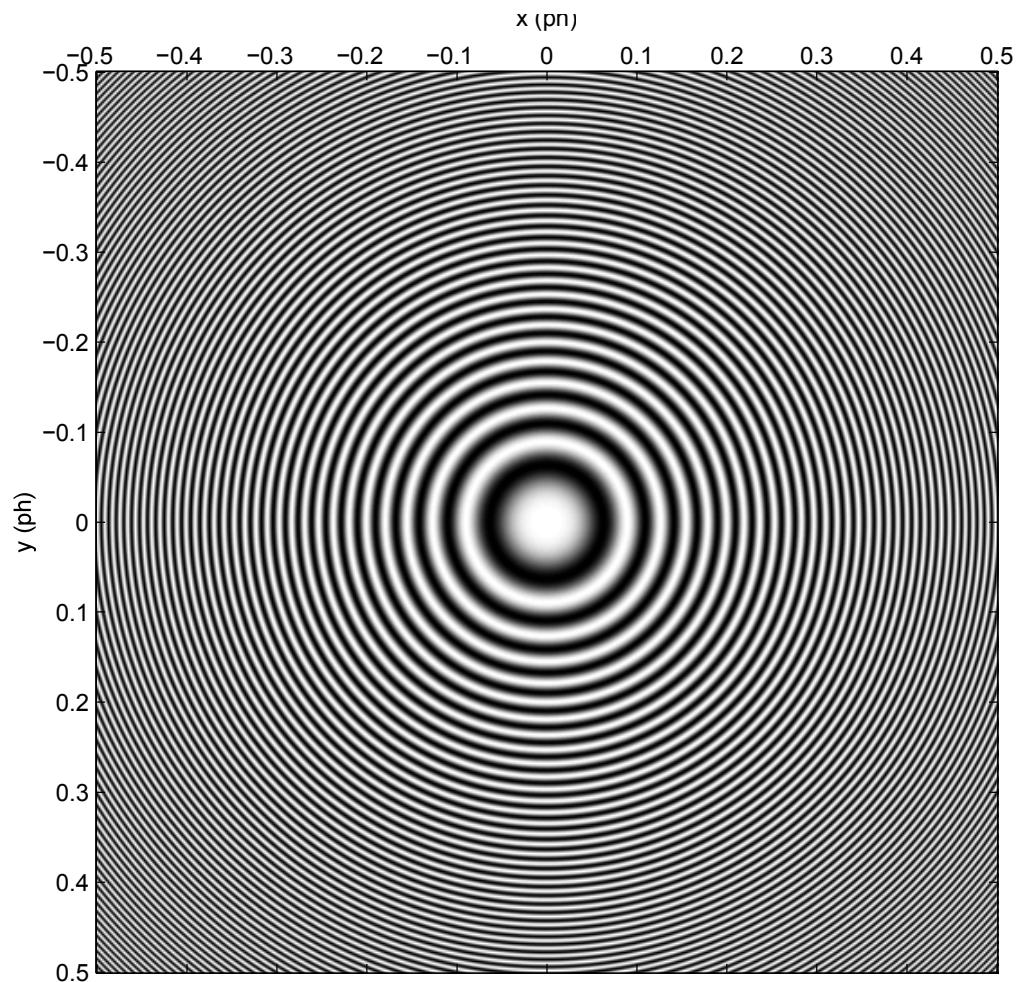


Figure 2.5: Illustration of the zoneplate.

We use these definitions to define the *local frequency* of a generalized sinusoidal signal $\cos(\phi(x,y))$. For the zone plate, we have $\phi(x,y) = \pi(x^2 + y^2)/r^2$ and so obtain

$$\begin{aligned} u(x,y) &= \frac{x}{r^2} \\ v(x,y) &= \frac{y}{r^2} \end{aligned}$$

which confirms that local horizontal and vertical frequency vary linearly with horizontal and vertical position respectively.

2.2.6 Separable and isotropic functions

A two-dimensional function $f(x,y)$ is said to be separable if it can be expressed as the product of one-dimensional functions,

$$f(x,y) = f_1(x)f_2(y). \quad (2.17)$$

Several of the 2D functions we have seen are separable, including $\text{rect}(x,y) = \text{rect}(x)\text{rect}(y)$, the complex exponential $\exp(j2\pi(ux+vy)) = \exp(j2\pi ux)\cdot\exp(j2\pi vy)$, and the exponential functions. Also, the 2D Dirac delta $\delta(x,y)$ can be considered to be separable: $\delta(x,y) = \delta(x)\delta(y)$. The extension to three-dimensional separable signals is evident. Separability is a convenient way to generate multidimensional signals from one-dimensional signals.

A two-dimensional signal is said to be isotropic (circularly symmetric) if it is only a function of the distance r from the origin,

$$f(x,y) = f_1(r) = f_1(\sqrt{x^2 + y^2}). \quad (2.18)$$

Examples of isotropic signals that we have seen are $\text{circ}(x,y)$, the Gaussian signal and the zone-plate. Again, the extension to multidimensional signals is evident.

2.3 Visualization of two-dimensional signals

It is easy to visualize a one-dimensional signal $f(t)$ by drawing its graph. If the graph is drawn to scale, we can derive numerical information by reading the graph,

e.g., $f(2.5) = 5.2$. There are various ways that we can visualize a 2D signal. The three main visualization techniques are

1. intensity image;
2. contour plot;
3. perspective plot.

Of course, method 1 is probably the most appropriate method to visualize a 2D signal that represents an image in the usual sense, but it may be useful in other cases as well. Contour and perspective plots are mainly useful for special 2D functions like the ones described above. Fig. 2.6 illustrates the three visualization methods for a 2D Gaussian function. MATLAB provides all the necessary tools to generate such figures. See Chapter 6 of [Hans 01] for a good description of how to generate such graphics in MATLAB.

2.4 Signal spaces and systems

A signal space \mathcal{S} is a collection of multidimensional signals defined on a specific domain \mathcal{D} and satisfying certain well-defined properties. For example, we can consider a space of all two-dimensional signals $f(x, y)$ defined for $-\infty < x < \infty$, $-\infty < y < \infty$, or we can define a space of 2D signals defined on the spatial window \mathcal{W} of Fig. 2.1. We can also impose additional constraints, such as boundedness, finite energy, or whatever constraints are appropriate for a given situation. An example of a specific signal space is

$$\mathcal{S} = \{f : \mathcal{W} \rightarrow \mathbb{R} \mid |f(\mathbf{x})| < \infty \text{ for all } \mathbf{x} \in \mathcal{W}\}.$$

As we will see later, we can also consider spaces of signals defined on discrete sets, corresponding to sampled images. We denote the fact that a given signal belongs to a signal space by

$$f \in \mathcal{S}.$$

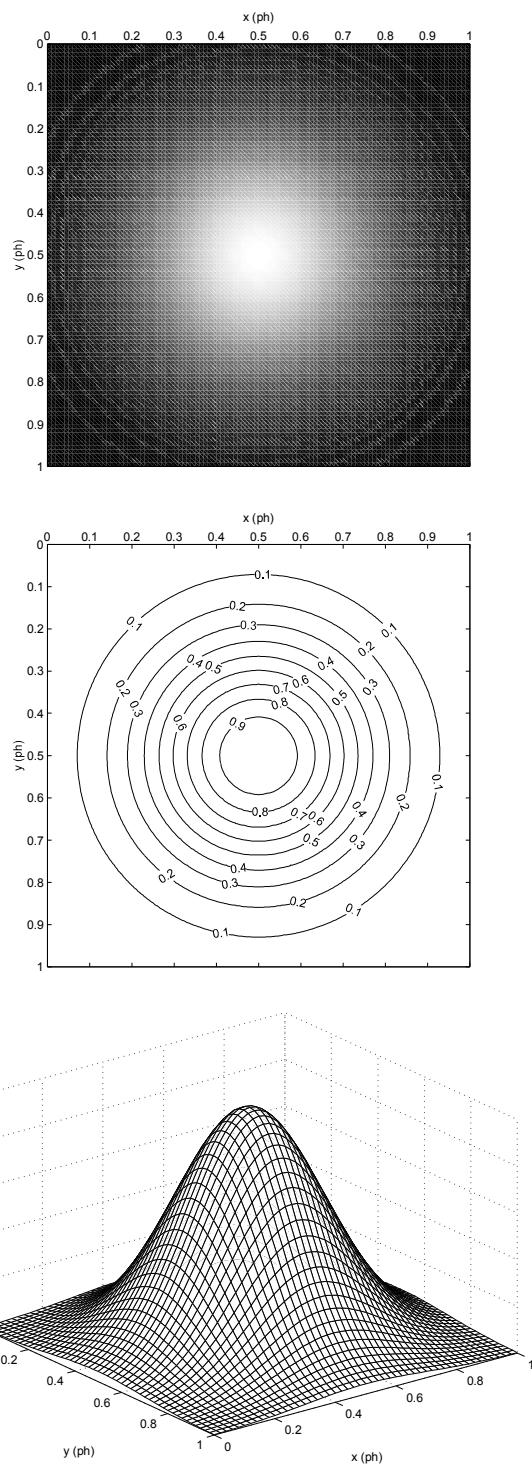


Figure 2.6: Visualization of a two-dimensional Gaussian function centered at (0.5,0.5) and scaled with $r = 0.2$. (a) Intensity plot. (b) Contour plot. (c) Perspective view.

The object f as a single entity is understood to encompass all the signal values $f(x, y)$ as (x, y) ranges over the specified domain of the signal space \mathcal{S} .

A system \mathcal{H} transforms elements of a signal space \mathcal{S}_1 into elements of a second signal space \mathcal{S}_2 according to some well-defined rule. We write

$$\mathcal{H} : \mathcal{S}_1 \rightarrow \mathcal{S}_2 : g = \mathcal{H}f \quad (2.19)$$

where $f \in \mathcal{S}_1$ and $g \in \mathcal{S}_2$. In most cases, \mathcal{S}_1 and \mathcal{S}_2 are the same, but sometimes they are different, e.g., for a system that samples a continuous-space image to produce a discrete-space image.

If we have two systems $\mathcal{H}_1 : \mathcal{S}_1 \rightarrow \mathcal{S}_2$ and $\mathcal{H}_2 : \mathcal{S}_2 \rightarrow \mathcal{S}_3$, then we can apply \mathcal{H}_1 and \mathcal{H}_2 successively to obtain a new system $\mathcal{H}_3 : \mathcal{S}_1 \rightarrow \mathcal{S}_3$ called the cascade of \mathcal{H}_1 and \mathcal{H}_2 . As shown in Fig. 2.7, $f_2 = \mathcal{H}_1 f_1$ and $f_3 = \mathcal{H}_2 f_2 = \mathcal{H}_2 \mathcal{H}_1 f_1$. Thus, we write $\mathcal{H}_3 = \mathcal{H}_2 \mathcal{H}_1$.

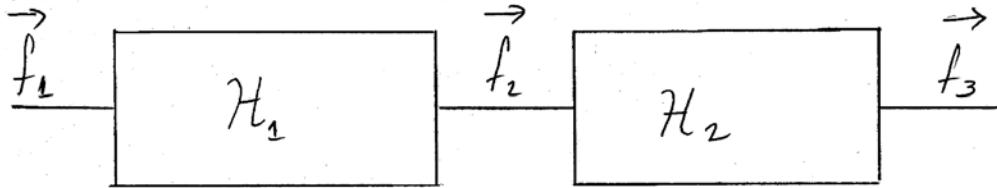


Figure 2.7: Cascade of two systems.

2.5 Continuous-domain linear systems

Let $\mathcal{H} : \mathcal{S} \rightarrow \mathcal{S}$ be a system defined on a space \mathcal{S} of continuous-domain multidimensional signals defined on some domain \mathcal{D} . Many systems of practical interest satisfy the key property of *linearity*. This is convenient since linear systems are normally simpler to analyze than general nonlinear systems.

We first make the assumption that the signal space \mathcal{S} has the properties of a vector space. This ensures that elements of the signal space can be added together,

and can be multiplied by a scalar constant, and that in each case the result also lies in the given signal space. This holds for most signal spaces of interest.

If $f_1 \in \mathcal{S}$ and $f_2 \in \mathcal{S}$, then the sum $g_1 = f_1 + f_2 \in \mathcal{S}$ is defined by $g_1(\mathbf{x}) = f_1(\mathbf{x}) + f_2(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{D}$. Similarly, $g_2 = \alpha f_1$, the multiplication by a scalar, is defined by $g_2(\mathbf{x}) = \alpha f_1(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{D}$. Given these conditions, a system \mathcal{H} is said to be linear if

$$\mathcal{H}(f_1 + f_2) = \mathcal{H}f_1 + \mathcal{H}f_2, \quad (2.20)$$

$$\mathcal{H}(\alpha f_1) = \alpha(\mathcal{H}f_1) \quad (2.21)$$

for all $f_1, f_2 \in \mathcal{S}$ and for all $\alpha \in \mathbb{R}$ (or for all $\alpha \in \mathbb{C}$ for complex signal spaces). The definition extends in an obvious fashion if \mathcal{S}_1 and \mathcal{S}_2 are different vector spaces. Linear systems are of particular interest because if we know the response of the system to a number of basic signals f_1, f_2, \dots, f_K , namely $g_i = \mathcal{H}f_i$, then we can determine the response to any linear combination of the f_i :

$$\mathcal{H}\left(\sum_{i=1}^K \alpha_i f_i\right) = \sum_{i=1}^K \alpha_i \mathcal{H}f_i = \sum_{i=1}^K \alpha_i g_i.$$

As a simple example of a linear system, consider the shift (or translation) operator $\mathcal{T}_{\mathbf{d}}$ for some fixed shift vector \mathbf{d} . If $g = \mathcal{T}_{\mathbf{d}}f$, then $g(\mathbf{x}) = f(\mathbf{x} - \mathbf{d})$. For this operation to be well defined, the domain of the signal space must be all of \mathbb{R}^D . In two dimensions, we would have $\mathbf{d} = (d_x, d_y)$, and $g(x, y) = f(x - d_x, y - d_y)$. It can easily be verified from the definitions that $\mathcal{T}_{\mathbf{d}}$ is a linear system. Fig. 2.8 illustrates the shift operator with $\mathbf{d} = [.25, -.25]^T$.

Another important class of linear systems consists of systems induced by an arbitrary nonsingular linear transformation of the domain \mathcal{D} . Let $\mathcal{A} : \mathcal{D} \rightarrow \mathcal{D} : \mathbf{x} \mapsto \mathbf{Ax}$ be such a transformation, where \mathbf{A} is a $D \times D$ nonsingular matrix. The induced system $\mathcal{M}_{\mathbf{A}}$ is defined by

$$\mathcal{M}_{\mathbf{A}} : \mathcal{S} \rightarrow \mathcal{S} : g = \mathcal{M}_{\mathbf{A}}f : g(\mathbf{x}) = f(\mathbf{Ax}). \quad (2.22)$$

Again, it is easily verified from the definitions that $\mathcal{M}_{\mathbf{A}}$ is a linear system. We will mainly use this category of systems for scaling and rotating basic signals such

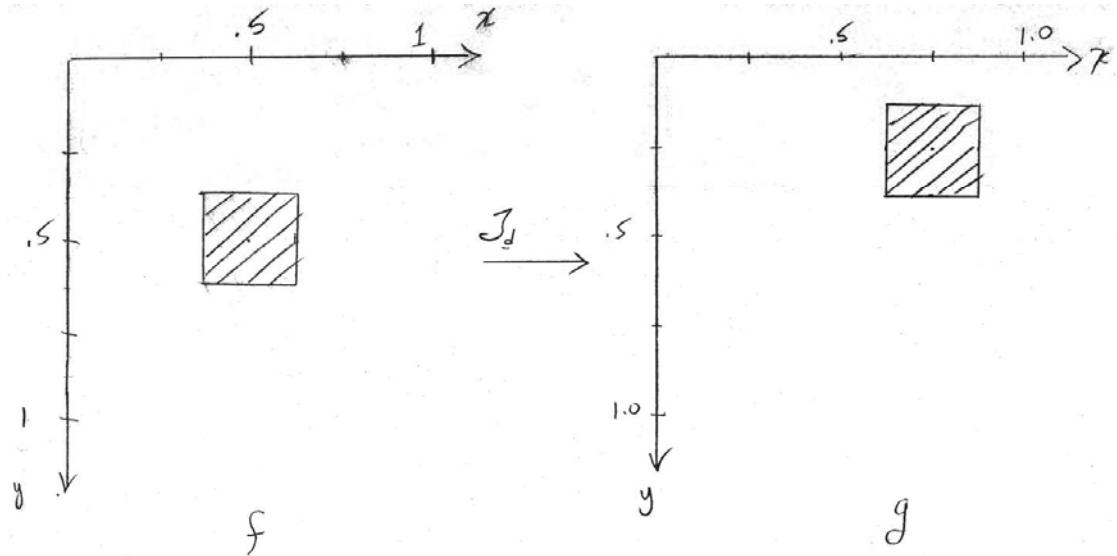


Figure 2.8: Shift operator $g = T_d f$ with $\mathbf{d} = [0.25, -0.25]^T$ giving $g(x, y) = f(x - 0.25, y + 0.25)$.

as those of Section 2.2, but more general cases are widely used as well. If \mathbf{A} is a diagonal matrix, the transformation $\mathcal{M}_{\mathbf{A}}$ carries out a separate scaling along each of the axes, as illustrated in two dimensions in Fig. 2.9(a). If \mathbf{A} is a two-dimensional rotation matrix,

$$\mathbf{A}_\theta = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}$$

the transformation $\mathcal{M}_{\mathbf{A}_\theta}$ rotates the signal in the clockwise direction, as illustrated in Fig. 2.9(b).

Another more general class of linear systems involves an affine transformation of the independent variables. One way to express this is

$$\mathcal{Q}_{\mathbf{A}, \mathbf{d}} : g = \mathcal{Q}_{\mathbf{A}, \mathbf{d}} f : g(\mathbf{x}) = f(\mathbf{A}(\mathbf{x} - \mathbf{d})),$$

where \mathbf{A} is a nonsingular $D \times D$ matrix. The affine transformation can be expressed as a cascade of the two preceding types of linear systems in two ways: $\mathcal{Q}_{\mathbf{A}, \mathbf{d}} =$

$\mathcal{T}_d \mathcal{M}_A = \mathcal{M}_A \mathcal{T}_{Ad}$. It can be shown that the cascade of any linear systems is also a linear system, and thus the system induced by an affine transformation of the domain is a linear system. As an example of an affine transformation, the Gaussian image of Fig. 2.6 can be obtained from the unit variance, zero-centered Gaussian $f(x, y) = \exp[-(x^2 + y^2)/2]$ using the affine transformation with

$$\mathbf{A} = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix} \quad \mathbf{d} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}.$$

An important subclass of linear systems is the class of *linear shift-invariant* (LSI) systems. In such a system, if the response to an input f is g , then if f is shifted by any amount \mathbf{d} , the resulting output is equal to g shifted by the same amount \mathbf{d} . Using the above terminology, if $\mathcal{H}f = g$, then $\mathcal{H}(\mathcal{T}_d f) = \mathcal{T}_d(\mathcal{H}f) = \mathcal{T}_d g$ for any \mathbf{d} . For an LSI system, $\mathcal{H}\mathcal{T}_d = \mathcal{T}_d\mathcal{H}$ for any \mathbf{d} . We say that the two systems \mathcal{H} and \mathcal{T}_d *commute*. The shift system \mathcal{T}_d is itself an LSI system, since $\mathcal{T}_d\mathcal{T}_s = \mathcal{T}_{d+s} = \mathcal{T}_{s+d} = \mathcal{T}_s\mathcal{T}_d$ for any \mathbf{s} . However, the general affine transformation system $\mathcal{Q}_{A,d}$ is *not* an LSI system if $\mathbf{A} \neq \mathbf{I}$, since $\mathcal{Q}_{A,d}\mathcal{T}_s = \mathcal{Q}_{A,0}\mathcal{T}_{Ad+s}$ but $\mathcal{T}_s\mathcal{Q}_{A,d} = \mathcal{T}_{s+d}\mathcal{Q}_{A,0} = \mathcal{Q}_{A,0}\mathbf{T}_{Ad+As}$.

2.5.1 Response of a linear system

The defining property of the Dirac delta function is given in equation (2.13):

$\int_{\mathbb{R}^D} \delta(\mathbf{x})f(\mathbf{x}) d\mathbf{x} = f(\mathbf{0})$. From this, we can derive the so-called *sifting property*

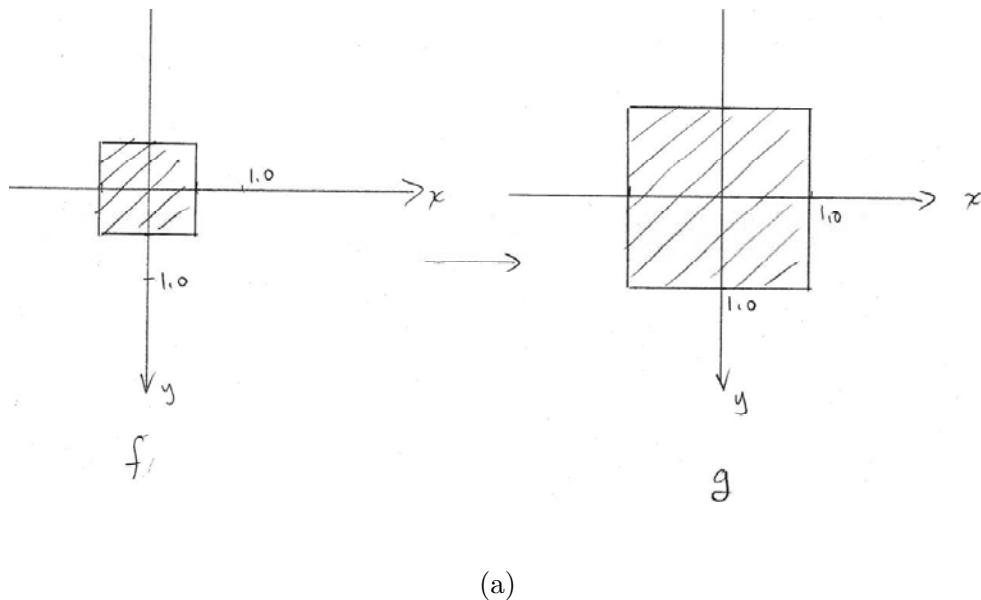
$$f(\mathbf{x}) = \int_{\mathbb{R}^D} f(\mathbf{s})\delta(\mathbf{x} - \mathbf{s}) d\mathbf{s}. \quad (2.23)$$

This follows from

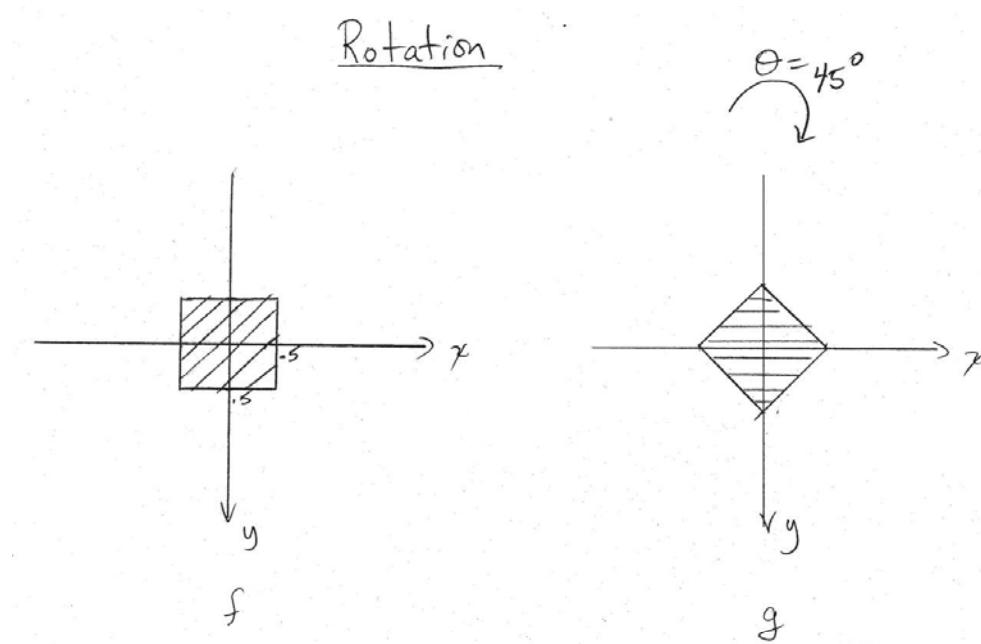
$$\begin{aligned} \int_{\mathbb{R}^D} f(\mathbf{s})\delta(\mathbf{x} - \mathbf{s}) d\mathbf{s} &= \int_{\mathbb{R}^D} f(\mathbf{s})\delta(\mathbf{s} - \mathbf{x}) d\mathbf{s} \quad (\text{from Eq. (2.14)}) \\ &= \int_{\mathbb{R}^D} f(\mathbf{t} + \mathbf{x})\delta(\mathbf{t}) d\mathbf{t} \quad (\text{letting } \mathbf{t} = \mathbf{s} - \mathbf{x}) \\ &= f(\mathbf{x}) \quad (\text{from eq(2.13)}). \end{aligned}$$

The sifting property (2.23) can be interpreted as the *synthesis* of the signal f by the superposition of shifted Dirac delta functions $\mathcal{T}_s\delta$ with weights $f(\mathbf{s})d\mathbf{s}$:

$$f = \int_{\mathbb{R}^D} (f(\mathbf{s}) d\mathbf{s}) \mathcal{T}_s\delta.$$



(a)



(b)

Figure 2.9: Transformations $g = \mathcal{M}_A f$ (a) Scale operator with $\mathbf{A} = \text{diag}(0.5, 0.5)$.
 (b) Rotation operator with $\theta = \pi/4$.

For a linear system \mathcal{H} , we can conclude that

$$g = \mathcal{H}f = \int_{\mathbb{R}^D} (f(\mathbf{s}) d\mathbf{s}) \mathcal{H}(\mathcal{T}_{\mathbf{s}}\delta)$$

where $\mathcal{H}(\mathcal{T}_{\mathbf{s}}\delta)$ is the response of the system to a Dirac delta function located at position \mathbf{s} . If we denote this impulse response $h_{\mathbf{s}}$, we obtain

$$g(\mathbf{x}) = \int_{\mathbb{R}^D} f(\mathbf{s}) h_{\mathbf{s}}(\mathbf{x}) d\mathbf{s}.$$

This equation describes the response of a general space-variant linear system. Most optical systems are indeed space variant, with the response to an impulse in the corner of the image being different than the response to an impulse in the center of the image, for example. However, the design goal is usually to have a system that is as close to being shift invariant as possible. Thus, shift invariant systems are an important class. In this case,

$$\mathcal{H}(\mathcal{T}_{\mathbf{s}}\delta) = \mathcal{T}_{\mathbf{s}}(\mathcal{H}\delta) = \mathcal{T}_{\mathbf{s}}h$$

where h is the response of the LSI system to an impulse at the origin, and so

$$g = \int_{\mathbb{R}^D} (f(\mathbf{s}) d\mathbf{s}) \mathcal{T}_{\mathbf{s}}h.$$

Evaluating at position \mathbf{x} gives

$$g(\mathbf{x}) = \int_{\mathbb{R}^D} f(\mathbf{s}) h(\mathbf{x} - \mathbf{s}) d\mathbf{s} \quad (2.24)$$

which is called the *convolution integral*, and is denoted

$$g = f * h.$$

By a simple change of variables in the integral (2.24), we can show that $f * h = h * f$, i.e., convolution is commutative.

It is important to recognize that while the above result is broadly valid and applies to essentially all cases of interest to us, the development is informal and there are many unstated assumptions. For example, the existence of $\mathcal{H}\delta$ and the applicability of the linearity condition to an integral relation are assumed. The

development can be treated more rigorously in several ways, such as assuming a specific signal space with a given metric, and assuming that the linear system is continuous. Then, an arbitrary signal can be approximated as a superposition of a finite number of pulse functions of the form $\delta_\Delta(\mathbf{x} - \mathbf{x}_i)$, and the output to this approximation determined. Taking the limit as $\Delta \rightarrow 0$ and the extent tends to infinity yields the desired result. More details of this approach can be found in Section 7.2 of [Barr 04].

Example 2.1. A common model for blurring due to a camera being out of focus is *uniform out-of-focus blur*, in which a point source anywhere in the scene is imaged as a small disk of radius R [Lage 09]. In this case the impulse response, or point spread function, of the system is given by $h(x, y) = \frac{1}{\pi R^2} \text{circ}\left(\frac{x}{R}, \frac{y}{R}\right)$. Suppose that we image a pure step function in the horizontal direction with this camera:

$$f(x, y) = u_0(x - 0.5) = \begin{cases} 0 & \text{if } x \leq 0.5, \\ 1 & \text{if } x > 0.5. \end{cases}$$

The output is computed as

$$\begin{aligned} g(x, y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(s_1, s_2) f(x - s_1, y - s_2) ds_1 ds_2 \\ &= \frac{1}{\pi R^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{circ}\left(\frac{s_1}{R}, \frac{s_2}{R}\right) u_0(x - s_1 - 0.5) ds_1 ds_2 \\ &= \frac{1}{\pi R^2} \int_{-\infty}^{\infty} \int_{-\infty}^{x-0.5} \text{circ}\left(\frac{s_1}{R}, \frac{s_2}{R}\right) ds_1 ds_2. \end{aligned}$$

This amounts to computing the area of the part of the circle overlapping the step. Clearly this area is zero if $x < 0.5 - R$ and it is one if $x > 0.5 + R$. Otherwise, using simple geometry we can determine the area to obtain the resulting image

$$g(x, y) = \begin{cases} 0 & x \leq 0.5 - R \\ \frac{1}{\pi R^2} \left[R^2 \cos^{-1} \left(\frac{0.5-x}{R} \right) - (0.5-x) \sqrt{R^2 - (0.5-x)^2} \right] & 0.5 - R < x \leq 0.5 \\ 1 - \frac{1}{\pi R^2} \left[R^2 \cos^{-1} \left(\frac{x-0.5}{R} \right) - (x-0.5) \sqrt{R^2 - (x-0.5)^2} \right] & 0.5 < x \leq 0.5 + R \\ 1 & x > 0.5 + R \end{cases}$$

2.5.2 Frequency response of an LSI system

Suppose that the input to an LSI system is the complex sinusoidal function $f(\mathbf{x}) = \exp(j2\pi\mathbf{u} \cdot \mathbf{x})$. According to equation (2.24), the corresponding output is

$$\begin{aligned} g(\mathbf{x}) &= \int_{\mathbb{R}^D} \exp(j2\pi\mathbf{u} \cdot \mathbf{s}) h(\mathbf{x} - \mathbf{s}) d\mathbf{s} \\ &= \int_{\mathbb{R}^D} \exp(j2\pi\mathbf{u} \cdot (\mathbf{x} - \mathbf{t})) h(\mathbf{t}) d\mathbf{t} \quad (\text{setting } \mathbf{t} = \mathbf{x} - \mathbf{s}) \\ &= \left(\int_{\mathbb{R}^D} h(\mathbf{t}) \exp(-j2\pi\mathbf{u} \cdot \mathbf{t}) d\mathbf{t} \right) \exp(j2\pi\mathbf{u} \cdot \mathbf{x}) \\ &= H(\mathbf{u}) \exp(j2\pi\mathbf{u} \cdot \mathbf{x}) \end{aligned} \quad (2.25)$$

where $H(\mathbf{u})$ is a complex scalar (assuming the integral converges). Thus, exactly as in one dimension, if the input to an LSI system is a complex sinusoidal signal with frequency vector \mathbf{u} , then the output is that same complex sinusoidal signal multiplied by the complex scalar $H(\mathbf{u})$. Taken as a function of the two or three-dimensional frequency vector, $H(\mathbf{u})$ is referred to as the frequency response of the LSI system. According to this observation, $\exp(j2\pi\mathbf{u} \cdot \mathbf{x})$ is called an *eigenfunction* of the linear system \mathcal{H} with corresponding *eigenvalue* $H(\mathbf{u})$.

Multiplication by $H(\mathbf{u}) = |H(\mathbf{u})| \exp(j\angle H(\mathbf{u}))$ amounts to multiplying the magnitude of $\exp(j2\pi\mathbf{u} \cdot \mathbf{x})$ by $|H(\mathbf{u})|$ and introducing a phase shift of $\angle H(\mathbf{u})$, i.e.,

$$g(\mathbf{x}) = |H(\mathbf{u})| \exp(j(2\pi\mathbf{u} \cdot \mathbf{x} + \angle H(\mathbf{u})))$$

2.6 The multidimensional Fourier transform

From equation (2.25) we identify

$$H(\mathbf{u}) = \int_{\mathbb{R}^D} h(\mathbf{x}) \exp(-j2\pi\mathbf{u} \cdot \mathbf{x}) d\mathbf{x} \quad (2.26)$$

as the multidimensional extension of the continuous-time Fourier transform. The multiD Fourier transform has properties that are completely analogous to the familiar properties of the 1D Fourier transform and so we simply present them without

proof in Table 2.1. In particular, the inverse Fourier transform is given by

$$h(\mathbf{x}) = \int_{\mathbb{R}^D} H(\mathbf{u}) \exp(j2\pi\mathbf{u} \cdot \mathbf{x}) d\mathbf{u}.$$

The Fourier transform can be applied to any signals in the signal space, not just the impulse response. We denote that $f(\mathbf{x})$ and $F(\mathbf{u})$ form a multidimensional Fourier transform pair by $f(\mathbf{x}) \xleftrightarrow{\text{CDFT}} F(\mathbf{u})$, where CDFT denotes continuous-domain Fourier transform.

The property that makes the Fourier transform so valuable in linear system analysis is the *convolution property* (property v): the Fourier transform of $f * h$ is $F(\mathbf{u})H(\mathbf{u})$. Thus, if the input f to an LSI system with frequency response $H(\mathbf{u})$ has Fourier transform $F(\mathbf{u})$, the output g has Fourier transform $G(\mathbf{u}) = F(\mathbf{u})H(\mathbf{u})$.

The proofs of the properties in Table 2.1 are straightforward and similar to analogous proofs for the one-dimensional Fourier transform, as given in many standard texts such as [Brac 00], [Gray 95]. They are left to the reader as an exercise. Since property (iv) relating to a linear transformation of the domain \mathbb{R}^D is somewhat different from the one-dimensional case, we provide the proof for this property as an illustration of the methods.

Linear transformation of the domain \mathbb{R}^D , property (iv). If $f(\mathbf{x}) \xleftrightarrow{\text{CDFT}} F(\mathbf{u})$, then $f(\mathbf{Ax}) \xleftrightarrow{\text{CDFT}} \frac{1}{|\det \mathbf{A}|} F(\mathbf{A}^{-T}\mathbf{u})$, where \mathbf{A} is a nonsingular $D \times D$ matrix and \mathbf{A}^{-T} denotes $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$.

Proof. If we denote $g(\mathbf{x}) = f(\mathbf{Ax})$, then

$$G(\mathbf{u}) = \int_{\mathbb{R}^D} f(\mathbf{Ax}) \exp(-j2\pi\mathbf{u} \cdot \mathbf{x}) d\mathbf{x}.$$

With the change of variables $\mathbf{s} = \mathbf{Ax}$, the Jacobian is $\frac{\partial(\mathbf{s})}{\partial(\mathbf{x})} = \det \mathbf{A}$, and using standard techniques for change of variables in an integral (e.g., [Kapl 84, Sect. 4.6]), we obtain

$$\begin{aligned} G(\mathbf{u}) &= \frac{1}{|\det \mathbf{A}|} \int_{\mathbb{R}^D} f(\mathbf{s}) \exp(-j2\pi\mathbf{u} \cdot (\mathbf{A}^{-1}\mathbf{s})) d\mathbf{s} \\ &= \frac{1}{|\det \mathbf{A}|} \int_{\mathbb{R}^D} f(\mathbf{s}) \exp(-j2\pi(\mathbf{A}^{-T}\mathbf{u}) \cdot \mathbf{s}) d\mathbf{s} \\ &= \frac{1}{|\det \mathbf{A}|} F(\mathbf{A}^{-T}\mathbf{u}) \end{aligned}$$

$f(\mathbf{x}) = \int_{\mathbb{R}^D} F(\mathbf{u}) \exp(j2\pi\mathbf{u} \cdot \mathbf{x}) d\mathbf{u}$	$F(\mathbf{u}) = \int_{\mathbb{R}^D} f(\mathbf{x}) \exp(-j2\pi\mathbf{u} \cdot \mathbf{x}) d\mathbf{x}$
(i) $af_1(\mathbf{x}) + bf_2(\mathbf{x})$	$aF_1(\mathbf{u}) + bF_2(\mathbf{u})$
(ii) $f(\mathbf{x} - \mathbf{x}_0)$	$F(\mathbf{u}) \exp(-j2\pi\mathbf{u} \cdot \mathbf{x}_0)$
(iii) $f(\mathbf{x}) \exp(j2\pi\mathbf{u}_0 \cdot \mathbf{x})$	$F(\mathbf{u} - \mathbf{u}_0)$
(iv) $f(\mathbf{Ax})$	$\frac{1}{ \det \mathbf{A} } F(\mathbf{A}^{-T}\mathbf{u})$
(v) $f_1(\mathbf{x}) * f_2(\mathbf{x})$	$F_1(\mathbf{u})F_2(\mathbf{u})$
(vi) $f_1(\mathbf{x})f_2(\mathbf{x})$	$F_1(\mathbf{u}) * F_2(\mathbf{u})$
(vii) $\nabla_{\mathbf{x}}f(\mathbf{x})$	$j2\pi\mathbf{u}F(\mathbf{u})$
(viii) $\mathbf{x}f(\mathbf{x})$	$\frac{j}{2\pi}\nabla_{\mathbf{u}}F(\mathbf{u})$
(ix) $F(\mathbf{x})$	$f(-\mathbf{u})$
(x) $f^*(\mathbf{x})$	$F^*(-\mathbf{u})$
(xi) $f_1(x)f_2(y)$	$F_1(u)F_2(v)$
(xii) $\int_{\mathbb{R}^D} f(\mathbf{x}) ^2 d\mathbf{x} = \int_{\mathbb{R}^D} F(\mathbf{u}) ^2 d\mathbf{u}$	

Table 2.1: Multidimensional Fourier transform properties.

where we use the identity $\mathbf{a} \cdot (\mathbf{C}\mathbf{b}) = (\mathbf{C}^T\mathbf{a}) \cdot \mathbf{b}$ (see problem). \square

Property (iv) can be used to determine the effect of independent scaling of x and y axes, of rotation of the image, or of an affine transformation of the independent variable (along with property (ii)).

2.6.1 Evaluation of multidimensional Fourier transforms

In general, the multiD Fourier transform is determined by direct evaluation of the defining integral (2.26) using standard methods of integral calculus. Simplifications are possible if the function $f(\mathbf{x})$ is separable or isotropic, and of course maximum use should be made of the Fourier transform properties of Table 2.1. A few examples follow, and Table 2.2 provides a number of useful multiD Fourier transforms.

Example 2.2. Compute the 2D Fourier transform of $f(x, y) = \text{rect}(x, y)$.

Solution.

$$\begin{aligned} F(u, v) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{rect}(x, y) \exp(-j2\pi(ux + vy)) dx dy \\ &= \int_{-0.5}^{0.5} \int_{-0.5}^{0.5} \exp(-j2\pi(ux + vy)) dx dy \\ &= \int_{-0.5}^{0.5} \exp(-j2\pi ux) dx \int_{-0.5}^{0.5} \exp(-j2\pi vy) dy \\ &= \frac{\sin \pi u}{\pi u} \frac{\sin \pi v}{\pi v} \\ &= \text{sinc}(u) \text{sinc}(v), \end{aligned}$$

where we introduce the standard function

$$\text{sinc}(t) = \frac{\sin \pi t}{\pi t}.$$

It is easy to show that $\text{sinc}(0) = 1$. We note that since $\text{rect}(x, y)$ is separable, the resulting 2D Fourier transform is the product of the corresponding 1D Fourier transforms. \square

Example 2.3. Compute the 2D Fourier transform of $f(x, y) = c \operatorname{rect}(ax, by)$, where $a > 0$ and $b > 0$.

Solution. Use property (i) and property (iv), with

$$\mathbf{A} = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}.$$

Then $|\det \mathbf{A}| = ab$ and

$$\mathbf{A}^{-T} = \begin{bmatrix} \frac{1}{a} & 0 \\ 0 & \frac{1}{b} \end{bmatrix}.$$

It follows that

$$\begin{aligned} F(u, v) &= \frac{c}{ab} \frac{\sin(\pi u/a)}{\pi u/a} \frac{\sin(\pi v/b)}{\pi v/b} \\ &= c \frac{\sin(\pi u/a) \sin(\pi v/b)}{\pi^2 uv} \\ &= \frac{c}{ab} \operatorname{sinc}(u/a) \operatorname{sinc}(v/b). \end{aligned}$$

□

Example 2.4. Compute the 2D Fourier transform of $f(x, y) = \operatorname{circ}(x, y)$.

Solution.

$$F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \operatorname{circ}(x, y) \exp(-j2\pi(ux + vy)) dx dy$$

This function is not separable but it is isotropic. We make the following changes of variables to polar coordinates:

$$\begin{array}{ll} x = r \cos \theta & y = r \sin \theta \\ u = w \cos \phi & v = w \sin \phi. \end{array}$$

The region of support of $\text{circ}(x, y)$ is the circle $0 \leq r \leq 1$, $-\pi \leq \theta < \pi$. Thus, denoting $F'(w, \phi) = F(w \cos \phi, w \sin \phi)$, we obtain

$$\begin{aligned} F'(w, \phi) &= \int_{-\pi}^{\pi} \int_0^1 \exp(-j2\pi(rw \cos \theta \cos \phi + rw \sin \theta \sin \phi)) r dr d\theta \\ &= \int_0^1 r \int_{-\pi}^{\pi} \exp(-j2\pi rw \cos(\theta - \phi)) d\theta dr \\ &= 2\pi \int_0^1 r J_0(2\pi rw) dr \\ &= \frac{1}{2\pi w^2} \int_0^{2\pi w} z J_0(z) dz \quad (z = 2\pi rw) \\ &= \frac{1}{w} J_1(2\pi w) \end{aligned}$$

where $J_0(\cdot)$ and $J_1(\cdot)$ are Bessel functions of the first kind of order zero and order one respectively [Poul 98]. We have used the properties that

$$\begin{aligned} J_0(x) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(jx \cos(\theta - \alpha)) d\theta \quad \text{for any } \alpha \\ x J_1(x) &= \int_0^x s J_0(s) ds \end{aligned}$$

Thus,

$$F(u, v) = \frac{1}{\sqrt{u^2 + v^2}} J_1(2\pi \sqrt{u^2 + v^2}).$$

We can now conclude using properties (i) and (iv) that the frequency response of the out-of-focus blur of example 2.1 with $h(x, y) = \frac{1}{\pi R^2} \text{circ}(\frac{x}{R}, \frac{y}{R})$ is

$$\begin{aligned} H(u, v) &= \frac{1}{\pi R^2} \cdot R^2 \frac{1}{R \sqrt{u^2 + v^2}} J_1(2\pi R \sqrt{u^2 + v^2}) \\ &= \frac{1}{\pi R \sqrt{u^2 + v^2}} J_1(2\pi R \sqrt{u^2 + v^2}). \end{aligned}$$

□

Example 2.5. Compute the impulse response of an ideal two-dimensional lowpass filter with frequency response $H(u, v) = \text{circ}(u/W, v/W)$, where W is the bandwidth of the filter.

$f(\mathbf{x}) = \int_{\mathbb{R}^D} F(\mathbf{u}) \exp(j2\pi\mathbf{u} \cdot \mathbf{x}) d\mathbf{u}$	$F(\mathbf{u}) = \int_{\mathbb{R}^D} f(\mathbf{x}) \exp(-j2\pi\mathbf{u} \cdot \mathbf{x}) d\mathbf{x}$
$\text{rect}(x, y)$	$\frac{\sin \pi u}{\pi u} \frac{\sin \pi v}{\pi v}$
$\text{circ}(x, y)$	$\frac{1}{\sqrt{u^2 + v^2}} J_1(2\pi\sqrt{u^2 + v^2})$
$\exp(-(x^2 + y^2)/2r^2)$	$2\pi r^2 \exp(-2\pi^2(u^2 + v^2)r^2)$
$\cos(\pi(x^2 + y^2)/r^2)$	$r^2 \sin(\pi(u^2 + v^2)r^2)$
$\exp(j\pi(x^2 + y^2)/r^2)$	$jr^2 \exp(-j\pi(u^2 + v^2)r^2)$
$\delta(\mathbf{x})$	1

Table 2.2: Multidimensional Fourier transform of selected functions.

Solution. From Example 2.4 and property (iv),

$$\text{circ}\left(\frac{x}{W}, \frac{y}{W}\right) \xleftrightarrow{\text{CDFT}} \frac{W}{\sqrt{u^2 + v^2}} J_1(2\pi W \sqrt{u^2 + v^2}).$$

Noting that circ is symmetric about the origin, the duality property (ix) yields

$$h(x, y) = \frac{W}{\sqrt{x^2 + y^2}} J_1(2\pi W \sqrt{x^2 + y^2}).$$

□

Example 2.6. A reasonable model for the impulse response of a vidicon video camera sensor is a separable Gaussian in the spatial dimensions and rect function in the temporal dimension, specifically

$$\begin{aligned} h(x, y, t) &= \exp(-(x^2 + y^2)/2r^2) \text{rect}(t/T - .5) \\ &= \exp(-x^2/2r^2) \exp(-y^2/2r^2) \text{rect}(t/T - 0.5) \end{aligned}$$

Using separability, and the standard results for the Fourier transform of 1D signal

[Brac 00]

$$\begin{aligned} \exp(-\pi x^2) &\xleftrightarrow{\text{CDFT}} \exp(-\pi u^2) \\ \text{rect}(t) &\xleftrightarrow{\text{CDFT}} \frac{\sin \pi w}{\pi w} \end{aligned}$$

gives

$$\begin{aligned} H(u, v, w) &= \sqrt{2\pi}r \exp(-2\pi^2 u^2 r^2) \sqrt{2\pi}r \exp(-2\pi^2 v^2 r^2) \frac{\sin \pi w T}{\pi w} \exp(-j\pi T w) \\ &= 2r^2 \exp(-j\pi T w) \exp(-2\pi^2(u^2 + v^2)r^2) \sin \pi w T / w \end{aligned}$$

2.6.2 Two-dimensional Fourier transform of polygonal zero-one functions

Polygonal zero-one functions are frequently encountered in the analysis of modern cameras and display devices, and their Fourier transform is required. For the rectangular region considered in Example 2.3, it is straightforward to compute the Fourier transform by direct evaluation of the integral. However, for other shapes such as hexagons, octagons, chevrons, etc., direct computation of the Fourier transform is more involved and tedious. It is possible to convert the area integral in the direct definition of the Fourier transform to a line integral along the boundary of the region \mathcal{A} using the two-dimensional version of Gauss's divergence theorem and thereby obtain a closed form expression for the Fourier transform in the case of polygonal regions.

Let $\mathcal{A} \subset \mathbb{R}^2$ be a bounded, simply-connected region in the plane and define $p_{\mathcal{A}}(\mathbf{x})$ as in Eq. (2.1). Then, the Fourier transform is given by

$$P_{\mathcal{A}}(\mathbf{u}) = \iint_{\mathcal{A}} \exp(-j2\pi \mathbf{u} \cdot \mathbf{x}) d\mathbf{x}. \quad (2.27)$$

Let $\partial\mathcal{A}$ be the boundary of \mathcal{A} , assumed to be piecewise smooth, traversed in the clockwise direction. Then let $\mathbf{Q}(\mathbf{x}) = (Q_1(\mathbf{x}), Q_2(\mathbf{x}))$ be a vector field defined on \mathcal{A} , assumed to be continuous with continuous first partial derivatives. The divergence theorem (see for example [Kapl 84, Sect. 5.11]) states that

$$\iint_{\mathcal{A}} \text{div} \mathbf{Q}(\mathbf{x}) d\mathbf{x} = \oint_{\partial\mathcal{A}} \mathbf{Q}(\mathbf{x}) \cdot \mathbf{n}_{\mathbf{x}} dS(\mathbf{x}), \quad (2.28)$$

where

$$\operatorname{div} \mathbf{Q}(\mathbf{x}) = \frac{\partial Q_1(\mathbf{x})}{\partial x} + \frac{\partial Q_2(\mathbf{x})}{\partial y}, \quad (2.29)$$

\mathbf{n}_x is a unit vector normal to $\partial\mathcal{A}$ at \mathbf{x} and pointing outward, and $S(\mathbf{x})$ denotes arc length along $\partial\mathcal{A}$ at \mathbf{x} . This result can be applied to computing the Fourier transform in Eq. (2.27) by choosing

$$\mathbf{Q}_{\mathbf{u}}(\mathbf{x}) = \frac{j\mathbf{u}}{2\pi\|\mathbf{u}\|^2} \exp(-j2\pi\mathbf{u} \cdot \mathbf{x}). \quad (2.30)$$

By applying the definition of divergence,

$$\operatorname{div} \mathbf{Q}_{\mathbf{u}}(\mathbf{x}) = \exp(-j2\pi\mathbf{u} \cdot \mathbf{x}). \quad (2.31)$$

We then find that

$$P_{\mathcal{A}}(\mathbf{u}) = \oint_{\partial\mathcal{A}} \left(\frac{j\mathbf{u}}{2\pi\|\mathbf{u}\|^2} \exp(-j2\pi\mathbf{u} \cdot \mathbf{x}) \right) \cdot \mathbf{n}_x dS(\mathbf{x}). \quad (2.32)$$

This result has been called the Abbe transform and was cited in the dissertation of R. Straubel in 1888 [Komr 82]. As shown in [Komr 82], the contour integral can easily be evaluated in closed form for a polygonal region, as follows.

Assume that \mathcal{A} is a polygon with K sides, with vertices $\mathbf{a}_1, \dots, \mathbf{a}_K$ in clockwise direction; for convenience, we denote $\mathbf{a}_{K+1} = \mathbf{a}_1$. We define the following quantities that are easily determined once the vertices are specified:

$$d_k = \|\mathbf{a}_{k+1} - \mathbf{a}_k\| \quad \text{length of side } k \quad (2.33)$$

$$\mathbf{c}_k = \frac{\mathbf{a}_k + \mathbf{a}_{k+1}}{2} \quad \text{midpoint of side } k \quad (2.34)$$

$$\mathbf{t}_k = \frac{\mathbf{a}_{k+1} - \mathbf{a}_k}{d_k} \quad \text{unit vector parallel to side } k \quad (2.35)$$

$$\mathbf{n}_k = \mathcal{R}\mathbf{t}_k \quad \text{unit normal to side } k \text{ pointing outward} \quad (2.36)$$

where \mathcal{R} rotates counterclockwise by 90° . With these definitions, the Fourier transform expression given in Eq. (2.32) can be written as a sum of the integrals over each of the polygon sides as follows.

$$\begin{aligned}
P_{\mathcal{A}}(\mathbf{u}) &= \sum_{k=1}^K \int_{\mathbf{a}_k}^{\mathbf{a}_{k+1}} \frac{j\mathbf{u} \cdot \mathbf{n}_k}{2\pi\|\mathbf{u}\|^2} \exp(-j2\pi\mathbf{u} \cdot \mathbf{x}) dS(\mathbf{x}) \\
&= \frac{j}{2\pi\|\mathbf{u}\|^2} \sum_{k=1}^K (\mathbf{u} \cdot \mathbf{n}_k) \int_{-d_k/2}^{d_k/2} \exp(-j2\pi\mathbf{u} \cdot (\mathbf{c}_k + s\mathbf{t}_k)) ds \\
&= \frac{j}{2\pi\|\mathbf{u}\|^2} \sum_{k=1}^K (\mathbf{u} \cdot \mathbf{n}_k) \exp(-j2\pi\mathbf{u} \cdot \mathbf{c}_k) \int_{-d_k/2}^{d_k/2} \exp(-j2\pi s\mathbf{u} \cdot \mathbf{t}_k) ds. \quad (2.37)
\end{aligned}$$

The integral can be easily evaluated to give the final result:

$$P_{\mathcal{A}}(\mathbf{u}) = \frac{j}{2\pi\|\mathbf{u}\|^2} \sum_{k=1}^K d_k (\mathbf{u} \cdot \mathbf{n}_k) \exp(-j\pi\mathbf{u} \cdot (\mathbf{a}_{k+1} + \mathbf{a}_k)) \frac{\sin(\pi\mathbf{u} \cdot (\mathbf{a}_{k+1} - \mathbf{a}_k))}{\pi\mathbf{u} \cdot (\mathbf{a}_{k+1} - \mathbf{a}_k)}. \quad (2.38)$$

In many (but not all) cases of interest, the polygon is symmetric about the origin, i.e., $\mathbf{x} \in \mathcal{A} \Rightarrow -\mathbf{x} \in \mathcal{A}$. In this case, the number of vertices and sides is necessarily even, and the terms corresponding to two opposite sides in Eq. (2.38) can be combined to yield a real-valued Fourier transform [Lu 09].

$$\begin{aligned}
P_{\mathcal{A}}(\mathbf{u}) &= \frac{1}{\pi\|\mathbf{u}\|^2} \sum_{k=1}^{K/2} d_k (\mathbf{u} \cdot \mathbf{n}_k) \sin(\pi\mathbf{u} \cdot (\mathbf{a}_{k+1} + \mathbf{a}_k)) \frac{\sin(\pi\mathbf{u} \cdot (\mathbf{a}_{k+1} - \mathbf{a}_k))}{\pi\mathbf{u} \cdot (\mathbf{a}_{k+1} - \mathbf{a}_k)} \\
&= \frac{1}{\pi\|\mathbf{u}\|^2} \sum_{k=1}^{K/2} d_k (\mathbf{u} \cdot \mathbf{n}_k) \sin(\pi\mathbf{u} \cdot (\mathbf{a}_{k+1} + \mathbf{a}_k)) \text{sinc}(\pi\mathbf{u} \cdot (\mathbf{a}_{k+1} - \mathbf{a}_k)). \quad (2.39)
\end{aligned}$$

This result has been extended to zero-one functions in more than two-dimensions where the region of support is a polytope [Bran 97], and applications in multi-dimensional signal processing have been described in [Lu 09].

It is very straightforward to apply this result to determine the Fourier transform of a rect function, and this is left as an exercise. The following shows the application to a regular hexagon with unit side.

Example 2.7. Let \mathcal{A} be a regular hexagon with unit side, as illustrated in Fig. 2.10. Since the hexagon is symmetric about the origin, we can apply Eq. (2.39). From the figure, we observe that

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \mathbf{a}_2 = \begin{bmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix} \quad \mathbf{a}_3 = \begin{bmatrix} -\frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix} \quad \mathbf{a}_4 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}.$$

By assumption, $d_k = 1$ for all k , and we can easily determine that

$$\mathbf{n}_1 = \begin{bmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix} \quad \mathbf{n}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \mathbf{n}_3 = \begin{bmatrix} -\frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix}.$$

Simply inserting these into Eq. (2.39) (and using symmetry of sinc) gives the result

$$P_{\mathcal{A}}(\mathbf{u}) = \frac{1}{\pi(u^2 + v^2)} \left(\frac{\sqrt{3}u + v}{2} \sin \left(\frac{\pi(3u + \sqrt{3}v)}{2} \right) \text{sinc} \left(\frac{-u + \sqrt{3}v}{2} \right) \right. \\ \left. + v \sin(\pi\sqrt{3}v) \text{sinc}(u) + \frac{-\sqrt{3}u + v}{2} \sin \left(\frac{\pi(-3u + \sqrt{3}v)}{2} \right) \text{sinc} \left(\frac{u + \sqrt{3}v}{2} \right) \right)$$

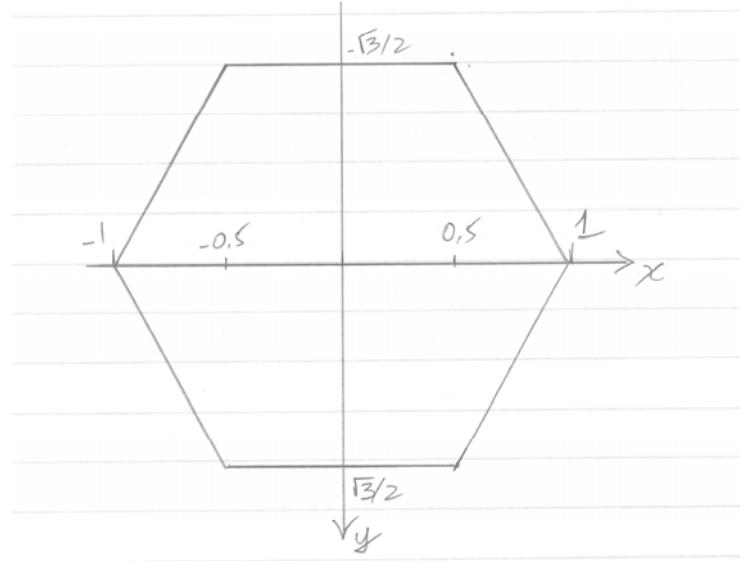


Figure 2.10: Regular hexagon with unit side.

2.6.3 Fourier transform of a translating still image

Assume that a still image $f_0(\mathbf{x})$ is moving with a uniform velocity ν to produce the time-varying image $f(\mathbf{x}, t) = f_0(\mathbf{x} - \nu t)$. We wish to relate the 3D Fourier transform

of $f(\mathbf{x}, t)$ to the 2D Fourier transform of $f_0(\mathbf{x})$.

$$\begin{aligned} F(\mathbf{u}, w) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_0(\mathbf{x} - \nu t) \exp(-j2\pi\mathbf{u} \cdot \mathbf{x}) d\mathbf{x} \exp(-j2\pi w t) dt \\ &= \int_{-\infty}^{\infty} \exp(-j2\pi w t) \exp(-j2\pi\mathbf{u} \cdot \nu t) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_0(\mathbf{s}) \exp(-j2\pi\mathbf{u} \cdot \mathbf{s}) d\mathbf{s} dt \\ &= F_0(\mathbf{u}) \int_{-\infty}^{\infty} \exp(-j2\pi(\mathbf{u} \cdot \nu + w)t) dt \\ &= F_0(\mathbf{u})\delta(\mathbf{u} \cdot \nu + w) \end{aligned}$$

Thus, the three-dimensional Fourier transform is concentrated on the plane $\mathbf{u} \cdot \nu + w = 0$. This leads us to conclude that the 3D Fourier transform of a typical time-varying image is not uniformly spread out in 3D frequency space, but will be largely concentrated near planes representing the dominant motions in the scene.

Problems

1. Consider a two-dimensional sinusoidal signal $f(x, y) = A \cos(2\pi(ux + vy) + \phi)$ where x and y are in ph and u and v are in c/ph. Form the one-dimensional signal $g(z)$ by tracing $f(x, y)$ along the line $y = cx$, where c is some real constant, as a function of distance along the line, $z = \sqrt{x^2 + y^2}$.
 - (a) Show that $g(z)$ is a sinusoidal signal $g(z) = A \cos(2\pi wz + \phi)$ and determine the spatial frequency w in c/ph, as a function of u , v and c .
 - (b) Explain what happens when $c = 0$ and when $c \rightarrow \infty$.
 - (c) Show that the spatial frequency w is greatest along the line $y = (v/u)x$, if $u \neq 0$. What is the value of this maximum spatial frequency? What happens if $u = 0$?
2. Show that for each of the following functions $\delta_{\Delta}(x, y)$,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta_{\Delta}(x, y) dx dy = 1$$

and

$$\lim_{\Delta \rightarrow 0} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta_{\Delta}(x, y) f(x, y) dx dy = f(0, 0)$$

for any function $f(x, y)$ that is continuous at $(x, y) = (0, 0)$.

- (a) $\delta_{\Delta}(x, y) = \frac{1}{\Delta^2} \text{rect}(x/\Delta, y/\Delta);$
- (b) $\delta_{\Delta}(x, y) = \frac{1}{\Delta^2} \exp(-\pi(x^2 + y^2)/\Delta^2);$
- (c) $\delta_{\Delta}(x, y) = \frac{1}{\pi\Delta^2} \text{circ}(x/\Delta, y/\Delta).$

3. Show that

$$\delta(ax, by) = \frac{1}{|ab|} \delta(x, y)$$

where $a, b \neq 0$.

4. Prove that the following systems are linear systems.

- (a) The shift system $\mathcal{T}_{\mathbf{d}}$ for any shift $\mathbf{d} \in \mathbb{R}^D$.
 - (b) The system induced by a nonsingular transformation of the domain, $\mathcal{M}_{\mathbf{A}} : g = \mathcal{M}_{\mathbf{A}} f : g(\mathbf{x}) = f(\mathbf{Ax})$, where \mathbf{A} is any nonsingular $D \times D$ matrix.
 - (c) The cascade of any two linear systems \mathcal{H}_1 and \mathcal{H}_2 . Thus, the system induced by an affine transformation of the domain is a linear system.
5. If \mathbf{A} is a $D \times D$ matrix and \mathbf{x}, \mathbf{y} are $D \times 1$ column matrices, show that $\mathbf{x} \cdot (\mathbf{Ay}) = (\mathbf{A}^T \mathbf{x}) \cdot \mathbf{y}$.
6. Prove the Fourier transform properties shown in Table 2.1.

7. Let $f(x, y) = 0.5 \text{rect}(4(x - 0.5), 2(y - 0.25))$ and $h(x, y) = \text{rect}(10x, 10y)$, where x and y are in ph.
- (a) Sketch the region of support of $f(x, y)$ and $h(x, y)$ in the XY-plane (i.e., the area where these two signals are nonzero).
 - (b) Compute the two-dimensional convolution $f(x, y) * h(x, y)$ from the definition using integration in the spatial domain.

- (c) Suppose that $f(x, y)$ is the input to a two-dimensional system, and the output of this system is computed as in (b). What can we say about this system?
- (d) Determine the continuous-space Fourier transforms $F(u, v)$, $H(u, v)$ and $G(u, v)$ of the above three signals. Make liberal use of Fourier transform properties. What are the units of u and v ?
- (e) Continuing with question (c), what is the interpretation of $H(u, v)$?

8. A two-dimensional continuous-space linear shift-invariant system has impulse response

$$h(x, y) = \begin{cases} \frac{1}{2\pi R_1 R_2}, & \left(\frac{x}{R_1}\right)^2 + \left(\frac{y}{R_2}\right)^2 \leq 1 \\ 0, & \text{otherwise,} \end{cases}$$

where $R_1 = 1/1000$ ph and $R_2 = 1/500$ ph.

- (a) Sketch the region of support of the impulse response in the XY-plane, following the conventions used in the course for the labelling of axes. Express $h(x, y)$ in terms of the circ function.
- (b) Find the frequency response $H(u, v)$ of this system, where u and v are in c/ph.
- (c) The image $f(x, y) = \text{rect}(5(x - .5), 2(y - .5))$ is filtered with this system to produce the output $g(x, y) = f(x, y)*h(x, y)$. Determine the Fourier transform of the output, $G(u, v)$.

9. Compute the two-dimensional continuous-space Fourier transform of the following signals:

- (a) The separable signal $f(x, y) = h_X^{(1)}(x)h_Y^{(1)}(y)$ where

$$h_T^{(1)}(t) = \begin{cases} 1 - \frac{|t|}{T} & |t| \leq T, \\ 0 & \text{otherwise.} \end{cases}$$

- (b) A Gaussian function $f(x, y) = \frac{1}{2\pi r_0^2} e^{-(x^2+y^2)/2r_0^2}$.

- (c) A real zoneplate, $f(x, y) = \cos(\pi(x^2 + y^2)/r_0^2)$. (Hint: Find the Fourier transform of the complex zoneplate $\exp(j\pi(x^2 + y^2)/r_0^2)$ and use linearity. You can use $\int_{-\infty}^{\infty} e^{jy^2} dy = \sqrt{\pi}e^{j\pi/4}$.)
- (d) Diamond-shaped pulse

$$f(x, y) = \begin{cases} 1 & |x| + |y| \leq 1, \\ 0 & |x| + |y| > 1. \end{cases}$$

(Hint: obtain this function from a rect function using a rotation transformation.)

- (e) Gabor function

$$f(x, y) = \cos(2\pi(u_0x + v_0y)) \exp\left(-\frac{(x - x_0)^2 + (y - y_0)^2}{2r_0^2}\right)$$

- (f) The two-dimensional zero-one function $p_{\mathcal{A}}(x, y)$ where \mathcal{A} is an elliptical region, with semi-minor axis X and semi-major axis $2X$, oriented at 45° as shown in the figure

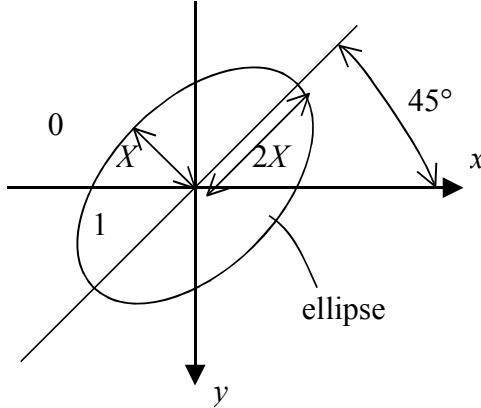


Figure 2.11: Elliptical region of support of two-dimensional zero-one function.

10. Derive the expression for the Fourier transform of a zero-one function on a polygon symmetric about the origin, as given in Eq. (2.39).
11. Use the expression in Eq. (2.39) to compute the Fourier transform of the rect function.
12. Use the expression in Eq. (2.39) to compute the Fourier transform of a zero-one function with a region \mathcal{A} that is a regular hexagon of unit area, with vertices on the y axis.
13. Use the expression in Eq. (2.39) to compute the Fourier transform of a zero-one function with a region \mathcal{A} that is a regular octagon of unit area, with two sides parallel to the x axis.

Chapter 3

Discrete-domain signals and systems

Although the original images projected on a sensor and the final images presented to a viewer's visual system are continuous in space and time, a discrete intermediate representation is required in order to carry out digital image and video processing operations. Thus images must be *sampled* in space and time for processing and eventually converted back to continuous form for presentation to the viewer. In this chapter we introduce discrete-space and discrete-space-time signals and systems, under the general name of discrete-domain signals and systems.

For the sampling of one-dimensional signals, all that needs to be specified is the sampling period (or equivalently, the sampling frequency), and possibly the sampling phase. In two or three dimensions, the situation is more complicated. We have to specify how the samples are arranged in space and time. The simplest arrangement is to lay out the samples on a rectangular grid, and this is the approach taken in most discussions of digital image processing. However, there are many important applications where this is not the case. For example, in standard broadcast television, the interlaced scanning method is used; scanning lines in each vertical pass of the image are midway between the scanning lines of the previous vertical pass. It follows that any scheme for sampling standard TV signals will lead to non-rectangular

sampling in 3D space-time. Another important example of non-rectangular sampling is the widely-used Bayer color filter array [Gunt 05]. Fig. 3.1 shows the layout of red, green and blue sensor elements in such an array. We see that the red, green and blue samples are each acquired on different sampling structures with different offsets, and that the green sampling structure is not rectangular. Our approach will be to develop sampling theory immediately in the general multidimensional setting which includes two and three-dimensional sampling on both rectangular and non-rectangular sampling structures. Rectangular sampling will be a special case. This is different from the usual approach of introducing rectangular sampling first, and then possibly extending the theory to non-rectangular sampling. The development will be based on the theory of *lattices*.

The application of lattices to the sampling of multidimensional signals was presented by Petersen and Middleton [Pete 62], and many of the basic results were presented in that work. Early applications of lattices to image and video sampling appeared in the late 1970s and early 1980s, some examples being [Ouel 81], [Kret 81], [Dubo 82]. The approach in these notes follows from the 1985 review of this topic in the Proceedings of the IEEE [Dubo 85]. The use of lattices for image and video sampling is discussed in several textbooks, including [Dudg 84], [Wang 02], [Wood 06]. An excellent treatment is also given in [Kalk 98]. Lattices have been successfully used in other aspects of information technology and communications. A review of lattice techniques in wireless communications is given in [Wubb 11] and applications in cryptography are given in [Micc 02], which also gives a very good, accessible overview of lattice definitions, properties and algorithms.

3.1 Lattices

3.1.1 Basic Definitions

The mathematical structure most useful in describing sampling of time-varying images is the *lattice*. A lattice Λ in D dimensions is a discrete set of points that can be expressed as the set of all linear combinations with *integer* coefficients of D linearly

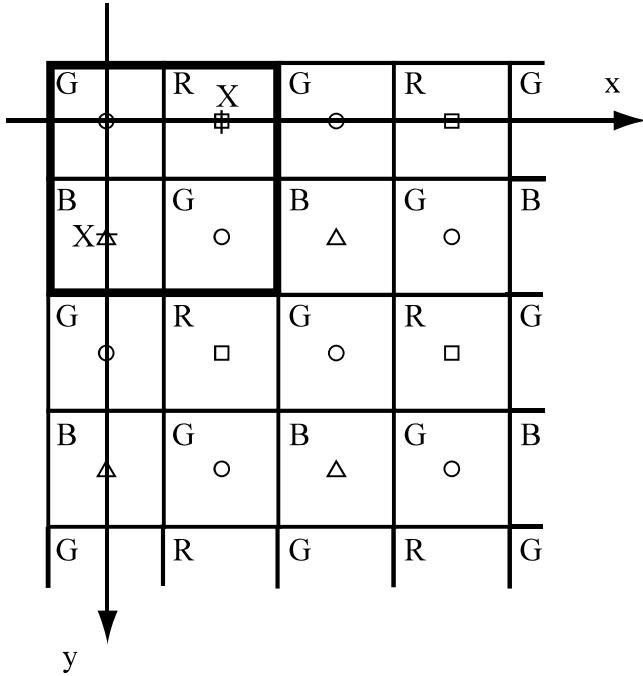


Figure 3.1: Bayer color filter array. The points labeled ‘G’ form a non-rectangular arrangements of samples.

independent vectors in \mathbb{R}^D (called basis vectors),

$$\Lambda = \{n_1 \mathbf{v}_1 + \cdots + n_D \mathbf{v}_D \mid n_i \in \mathbb{Z}\}, \quad (3.1)$$

where \mathbb{Z} is the set of integers. For our purposes, D will be 1, 2 or 3 dimensions. Fig. 3.2 shows an example of a lattice in two dimensions, with basis vectors $\mathbf{v}_1 = [2X \ 0]^T$ and $\mathbf{v}_2 = [X \ Y]^T$. A convenient way to represent a lattice is the *sampling matrix* $\mathbf{V} = [\mathbf{v}_1 \mid \mathbf{v}_2 \mid \cdots \mid \mathbf{v}_D]$ whose columns are the basis vectors \mathbf{v}_i represented as column matrices. We denote the lattice Λ determined by the sampling matrix \mathbf{V} by $\Lambda = \text{LAT}(\mathbf{V})$. The sampling matrix for the lattice of Fig 3.2 with respect to the

given basis vectors is

$$\mathbf{V} = \begin{bmatrix} 2X & X \\ 0 & Y \end{bmatrix}. \quad (3.2)$$

The basis or sampling matrix for any given lattice is not unique. For example, we can easily verify by inspection that the sampling matrix

$$\mathbf{V}_1 = \begin{bmatrix} X & -X \\ Y & Y \end{bmatrix}$$

also generates the lattice of Fig. 3.2. It can be shown that $\text{LAT}(\mathbf{V}) = \text{LAT}(\mathbf{VE})$ where \mathbf{E} is any unimodular ($|\det \mathbf{E}| = 1$) integer matrix. In the above example, $\mathbf{V}_1 = \mathbf{VE}$ where

$$\mathbf{E} = \mathbf{V}^{-1}\mathbf{V}_1 = \begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix}$$

and $\det \mathbf{E} = 1$. Alternatively, we can conclude that two sampling matrices \mathbf{V} and \mathbf{V}_1 represent the same lattice if and only if $\mathbf{V}^{-1}\mathbf{V}_1$ is an integer matrix with $|\det(\mathbf{V}^{-1}\mathbf{V}_1)| = 1$.

If we represent points in \mathbb{R}^D as column matrices, the points in the lattice are given by

$$\Lambda = \{\mathbf{V}\mathbf{n} \mid \mathbf{n} \in \mathbb{Z}^D\}.$$

For example, for the hexagonal lattice defined by the sampling matrix of equation 3.2, we have

$$\begin{aligned} \Lambda &= \left\{ \begin{bmatrix} 2Xn_1 + Xn_2 \\ Yn_2 \end{bmatrix} \mid n_1, n_2 \in \mathbb{Z} \right\} \\ &= \left\{ \begin{bmatrix} (2n_1 + n_2)X \\ n_2Y \end{bmatrix} \mid n_1, n_2 \in \mathbb{Z} \right\}. \end{aligned}$$

A *unit cell* of a lattice Λ is a set $\mathcal{P} \subset \mathbb{R}^D$ such that copies of \mathcal{P} centered on each lattice point tile the whole space without overlap: $(\mathcal{P} + \mathbf{s}_1) \cap (\mathcal{P} + \mathbf{s}_2) = \emptyset$ for $\mathbf{s}_1, \mathbf{s}_2 \in \Lambda$, $\mathbf{s}_1 \neq \mathbf{s}_2$, and $\cup_{\mathbf{s} \in \Lambda} (\mathcal{P} + \mathbf{s}) = \mathbb{R}^D$. The volume of a unit cell is $d(\Lambda) = |\det \mathbf{V}|$, which is independent of the particular choice of sampling matrix.

We can imagine that there is a region congruent to \mathcal{P} of volume $d(\Lambda)$ associated with each sample in Λ , so that $d(\Lambda)$ is the reciprocal of the sampling density. The quantity $d(\Lambda)$ is often referred to as the *determinant* of the lattice. The unit cell of a lattice is not unique. In Fig. 3.2, the shaded hexagonal region centered at the origin is a unit cell, of area $d(\Lambda) = 2XY$. The shaded parallelogram in the upper right is also a possible unit cell. The hexagonal-shaped unit cell is an example of a *Voronoi* unit cell, consisting of all points closer to the origin than to any other lattice point. In the two-dimensional case, the Voronoi unit cell is a polygon whose edges are the perpendicular bisectors of lines from the origin to the nearest lattice points. Note that we are usually informal about how we treat the boundary of the unit cell, since it is usually not very important, unless there is a singularity on the boundary.

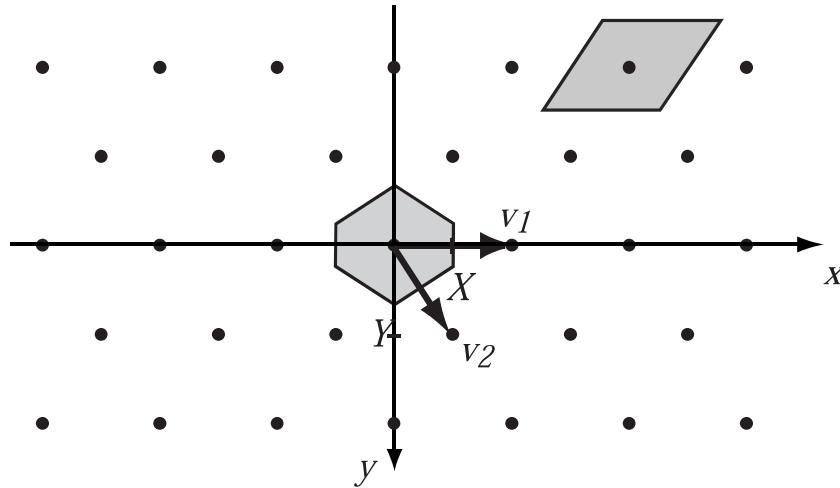


Figure 3.2: Example of a lattice in two dimensions with two possible unit cells.

3.1.2 Properties of Lattices

The following properties of lattices are easily seen from the definition and are key to the theory of processing signals defined on lattices. Let Λ be any lattice.

- (i) $\mathbf{0} \in \Lambda$; the origin belongs to *any* lattice.

- (ii) If $\mathbf{x} \in \Lambda$ and $\mathbf{y} \in \Lambda$ then $\mathbf{x} + \mathbf{y} \in \Lambda$.
- (iii) If $\mathbf{d} \in \Lambda$ then $\Lambda + \mathbf{d} = \Lambda$ where $\Lambda + \mathbf{d} = \{\mathbf{x} + \mathbf{d} \mid \mathbf{x} \in \Lambda\}$

The proofs are left as an exercise.

3.1.3 Examples of 2D and 3D Lattices

This section presents several lattices that have been used for image sampling.

2D rectangular (or orthogonal) lattice (Fig. 3.3).

The rectangular lattice is defined by perpendicular basis vectors, and thus can be generated by a diagonal sampling matrix.

$$\mathbf{V} = \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix}, \quad d(\Lambda) = XY.$$

If $X = Y$, we call it a square lattice.

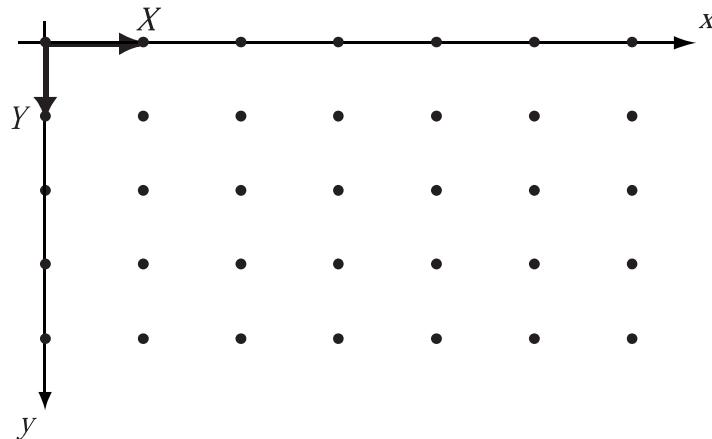


Figure 3.3: Two-dimensional rectangular lattice

2D hexagonal lattice (Fig. 3.4). The hexagonal lattice is so named because the six nearest neighbors of any lattice point form the vertices of a hexagon. The standard

form of the sampling matrix is

$$\mathbf{V} = \begin{bmatrix} X & X/2 \\ 0 & Y \end{bmatrix}, \quad d(\Lambda) = XY.$$

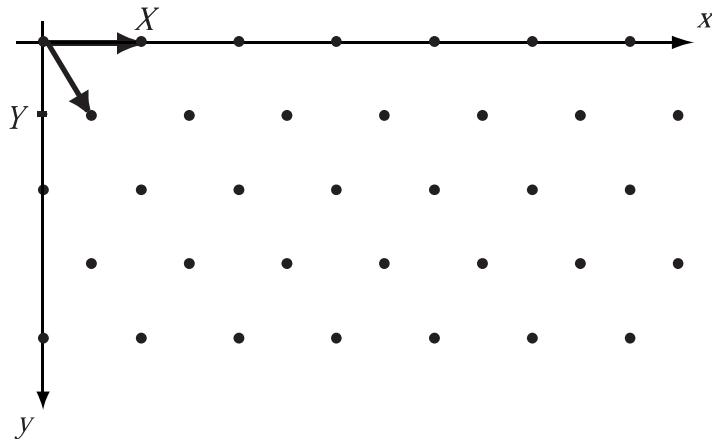


Figure 3.4: Two-dimensional hexagonal lattice

3D rectangular lattice (Fig. 3.5). The three-dimensional rectangular lattice is defined by three perpendicular basis vectors, and so is defined by a 3×3 diagonal matrix.

$$\mathbf{V} = \begin{bmatrix} X & 0 & 0 \\ 0 & Y & 0 \\ 0 & 0 & T \end{bmatrix}, \quad d(\Lambda) = XYT.$$

3D interlaced lattice (Fig. 3.6). The three-dimensional interlaced lattice arises when an interlaced signal like any standard television signal is sampled with all samples aligned vertically. The sampling matrix is

$$\mathbf{V} = \begin{bmatrix} X & 0 & 0 \\ 0 & 2Y & Y \\ 0 & 0 & T/2 \end{bmatrix}, \quad d(\Lambda) = XYT.$$

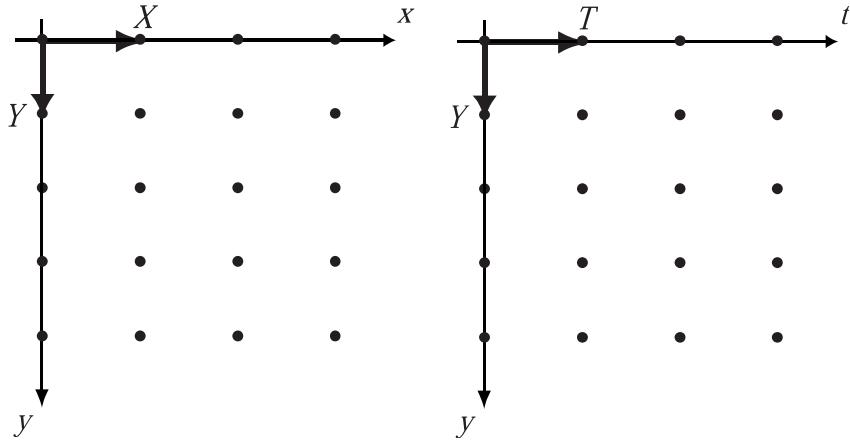


Figure 3.5: Three-dimensional rectangular lattice

3.2 Sampling Structures

Most sampling structures of interest for still and time-varying images can be constructed using a lattice. The sampling structure is the set of points in \mathbb{R}^D on which the image is defined. Let us consider still images first. The sampling structure determines how the sample points are laid out on the image plane. By far, the most widely used structure is the rectangular lattice of Fig. 3.3. However, the hexagonal lattice of Fig. 3.4 is also used in a number of important applications. The sampling structure of an image is normally confined to the image window \mathcal{W} of Fig. 2.1. However, for convenience of analysis, we often consider the sampling structure to be of infinite extent. Outside \mathcal{W} , the image can be considered to be zero, periodically extended, or extrapolated in some way. This has an impact on the performance of image processing algorithms near the boundary of the image, and may or may not be important.

If the sampling phase is not important, we can assume that one point of the sampling structure lies at the origin of the coordinate system and that the sampling structure is a lattice Λ . However, if we consider the Bayer color filter array of Fig. 3.1 to simultaneously sample the red, green and blue components of a color image, we see that the sampling structures for R, G and B have no points in common. Thus,

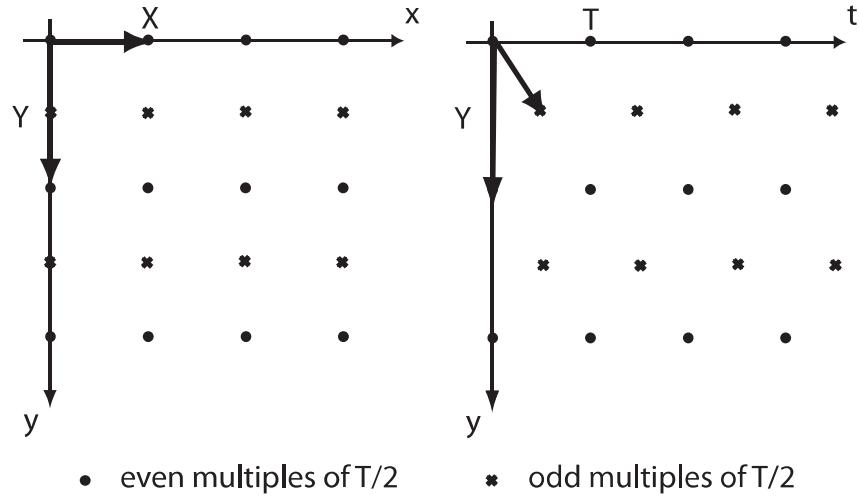


Figure 3.6: Three-dimensional interlaced lattice

only one of them can contain the origin and be a lattice. The other two are *shifted* lattices. Assume that sample locations correspond to the centers of the cells in Fig. 3.1, which are assumed to be squares of size X by X . The resulting sampling structures are indicated in Fig. 3.1.

The sampling structure for the green component, Ψ_G , contains the origin and is the hexagonal lattice

$$\Psi_G = \text{LAT} \left(\begin{bmatrix} 2X & X \\ 0 & X \end{bmatrix} \right).$$

The sampling structures for the red and blue channels are *shifted* rectangular lattices:

$$\begin{aligned} \Psi_R &= \begin{bmatrix} X \\ 0 \end{bmatrix} + \text{LAT} \left(\begin{bmatrix} 2X & 0 \\ 0 & 2X \end{bmatrix} \right) \\ \Psi_B &= \begin{bmatrix} 0 \\ X \end{bmatrix} + \text{LAT} \left(\begin{bmatrix} 2X & 0 \\ 0 & 2X \end{bmatrix} \right). \end{aligned}$$

Notice that the sampling structure for G has twice the sampling density as those for R and B ($\frac{1}{2X^2}$ versus $\frac{1}{4X^2}$). The sampling density of a shifted lattice $\mathbf{d} + \Lambda$ is $1/d(\Lambda)$, i.e., the shift does not change the sampling density.

It is possible to construct a sampling structure that is the union of two or more shifted lattices, and this has found a few applications, for example in some recent digital cameras.

Similarly, sampling structures for time-varying images can be constructed from three-dimensional lattices. The two most widely used correspond to the lattices of Fig. 3.5 and Fig. 3.6. The rectangular lattice would normally be used whenever motion-picture film is digitized, and is now used with many video sources. However, many video cameras today still use *interlaced* scanning; each image *frame* consists of two *fields*. Each vertical pass of the scanning beam captures half of the lines forming the entire frame. This scanning structure is used in the three main analog TV formats (NTSC, PAL, SECAM) as well as in some new HDTV formats. Typical parameters for the North American video signal in interlaced format in Fig. 3.6 are $X = \frac{1}{540}$ ph, $Y = \frac{1}{480}$ ph and $T = \frac{1}{29.97}$ s. The value of the parameter Y follows directly from the fact that there are 480 active scanning lines (out of a total of 525) forming an image frame.

It is possible to consider *partially* sampled signals. For example, a motion picture film sequence can be considered to be continuous in the spatial domain and discrete in the temporal domain.

3.3 Signals Defined on Lattices

The extension of discrete-time signal theory to two and three dimensions is accomplished by considering signals defined on lattices. In fact, ordinary discrete-time signals can be thought of as signals defined on the one-dimensional lattice $\text{LAT}([T])$. Let Λ be a lattice in D dimensions. Then a real signal defined on the lattice Λ is denoted $f[\mathbf{x}]$, $\mathbf{x} \in \Lambda$. The signal f is understood to only be defined on the points of Λ ; it is undefined elsewhere in \mathbb{R}^D (it is not 0). Note that, following the convention of Oppenheim and Willsky [Opp 97], we use *square* brackets to enclose the independent variables \mathbf{x} for signals defined on a lattice, as opposed to round parentheses for continuous-domain signals.

As in Section 2.4, we can consider a signal space \mathcal{S} of signals defined on Λ and

denote an element of this space f . For example, the space of signals of finite energy defined on Λ is

$$l^2(\Lambda) = \{f \mid \sum_{\mathbf{x} \in \Lambda} |f[\mathbf{x}]|^2 < \infty\}.$$

The notation $\sum_{\mathbf{x} \in \Lambda} \cdot$ indicates a summation over all elements of the lattice. Using the definition of a lattice generated by the basis vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_D\}$, this summation can be written explicitly as

$$\sum_{\mathbf{x} \in \Lambda} |f[\mathbf{x}]|^2 = \sum_{n_1=-\infty}^{\infty} \cdots \sum_{n_D=-\infty}^{\infty} |f[n_1 \mathbf{v}_1 + \cdots + n_D \mathbf{v}_D]|^2.$$

In the most common case of a rectangular lattice, a signal is often written explicitly in terms of its coordinates $f[n_1 X, n_2 Y]$ or $f[n_1 X, n_2 Y, n_3 T]$, and the summation is expressed, in the 3D case, as

$$\sum_{\mathbf{x} \in \Lambda} |f[\mathbf{x}]|^2 = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} \sum_{n_3=-\infty}^{\infty} |f[n_1 X, n_2 Y, n_3 T]|^2.$$

In the case of a *square* spatial lattice with horizontal and vertical spacing X , it is common to choose X as the unit of length. We shall call this length the pixel height, denoted px . In this case, the sampling matrix is the identity matrix $\mathbf{V} = \text{diag}(1, 1)$ and $\Lambda = \mathbb{Z}^2$.

3.4 Special Multidimensional Signals on a Lattice

As in continuous space-time, many analytically defined two and three-dimensional functions over a lattice are useful in image processing theory. A number of these are described here.

3.4.1 Unit sample

The unit sample function is defined as

$$\delta[\mathbf{x}] = \begin{cases} 1 & \text{if } \mathbf{x} = \mathbf{0} \\ 0 & \text{if } \mathbf{x} \in \Lambda \setminus \mathbf{0}. \end{cases} \quad (3.3)$$

The notation $\mathbf{x} \in \Lambda \setminus \mathbf{0}$ signifies that \mathbf{x} is any element of Λ except $\mathbf{0}$; we know that $\mathbf{0}$ is an element of *any* lattice. The unit sample has many characteristics in common with the Dirac delta function $\delta(\mathbf{x})$ of continuous space-time. However, whereas the latter is a rather complex generalized function, the unit sample is probably the simplest of signals over a lattice. It will be key in the characterization of linear shift-invariant systems over a lattice. Note that we use the same symbol for the unit sample and the Dirac function; they are distinguished however by the use of square brackets versus parentheses, and the particular lattice Λ is assumed.

3.4.2 Sinusoidal signals

The real and complex forms of a sinusoidal signal defined on a lattice are

$$\begin{aligned} f_r[\mathbf{x}] &= A \cos(2\pi \mathbf{u} \cdot \mathbf{x} + \phi), \\ f_c[\mathbf{x}] &= C \exp(j2\pi \mathbf{u} \cdot \mathbf{x}), \quad \mathbf{x} \in \Lambda, \end{aligned}$$

where as before \mathbf{u} is a two or three-dimensional frequency vector. We recall that for one-dimensional discrete-time signals, different frequencies can give the same discrete-time sinusoid. Specifically,

$$\begin{aligned} f_1[nT] &= C \exp(j2\pi u n T) \\ \text{and } f_2[nT] &= C \exp(j2\pi(u + \frac{k}{T})n T) \end{aligned}$$

are the same discrete-time sinusoid for any integer k . Thus all frequencies $u + \frac{k}{T}$, $k \in \mathbb{Z}$, define the same discrete-time sinusoid. There is a similar property for multidimensional sinusoids on a lattice. Notice that the one-dimensional signal is defined on the 1D lattice $\Lambda = \{nT \mid n \in \mathbb{Z}\}$. Also notice that the frequencies u and $u + r$ define the same discrete-time sinusoid if and only if $r \in \{k \cdot \frac{1}{T} \mid k \in \mathbb{Z}\}$. We see that this is also a lattice that we call the reciprocal lattice Λ^* . This is the notion that extends in a straightforward fashion to higher-dimensional lattices.

The two multidimensional sinusoidal signals

$$\begin{aligned} f_1[\mathbf{x}] &= C \exp(j2\pi \mathbf{u} \cdot \mathbf{x}) \\ \text{and } f_2[\mathbf{x}] &= C \exp(j2\pi(\mathbf{u} + \mathbf{r}) \cdot \mathbf{x}) \end{aligned}$$

will be identical if $\mathbf{r} \cdot \mathbf{x}$ is an integer for all $\mathbf{x} \in \Lambda$. The set of all such points \mathbf{r} is called the reciprocal lattice:

$$\Lambda^* = \{\mathbf{r} \in \mathbb{R}^D \mid \mathbf{r} \cdot \mathbf{x} \in \mathbb{Z} \text{ for all } \mathbf{x} \in \Lambda\}. \quad (3.4)$$

Theorem 3.1. If $\Lambda = \text{LAT}(\mathbf{V})$, then

$$\{\mathbf{r} \mid \mathbf{r} \cdot \mathbf{x} \in \mathbb{Z} \text{ for all } \mathbf{x} \in \Lambda\} = \text{LAT}(\mathbf{V}^{-T}).$$

Proof. Let $\mathcal{R} = \{\mathbf{r} \mid \mathbf{r} \cdot \mathbf{x} \in \mathbb{Z} \text{ for all } \mathbf{x} \in \Lambda\}$. Suppose that $\mathbf{r} \in \text{LAT}(\mathbf{V}^{-T})$. Then $\mathbf{r} = \mathbf{V}^{-T}\mathbf{k}$ for some integer vector $\mathbf{k} \in \mathbb{Z}^D$. Thus,

$$\mathbf{r} \cdot \mathbf{x} = \mathbf{k}^T \mathbf{V}^{-1} \mathbf{V} \mathbf{x} = \mathbf{k}^T \mathbf{x} \in \mathbb{Z}$$

so that $\mathbf{r} \in \mathcal{R}$ and we can conclude that $\text{LAT}(\mathbf{V}^{-T}) \subset \mathcal{R}$.

Now suppose that $\mathbf{r} \in \mathcal{R}$. Let $\alpha = \mathbf{V}^T \mathbf{r}$. Then

$$\mathbf{r} \cdot \mathbf{x} = \alpha^T \mathbf{V}^{-1} \mathbf{V} \mathbf{x} = \alpha^T \mathbf{x}.$$

This must be an integer for all possible $\mathbf{x} \in \Lambda$, and therefore for all possible $\mathbf{n} \in \mathbb{Z}^D$. In particular, if \mathbf{n} is 1 in position k and zero elsewhere, then it follows that $\alpha_k \in \mathbb{Z}$ for any k , i.e., $\alpha \in \mathbb{Z}^D$ and so $\mathbf{r} \in \text{LAT}(\mathbf{V}^{-T})$. Thus $\mathcal{R} \subset \text{LAT}(\mathbf{V}^{-T})$, and combining with the first result, $\mathcal{R} = \text{LAT}(\mathbf{V}^{-T})$. \square

The significance of the reciprocal lattice will become more apparent when we introduce the Fourier transform of signals defined on a lattice. Note that

$$\begin{aligned} d(\Lambda^*) &= |\det(\mathbf{V}^{-T})| = \frac{1}{|\det(\mathbf{V})|} \\ &= \frac{1}{d(\Lambda)} \end{aligned}$$

using standard properties of determinants. Thus, the density of the reciprocal lattice is the reciprocal of the density of the original lattice.

For the 2D examples presented previously, we have:

rectangular lattice

$$\mathbf{V} = \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix} \quad d(\Lambda) = XY$$

$$\mathbf{V}^{-T} = \begin{bmatrix} \frac{1}{X} & 0 \\ 0 & \frac{1}{Y} \end{bmatrix} \quad d(\Lambda^*) = \frac{1}{XY}$$

hexagonal lattice

$$\mathbf{V} = \begin{bmatrix} X & \frac{X}{2} \\ 0 & Y \end{bmatrix} \quad d(\Lambda) = XY$$

$$\mathbf{V}^{-T} = \begin{bmatrix} \frac{1}{X} & 0 \\ -\frac{1}{2Y} & \frac{1}{Y} \end{bmatrix} \quad d(\Lambda^*) = \frac{1}{XY}$$

It follows that a set of frequencies corresponding to *distinct* sinusoidal signals on a lattice Λ forms a unit cell of the reciprocal lattice Λ^* . For the hexagonal lattice we have considered, this is illustrated in Fig. 3.7. For consistency with our convention in the spatial domain, the vertical frequency axis is oriented downward.

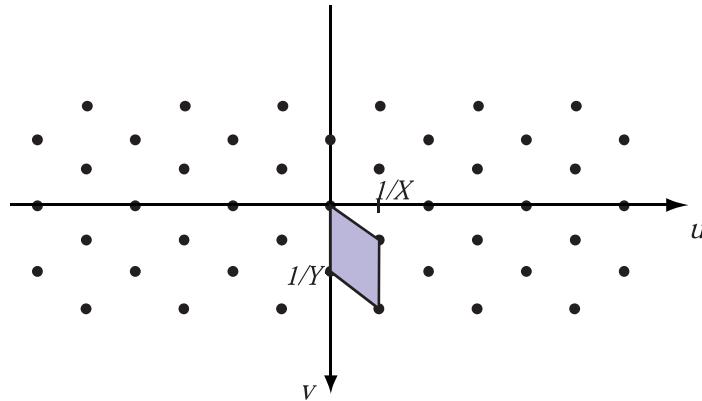


Figure 3.7: Reciprocal lattice for the hexagonal lattice with a possible unit cell (shaded).

3.5 Linear Systems over Lattices

We now extend the familiar concept of linear systems to signals defined on a lattice. The notation of Section 2.5 for continuous space-time linear systems can be used largely unchanged. Thus, \mathcal{S} denotes a vector space of signals defined on a lattice Λ and \mathcal{H} is a system mapping signals in \mathcal{S} to new signals in \mathcal{S} . The conditions for linearity are identical to the ones given in Eq. (2.20) and Eq. (2.21).

Similarly, we can define the shift operator

$$\mathcal{T}_d : g = \mathcal{T}_d f : g[\mathbf{x}] = f[\mathbf{x} - \mathbf{d}], \quad \mathbf{d} \in \Lambda,$$

where we impose the constraint that the shift \mathbf{d} is an element of the lattice Λ . This ensures that for every $\mathbf{x} \in \Lambda$, $\mathbf{x} - \mathbf{d}$ is also in Λ and so the signal with values $f[\mathbf{x} - \mathbf{d}]$ is well defined. The shift operator is again a linear system. The definition of a linear shift-invariant system is also unchanged: \mathcal{H} is shift-invariant if $\mathcal{H}(\mathcal{T}_d f) = \mathcal{T}_d(\mathcal{H}f)$ for any shift $\mathbf{d} \in \Lambda$.

3.5.1 Response of a Linear System

The response of a linear system to an arbitrary input is found in a similar fashion to continuous space-time systems. An arbitrary signal defined on the lattice Λ can be expressed

$$f = \sum_{\mathbf{y} \in \Lambda} f[\mathbf{y}] \mathcal{T}_{\mathbf{y}} \delta$$

or alternatively

$$f[\mathbf{x}] = \sum_{\mathbf{y} \in \Lambda} f[\mathbf{y}] \delta[\mathbf{x} - \mathbf{y}].$$

Applying linearity, for a linear system we have

$$\begin{aligned} \mathcal{H}f &= \sum_{\mathbf{y} \in \Lambda} f[\mathbf{y}] \mathcal{H}(\mathcal{T}_{\mathbf{y}} \delta) \\ &= \sum_{\mathbf{y} \in \Lambda} f[\mathbf{y}] h_{\mathbf{y}} \end{aligned}$$

where $h_{\mathbf{y}}$ is the response of the linear system to a unit sample at position \mathbf{y} .

For a linear shift-invariant system, we have $\mathcal{H}\mathcal{T}_y = \mathcal{T}_y\mathcal{H}$, and so

$$\mathcal{H}(\mathcal{T}_y\delta) = \mathcal{T}_y\mathcal{H}\delta = \mathcal{T}_y h$$

where h is the unit sample response of the system. Written explicitly, if $g = \mathcal{H}f$, then

$$g[\mathbf{x}] = \sum_{\mathbf{y} \in \Lambda} f[\mathbf{y}]h[\mathbf{x} - \mathbf{y}].$$

This expression defines *convolution* over a lattice, which we denote $g = f * h$. Once again, convolution is commutative, $f * h = h * f$, and so we also have

$$g[\mathbf{x}] = \sum_{\mathbf{y} \in \Lambda} h[\mathbf{y}]f[\mathbf{x} - \mathbf{y}].$$

3.5.2 Frequency Response

Let the input to an LSI system \mathcal{H} be a complex sinusoid of frequency \mathbf{u} ,

$$f[\mathbf{x}] = \exp(j2\pi\mathbf{u} \cdot \mathbf{x}), \quad \mathbf{x} \in \Lambda.$$

Then

$$\begin{aligned} g[\mathbf{x}] &= \sum_{\mathbf{y} \in \Lambda} h[\mathbf{y}] \exp(j2\pi\mathbf{u} \cdot (\mathbf{x} - \mathbf{y})) \\ &= \exp(j2\pi\mathbf{u} \cdot \mathbf{x}) \sum_{\mathbf{y} \in \Lambda} h[\mathbf{y}] \exp(-j2\pi\mathbf{u} \cdot \mathbf{y}) \\ &= H(\mathbf{u}) \exp(j2\pi\mathbf{u} \cdot \mathbf{x}), \end{aligned}$$

where

$$H(\mathbf{u}) = \sum_{\mathbf{y} \in \Lambda} h[\mathbf{y}] \exp(-j2\pi\mathbf{u} \cdot \mathbf{y})$$

(assuming that the sum converges). Once again, the complex sinusoid is an eigenfunction of an LSI system over a lattice, with eigenvalue $H(\mathbf{u})$. We call $H(\mathbf{u})$, as a function of the frequency vector \mathbf{u} , the *frequency response* of the system. It is also the discrete space(-time) Fourier transform over the lattice Λ of the unit sample response $h[\mathbf{x}]$.

3.6 Discrete-domain Fourier transforms over a lattice

Let $f[\mathbf{x}]$ be defined on the lattice Λ . We define the discrete-domain Fourier transform of $f[\mathbf{x}]$ by

$$F(\mathbf{u}) = \sum_{\mathbf{x} \in \Lambda} f[\mathbf{x}] \exp(-j2\pi\mathbf{u} \cdot \mathbf{x}). \quad (3.5)$$

The property that $\exp(-j2\pi\mathbf{u} \cdot \mathbf{x})$ and $\exp(-j2\pi(\mathbf{u} + \mathbf{r}) \cdot \mathbf{x})$ are the same signal if $\mathbf{r} \in \Lambda^*$ determines the periodicity property of the Fourier transform on Λ :

$$F(\mathbf{u}) = F(\mathbf{u} + \mathbf{r}) \quad \text{for all } \mathbf{r} \in \Lambda^*.$$

We say that $F(\mathbf{u})$ is periodic with periodicity lattice Λ^* . It follows that $F(\mathbf{u})$ is completely determined by its values over one unit cell of Λ^* , that we denote \mathcal{P}^* (or sometimes \mathcal{P}_{Λ^*}).

It can be shown that the inverse Fourier transform is given by

$$\begin{aligned} f[\mathbf{x}] &= \frac{1}{|\mathcal{P}^*|} \int_{\mathcal{P}^*} F(\mathbf{u}) \exp(j2\pi\mathbf{u} \cdot \mathbf{x}) d\mathbf{u} \\ &= d(\Lambda) \int_{\mathcal{P}^*} F(\mathbf{u}) \exp(j2\pi\mathbf{u} \cdot \mathbf{x}) d\mathbf{u}. \end{aligned}$$

The notation $\int_{\mathcal{P}^*}$ means integration over one unit cell of the reciprocal lattice Λ^* .

Consider the most common case of a 2D rectangular lattice

$$\Lambda = \text{LAT} \left(\begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix} \right), \quad \Lambda^* = \text{LAT} \left(\begin{bmatrix} \frac{1}{X} & 0 \\ 0 & \frac{1}{Y} \end{bmatrix} \right).$$

Then

$$F(u, v) = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} f[n_1 X, n_2 Y] \exp(-j2\pi(un_1 X + vn_2 Y)).$$

A suitable unit cell of Λ^* is

$$\mathcal{P}^* = \{(u, v) \mid -\frac{1}{2X} \leq u \leq \frac{1}{2X}, -\frac{1}{2Y} \leq v \leq \frac{1}{2Y}\}$$

and so the inverse Fourier transform can be written

$$f[n_1 X, n_2 Y] = XY \int_{-\frac{1}{2X}}^{\frac{1}{2X}} \int_{-\frac{1}{2Y}}^{\frac{1}{2Y}} F(u, v) \exp(j2\pi(un_1 X + vn_2 Y)) du dv.$$

It is common in the literature to write

$$f'[n_1, n_2] = f[n_1 X, n_2 Y]$$

and to use the normalized frequency variables $\omega_1 = 2\pi uX$ and $\omega_2 = 2\pi vX$. We then find

$$\begin{aligned} F'(\omega_1, \omega_2) &= F\left(\frac{\omega_1}{2\pi X}, \frac{\omega_2}{2\pi Y}\right) \\ &= \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} f'[n_1, n_2] \exp(-j(\omega_1 n_1 + \omega_2 n_2)) \\ &\quad -\pi \leq \omega_1, \omega_2 \leq \pi. \end{aligned}$$

Although this simplifies some expressions, we lose the physical dimensions of space and spatial frequency. Furthermore, if $X \neq Y$ as is often the case (e.g., television), we also introduce geometric distortion. For these reasons, we prefer the non-normalized definition of the Fourier transform.

3.6.1 Properties of the multidimensional Fourier transform over a lattice Λ

The Fourier transform over a lattice has very similar properties to the continuous-domain Fourier transform, as well as to the one-dimensional discrete-time Fourier transform. In fact, these are all just different versions of properties of a general Fourier transform in an algebraic setting. Table 3.1 lists a few of these properties.

As an example of how to prove these properties, we give a proof of the multiplication property (v). The proofs of the other properties are left as an exercise for the reader.

Multiplication property (v) If $g[\mathbf{x}] = f_1[\mathbf{x}]f_2[\mathbf{x}]$, then

$$G(\mathbf{u}) = d(\Lambda) \int_{\mathcal{P}^*} F_1(\mathbf{r})F_2(\mathbf{u} - \mathbf{r}) d\mathbf{r}.$$

$f[\mathbf{x}] = d(\Lambda) \int_{\mathcal{P}^*} F(\mathbf{u}) \exp(j2\pi\mathbf{u} \cdot \mathbf{x}) d\mathbf{u}$	$F(\mathbf{u}) = \sum_{\mathbf{x} \in \Lambda} f[\mathbf{x}] \exp(-j2\pi\mathbf{u} \cdot \mathbf{x})$
(i) $af_1[\mathbf{x}] + bf_2[\mathbf{x}]$	$aF_1(\mathbf{u}) + bF_2(\mathbf{u})$
(ii) $f[\mathbf{x} - \mathbf{x}_0]$	$F(\mathbf{u}) \exp(-j2\pi\mathbf{u} \cdot \mathbf{x}_0)$
(iii) $f[\mathbf{x}] \exp(j2\pi\mathbf{u}_0 \cdot \mathbf{x})$	$F(\mathbf{u} - \mathbf{u}_0)$
(iv) $f_1[\mathbf{x}] * f_2[\mathbf{x}]$	$F_1(\mathbf{u})F_2(\mathbf{u})$
(v) $f_1[\mathbf{x}]f_2[\mathbf{x}]$	$d(\Lambda) \int_{\mathcal{P}^*} F_1(\mathbf{r})F_2(\mathbf{u} - \mathbf{r}) d\mathbf{r}$
(vi) $\mathbf{x}f[\mathbf{x}]$	$\frac{j}{2\pi} \nabla_{\mathbf{u}} F(\mathbf{u})$
(vii) $f^*[\mathbf{x}]$	$F^*(-\mathbf{u})$
(viii)	$\sum_{\mathbf{x} \in \Lambda} f[\mathbf{x}] ^2 = d(\Lambda) \int_{\mathcal{P}^*} F(\mathbf{u}) ^2 d\mathbf{u}$

Table 3.1: Properties of the multidimensional Fourier transform over a lattice Λ .

Proof.

$$\begin{aligned}
G(\mathbf{u}) &= \sum_{\mathbf{x} \in \Lambda} f_1[\mathbf{x}]f_2[\mathbf{x}] \exp(-j2\pi\mathbf{u} \cdot \mathbf{x}) \\
&= \sum_{\mathbf{x} \in \Lambda} \left(d(\Lambda) \int_{\mathcal{P}^*} F_1(\mathbf{r}) \exp(j2\pi\mathbf{r} \cdot \mathbf{x}) d\mathbf{r} \right) f_2[\mathbf{x}] \exp(-j2\pi\mathbf{u} \cdot \mathbf{x}) \\
&= d(\Lambda) \int_{\mathcal{P}^*} F_1(\mathbf{r}) \left(\sum_{\mathbf{x} \in \Lambda} f_2[\mathbf{x}] \exp(-j2\pi(\mathbf{u} - \mathbf{r}) \cdot \mathbf{x}) \right) d\mathbf{r} \\
&= d(\Lambda) \int_{\mathcal{P}^*} F_1(\mathbf{r})F_2(\mathbf{u} - \mathbf{r}) d\mathbf{r}.
\end{aligned}$$

□

3.7 Finite Impulse Response (FIR) Filters

A multidimensional filter over a lattice Λ is said to be a finite impulse filter if its unit sample response $h[\mathbf{x}]$ has a finite number of non-zero values, i.e., the set

$$\mathcal{B} = \{\mathbf{x} \in \Lambda \mid h[\mathbf{x}] \neq 0\}$$

has a finite number of elements. In this case

$$H(\mathbf{u}) = \sum_{\mathbf{x} \in \mathcal{B}} h[\mathbf{x}] \exp(-j2\pi \mathbf{u} \cdot \mathbf{x}).$$

The set \mathcal{B} is called the region of support of the FIR filter. If \mathcal{B} has an infinite number of elements, the filter is called infinite impulse response (IIR). In image processing, FIR filters are used almost exclusively, mainly because they do not have stability problems and they can be designed to have linear phase, or more often, zero phase.

Example 3.1. Let $\Lambda = \text{LAT} \left(\begin{bmatrix} X & 0 \\ 0 & X \end{bmatrix} \right)$ and

$$h[n_1X, n_2X] = \begin{cases} \frac{1}{2} & \text{if } n_1 = n_2 = 0; \\ \frac{1}{8} & \text{if } (n_1, n_2) = (\pm 1, 0) \text{ or } (0, \pm 1); \\ 0 & \text{otherwise.} \end{cases}$$

The unit sample response is illustrated in Fig. 3.8.

If the input to the filter is $f[\mathbf{x}]$, the output $g[\mathbf{x}]$ can be computed by direct evaluation of the convolution

$$g[\mathbf{x}] = \sum_{\mathbf{s} \in \Lambda} h[\mathbf{s}] f[\mathbf{x} - \mathbf{s}]$$

where there are only five nonzero terms in the summation. Explicitly

$$\begin{aligned} g[x, y] = & \frac{1}{2}f[x, y] + \frac{1}{8}f[x - X, y] + \frac{1}{8}f[x + X, y] + \frac{1}{8}f[x, y - X] \\ & + \frac{1}{8}f[x, y + X], \quad (x, y) \in \Lambda \end{aligned}$$

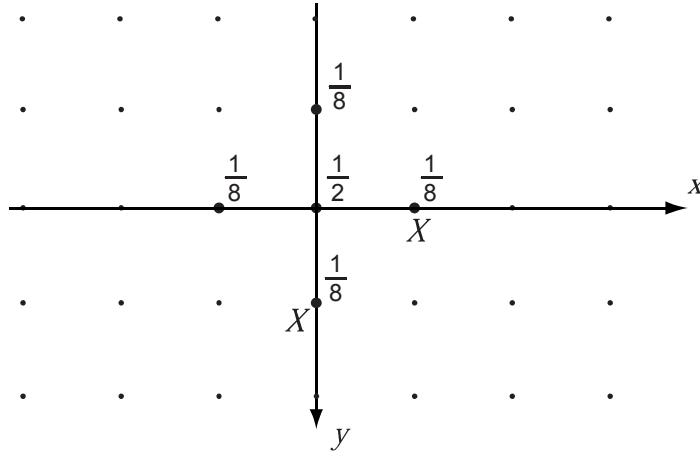


Figure 3.8: Unit sample response of a simple two-dimensional FIR filter. Small dots have a value of 0.0.

or alternatively,

$$\begin{aligned} g[n_1X, n_2X] &= \frac{1}{2}f[n_1X, n_2X] + \frac{1}{8}f[(n_1 - 1)X, n_2X] + \frac{1}{8}f[(n_1 + 1)X, n_2X] \\ &\quad + \frac{1}{8}f[n_1X, (n_2 - 1)X] + \frac{1}{8}f[n_1X, (n_2 + 1)X], \\ (n_1, n_2) &\in \mathbb{Z}^2. \end{aligned}$$

The frequency response of this filter is

$$H(\mathbf{u}) = \sum_{\mathbf{x} \in \Lambda} h[\mathbf{x}] \exp(-j2\pi \mathbf{u} \cdot \mathbf{x})$$

or alternatively

$$H(u, v) = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} h[n_1X, n_2X] \exp(-j2\pi(un_1X + vn_2X))$$

where there are only five nonzero terms in the double summation. Thus,

$$\begin{aligned} H(u, v) &= \frac{1}{2} + \frac{1}{8} \exp(-j2\pi uX) + \frac{1}{8} \exp(j2\pi uX) + \frac{1}{8} \exp(-j2\pi vX) + \frac{1}{8} \exp(j2\pi vX) \\ &= \frac{1}{2} + \frac{1}{4} \cos(2\pi uX) + \frac{1}{4} \cos(2\pi vX). \end{aligned}$$

Note that $H(0, 0) = 1$. A constant input is unchanged by this filter; we say that the *DC gain* of the filter is one. Also note that $H(\frac{1}{2X}, \frac{1}{2X}) = 0$. The input

$f[n_1 X, n_2 X] = \cos(\pi(n_1 + n_2)) = (-1)^{n_1+n_2}$ yields an identically zero output. The frequency response of this filter is real and positive, so $|H(u, v)| = H(u, v)$ and $H(u, v) = 0$. We call this a zero-phase filter.

Any real FIR filter satisfying the symmetry condition $h[\mathbf{x}] = h[-\mathbf{x}]$ will have a purely real frequency response, since combining terms in expression for the frequency response

$$\begin{aligned} h[\mathbf{x}] \exp(-j2\pi\mathbf{u} \cdot \mathbf{x}) + h[-\mathbf{x}] \exp(-j2\pi\mathbf{u} \cdot (-\mathbf{x})) \\ = h[\mathbf{x}] \exp(-j2\pi\mathbf{u} \cdot \mathbf{x}) + h[\mathbf{x}] \exp(j2\pi\mathbf{u} \cdot \mathbf{x}) \\ = 2h[\mathbf{x}] \cos(2\pi\mathbf{u} \cdot \mathbf{x}) \end{aligned}$$

which is real, so the overall frequency response is real.

Example 3.2. Another example of a 2D FIR filter on the same sampling lattice has unit sample response

$$h[n_1 X, n_2 X] = \begin{matrix} & -2 & -1 & 0 & 1 & 2 & n_1 \\ -2 & \left[\begin{matrix} 1 & 1 & 1 & 1 & 1 \end{matrix} \right] \\ -1 & \left[\begin{matrix} 1 & 2 & 2 & 2 & 1 \end{matrix} \right] \\ 0 & \left[\begin{matrix} 1 & 2 & 3 & 2 & 1 \end{matrix} \right] /35 \\ 1 & \left[\begin{matrix} 1 & 2 & 2 & 2 & 1 \end{matrix} \right] \\ 2 & \left[\begin{matrix} 1 & 1 & 1 & 1 & 1 \end{matrix} \right] \\ n_2 & \end{matrix}$$

The frequency response of this filter is shown in perspective and contour views in Fig. 3.9. The response is illustrated over one unit cell of the reciprocal lattice, namely

$$\mathcal{P}^* = \{(u, v) \mid -\frac{1}{2X} \leq u, v \leq \frac{1}{2X}\}.$$

Fig. 3.10 shows the result of filtering a zoneplate test pattern with this filter, while the result of filtering the Barbara image is shown in Fig. 3.11. The low-pass characteristic of the filter can be clearly seen from the result on the zoneplate. The result on the Barbara image demonstrates the blurring effect associated with a low-pass filter.

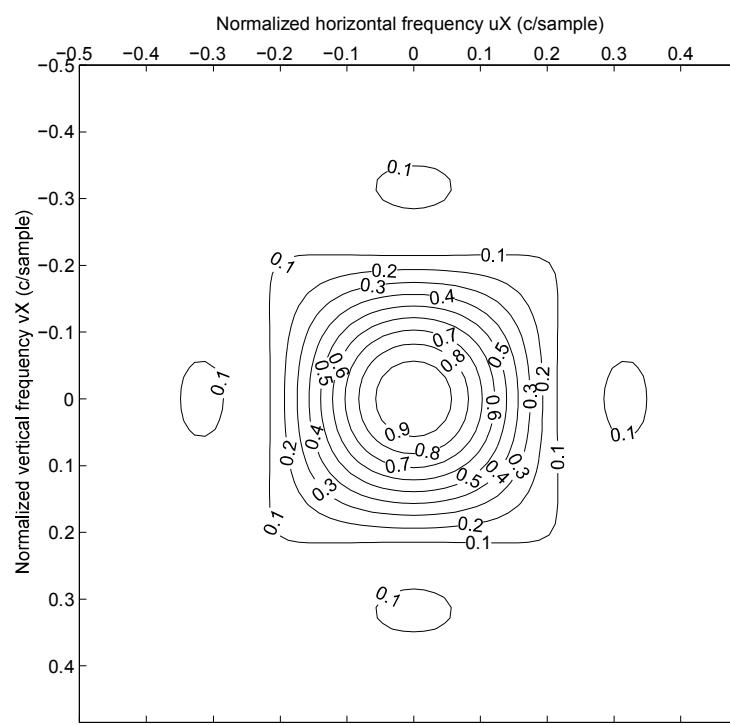
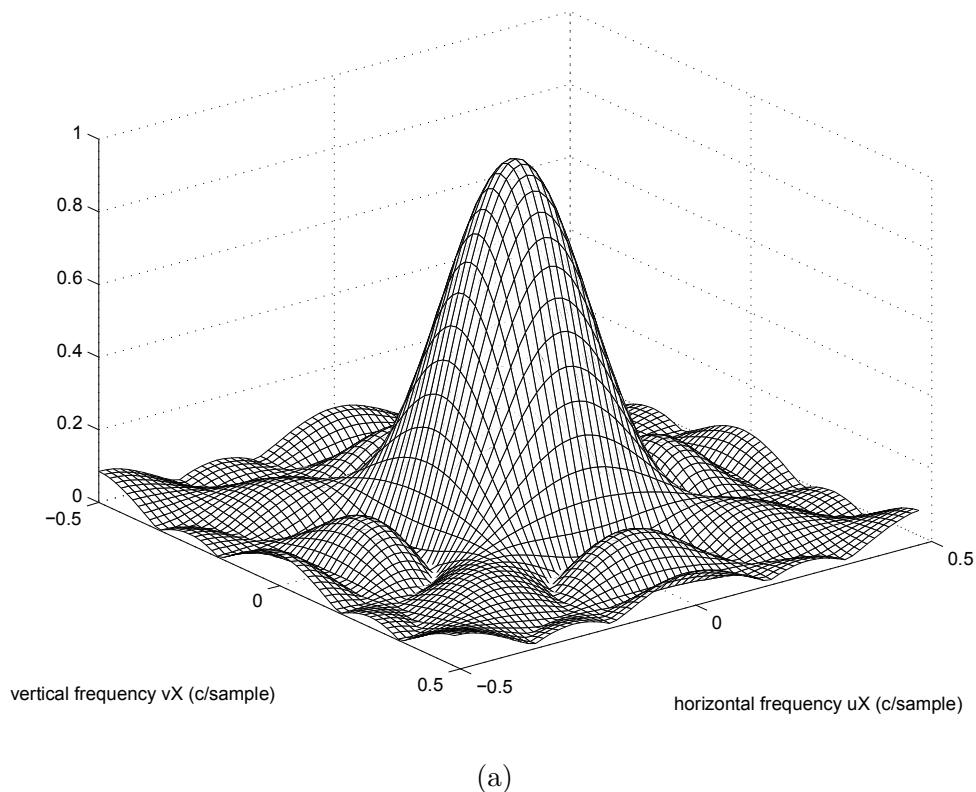
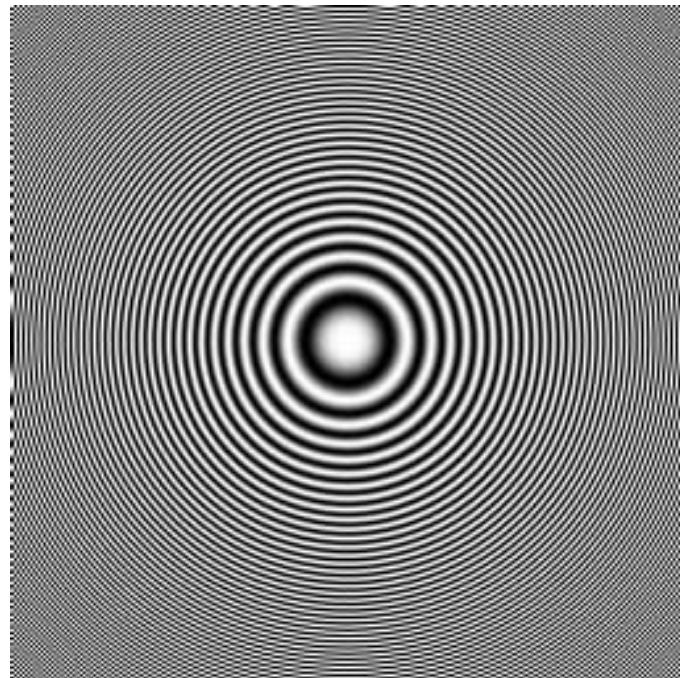
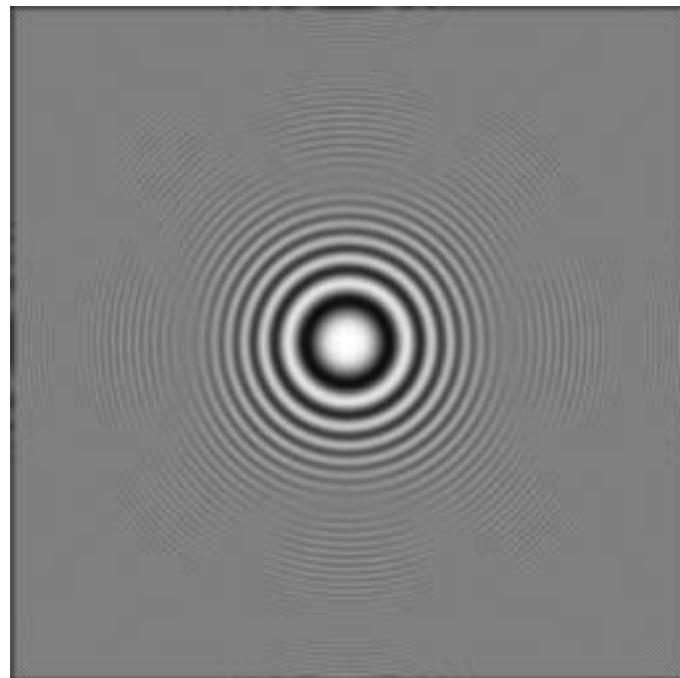


Figure 3.9: Frequency response of FIR filter. (a) Perspective view. (b) Contour plot.



(a)



(b)

Figure 3.10: Result of filtering zoneplate test pattern with FIR filter (a) Original zoneplate. (b) Filtered zoneplate.



(a)



(b)

Figure 3.11: Result of filtering Barbara image with FIR filter
(a) Original image.
(b) Filtered image.

3.7.1 Separable filters

We consider here filters defined on a 2D rectangular lattice $\Lambda = \text{LAT}([\begin{smallmatrix} X & 0 \\ 0 & Y \end{smallmatrix}])$. A filter is separable if its unit sample response satisfies $h[n_1X, n_2Y] = h_1[n_1X] \cdot h_2[n_2Y]$ for two 1D filters h_1 and h_2 . Separability can significantly reduce the implementation complexity of a filter, as we will see. The frequency response of a separable filter is given by

$$\begin{aligned} H(u, v) &= \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} h[n_1X, n_2Y] \exp(-j2\pi(un_1X + vn_2Y)) \\ &= \sum_{n_1=-\infty}^{\infty} h_1[n_1X] \exp(-j2\pi un_1X) \sum_{n_2=-\infty}^{\infty} h_2[n_2Y] \exp(-j2\pi vn_2Y) \\ &= H_1(u)H_2(v) \end{aligned}$$

which is also a separable function in the frequency domain.

Now consider the implementation of the 2D convolution for a separable FIR system. Assume that the region of support of h_1 is $[P_1X, Q_1X]$ and the region of support of h_2 is $[P_2Y, Q_2Y]$, so that the region of support of the separable filter is

$$\begin{aligned} \mathcal{B} &= [P_1X, Q_1X] \times [P_2Y, Q_2Y] \\ &= \{(n_1X, n_2Y) \mid P_1 \leq n_1 \leq Q_1, P_2 \leq n_2 \leq Q_2\}. \end{aligned}$$

The filter output, by direct implementation of the convolution, is

$$g[n_1X, n_2Y] = \sum_{k_1=P_1}^{Q_1} \sum_{k_2=P_2}^{Q_2} h[k_1X, k_2Y] f[(n_1 - k_1)X, (n_2 - k_2)Y].$$

This double sum contains $(Q_1 - P_1 + 1)(Q_2 - P_2 + 1)$ terms so that direct implementation requires $(Q_1 - P_1 + 1)(Q_2 - P_2 + 1)$ multiplications per output point and $(Q_1 - P_1 + 1)(Q_2 - P_2 + 1) - 1$ additions.

Now, using separability, we can express the filter output as

$$\begin{aligned} g[n_1X, n_2Y] &= \sum_{k_1=P_1}^{Q_1} \sum_{k_2=P_2}^{Q_2} h_1[k_1X] h_2[k_2Y] f[(n_1 - k_1)X, (n_2 - k_2)Y] \\ &= \sum_{k_1=P_1}^{Q_1} h_1[k_1X] \sum_{k_2=P_2}^{Q_2} h_2[k_2Y] f[(n_1 - k_1)X, (n_2 - k_2)Y]. \end{aligned}$$

Let us define

$$r[jX, n_2Y] = \sum_{k_2=P_2}^{Q_2} h_2[k_2Y]f[jX, (n_2 - k_2)Y].$$

This intermediate image r is obtained by filtering each *column* of the input image f with the 1D filter h_2 , and requires $Q_2 - P_2 + 1$ multiplications per output point. Then, the filter output is

$$g[n_1X, n_2Y] = \sum_{k_1=P_1}^{Q_1} h_1[k_1X]r[(n_1 - k_1)X, n_2Y]$$

which is obtained by filtering each *row* of the intermediate image r with the 1D filter h_1 , requiring $Q_1 - P_1 + 1$ multiplications per output point. Thus overall, separable filtering requires $(Q_1 - P_1 + 1) + (Q_2 - P_2 + 1)$ multiplications per output point. This can be significantly smaller than the $(Q_1 - P_1 + 1)(Q_2 - P_2 + 1)$ required for the non-separable implementation, especially as the filter size increases. For example, if $P_1 = P_2 = -5$ and $Q_1 = Q_2 = 5$, we have an 11×11 filter impulse response. The non-separable implementation requires 121 multiplications per output point, while the separable implementation requires 22 multiplications, a factor of 5.5 less.

Problems

1. Prove the properties of lattices given in Section 3.1.2.
2. Prove the Fourier transform properties shown in Table 3.1.
3. Prove that convolution on a lattice is commutative, $f * h = h * f$.
4. A linear shift-invariant filter defined on the hexagonal lattice

$$\Lambda = \text{LAT} \left(\begin{bmatrix} 2X & X \\ 0 & 1.5X \end{bmatrix} \right)$$

has unit-sample response given by

$$h[\mathbf{x}] = \begin{cases} \frac{1}{4} & \mathbf{x} = (0, 0) \\ \frac{1}{8} & \mathbf{x} = (X, 1.5X) \text{ or } (-X, -1.5X) \text{ or } (0, 3X) \text{ or } (0, -3X) \\ \frac{1}{16} & \mathbf{x} = (2X, 0) \text{ or } (-2X, 0) \text{ or } (X, -1.5X) \text{ or } (-X, 1.5X) \\ 0 & \text{otherwise} \end{cases}$$

Determine the frequency response $H(u, v)$ of this filter. Express it in real form. What is the DC gain of this filter?

5. For each of the following two-dimensional lattices Λ given by their sampling matrix, sketch the lattice to scale in the space domain, determine and sketch the reciprocal lattice and a Voronoi unit cell of the reciprocal lattice.

$$(a) V_\Lambda = \begin{bmatrix} 2X & 0 \\ 0 & 2X \end{bmatrix}$$

$$(b) V_\Lambda = \begin{bmatrix} 3X & X \\ 0 & X \end{bmatrix}$$

$$(c) V_\Lambda = \begin{bmatrix} X & X \\ X & -X \end{bmatrix}$$

6. A two-dimensional FIR filter defined on the rectangular lattice $\Lambda = \text{LAT}(\text{diag}(X, X))$ has unit-sample response shown in Fig. 3.12.

- (a) Compute the frequency response $H(u, v)$. Express it in real form.
 (b) What is the output $g[x, y]$ of this filter if the input is

$$f[x, y] = \delta[x - X, y + X] - \delta[x + X, y - X]?$$

Carefully sketch the output signal $g[x, y]$ to scale in the same manner as in Fig. 3.12.

7. Consider an ideal discrete-space circularly symmetric lowpass filter defined on the rectangular lattice with horizontal and vertical sample spacing X and

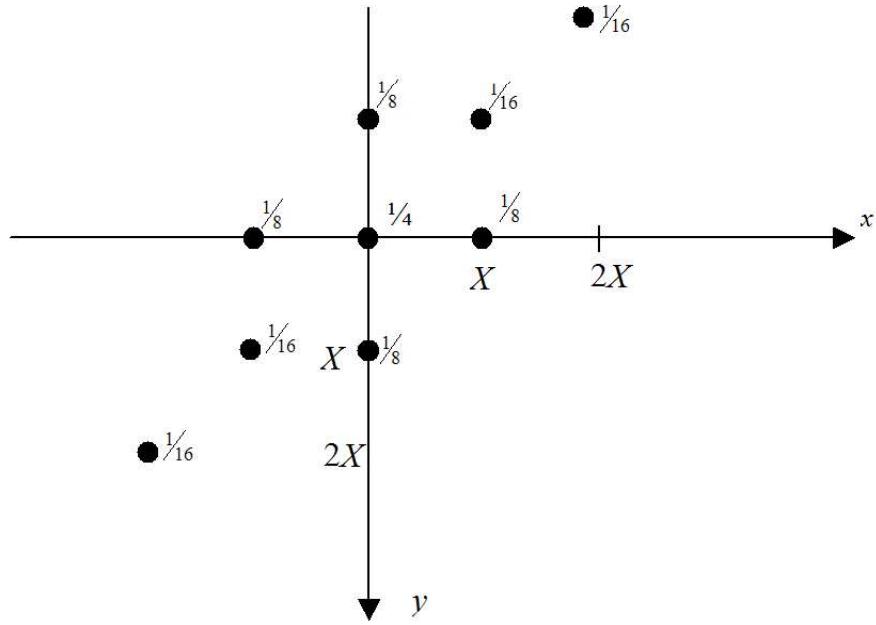


Figure 3.12: Unit-sample response $h[x, y]$. Non-zero values are shown; all others are zero.

Y . The passband is $C_W = \{(u, v) \mid u^2 + v^2 \leq W^2\}$ and the unit cell of the reciprocal lattice is $P^* = \{(u, v) \mid -1/2X \leq u < 1/2X, -1/2Y \leq v < 1/2Y\}$. Assume that $W < \min(1/2X, 1/2Y)$.

$$H(u, v) = \begin{cases} 1 & (u, v) \in C_W \\ 0 & (u, v) \in P^* \setminus C_W \end{cases}$$

where of course $H(u + k/X, v + l/Y) = H(u, v)$ for all integers k, l . Show that the unit sample response of this filter is given by

$$h[mX, nY] = \frac{WXY}{\sqrt{X^2m^2 + Y^2n^2}} J_1(2\pi W \sqrt{X^2m^2 + Y^2n^2})$$

where $J_1(s)$ is the Bessel function of the first kind and first order. You may use the following identities:

$$J_0(s) = \frac{1}{2\pi} \int_0^{2\pi} \exp [js \cos(\theta + \phi)] d\theta, \quad \text{for any } \phi$$

$$\int s J_0(s) ds = s J_1(s)$$

Simplify the expression in the case $X = Y$.

Chapter 4

Sampling and reconstruction of multidimensional signals

Physical real-world images are continuous in space and time, while we must use sampled signals for digital image processing. Thus we must be able to convert continuous-domain signals to discrete-domain signals, a process known as sampling. Likewise, discrete-domain signals may need to be converted to the continuous domain for viewing, a process known as reconstruction. This chapter presents the theory of ideal sampling and reconstruction, which is strongly based on Fourier analysis. Some of the issues involved in practical sampling and reconstruction are then discussed.

4.1 Ideal sampling and reconstruction

Let $f_c(\mathbf{x})$ be a continuous-domain multidimensional signal, which could be for example a still image $f_c(x, y)$ or time-varying video $f_c(x, y, t)$. We wish to convert f_c to a discrete-domain signal for digital processing and subsequent conversion back to continuous form. The ideal sampling operation can be simply expressed

$$f[\mathbf{x}] = f_c(\mathbf{x}), \quad \mathbf{x} \in \Lambda \tag{4.1}$$

where Λ is the sampling lattice. Assume that $f_c(\mathbf{x})$ has Fourier transform

$$F_c(\mathbf{u}) = \int_{\mathbb{R}^D} f_c(\mathbf{x}) \exp(-j2\pi\mathbf{u} \cdot \mathbf{x}) d\mathbf{x}, \quad \mathbf{u} \in \mathbb{R}^D \quad (4.2)$$

with corresponding inverse Fourier transform

$$f_c(\mathbf{x}) = \int_{\mathbb{R}^D} F_c(\mathbf{u}) \exp(j2\pi\mathbf{u} \cdot \mathbf{x}) d\mathbf{u}, \quad \mathbf{x} \in \mathbb{R}^D. \quad (4.3)$$

Let \mathcal{P}^* be a unit cell of the reciprocal lattice Λ^* . We know that \mathcal{P}^* tiles \mathbb{R}^D when centered on the points of Λ^* , so that we can write

$$f[\mathbf{x}] = \sum_{\mathbf{r} \in \Lambda^*} \int_{\mathcal{P}^*} F_c(\mathbf{u} + \mathbf{r}) \exp(j2\pi(\mathbf{u} + \mathbf{r}) \cdot \mathbf{x}) d\mathbf{u}, \quad \mathbf{x} \in \Lambda. \quad (4.4)$$

Since $\mathbf{r} \cdot \mathbf{x} \in \mathbb{Z}$ for all $\mathbf{r} \in \Lambda^*$, $\mathbf{x} \in \Lambda$, this becomes

$$\begin{aligned} f[\mathbf{x}] &= \sum_{\mathbf{r} \in \Lambda^*} \int_{\mathcal{P}^*} F_c(\mathbf{u} + \mathbf{r}) \exp(j2\pi\mathbf{u} \cdot \mathbf{x}) d\mathbf{u} \\ &= \int_{\mathcal{P}^*} \left(\sum_{\mathbf{r} \in \Lambda^*} F_c(\mathbf{u} + \mathbf{r}) \right) \exp(j2\pi\mathbf{u} \cdot \mathbf{x}) d\mathbf{u}, \quad \mathbf{x} \in \Lambda. \end{aligned} \quad (4.5)$$

We know that

$$f[\mathbf{x}] = d(\Lambda) \int_{\mathcal{P}^*} F(\mathbf{u}) \exp(j2\pi\mathbf{u} \cdot \mathbf{x}) d\mathbf{u}, \quad \mathbf{x} \in \Lambda \quad (4.6)$$

and that the Fourier transform is unique. It follows that

$$F(\mathbf{u}) = \frac{1}{d(\Lambda)} \sum_{\mathbf{r} \in \Lambda^*} F_c(\mathbf{u} + \mathbf{r}). \quad (4.7)$$

We assume in the above that the sums and integrals converge and that exchanging the order of the sum and integrals is permitted; these conditions hold in the cases of interest to us. A more rigorous but accessible discussion can be found in standard books on Fourier analysis such as [Kamm 00].

Fig. 4.1 illustrates the sampling on a hexagonal lattice of a 2D continuous-space signal, whose Fourier transform is limited to a circular region. If the signal $f_{c1}(\mathbf{x})$ whose Fourier transform $F_{c1}(\mathbf{u})$ is limited to the circular region in Fig. 4.1(a) is

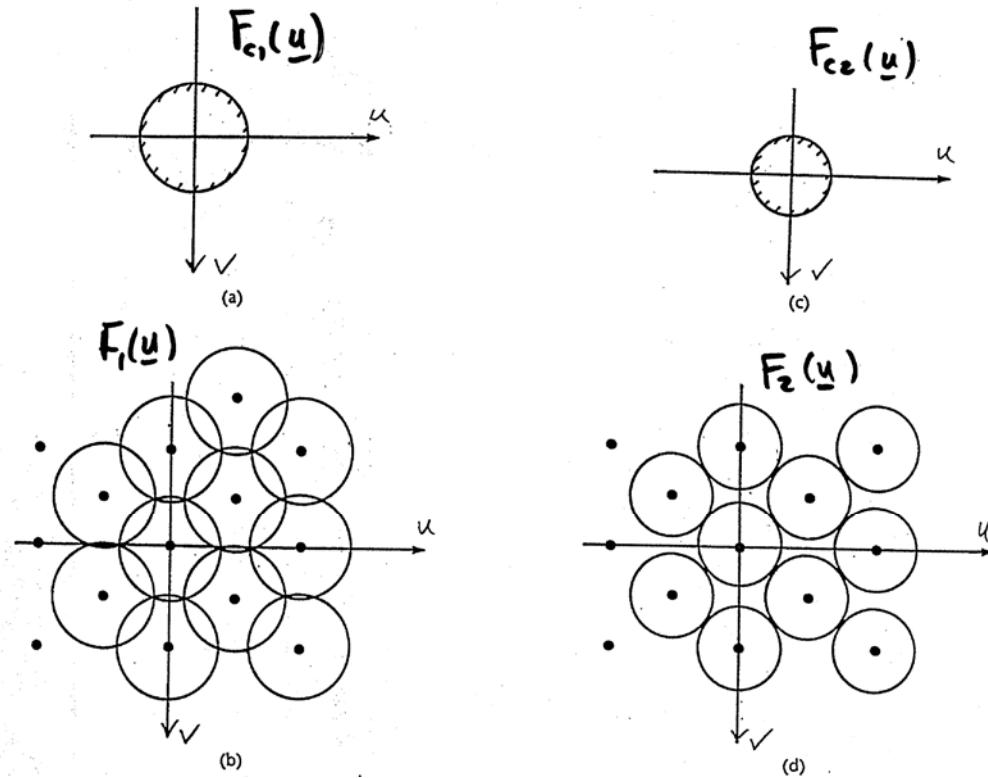


Figure 4.1: Sampling of a 2D continuous-space signal on a hexagonal lattice.

sampled on this lattice, the replicas on the points of the reciprocal lattice will overlap, as shown in Fig. 4.1(b). If the signal $f_{c2}(\mathbf{x})$ whose Fourier transform $F_{c2}(\mathbf{u})$ is limited to the smaller circular region in Fig. 4.1(c) is sampled on the same lattice, the replicas on the points of the reciprocal lattice will not overlap, as shown in Fig. 4.1(d).

It will be possible to reconstruct $f_c(\mathbf{x})$ exactly from the samples if the shifted copies $F_c(\mathbf{u} + \mathbf{r})$ in the sum do not overlap. Assume that $F_c(\mathbf{u}) = 0$ if $\mathbf{u} \notin \mathcal{V}$, where \mathcal{V} is some bounded set in the frequency domain. If the lattice Λ is chosen so that $\mathcal{V} \subset \mathcal{P}^*$ for *some* unit cell \mathcal{P}^* of Λ^* , then $F_c(\mathbf{u}) = 0$ for $\mathbf{u} \notin \mathcal{P}^*$. It follows that $F_c(\mathbf{u})$ and $F_c(\mathbf{u} + \mathbf{r})$ do not overlap for $\mathbf{r} \in \Lambda^* \setminus \mathbf{0}$. This implies that

$$F(\mathbf{u}) = \frac{1}{d(\Lambda)} F_c(\mathbf{u}) \quad \text{for } \mathbf{u} \in \mathcal{P}^*. \quad (4.8)$$

Thus

$$\begin{aligned}
f_c(\mathbf{x}) &= \int_{\mathbb{R}^D} F_c(\mathbf{u}) \exp(j2\pi\mathbf{u} \cdot \mathbf{x}) d\mathbf{u} \\
&= \int_{\mathcal{P}^*} F_c(\mathbf{u}) \exp(j2\pi\mathbf{u} \cdot \mathbf{x}) d\mathbf{u} \\
&= d(\Lambda) \int_{\mathcal{P}^*} F(\mathbf{u}) \exp(j2\pi\mathbf{u} \cdot \mathbf{x}) d\mathbf{u} \\
&= d(\Lambda) \int_{\mathcal{P}^*} \left(\sum_{\mathbf{s} \in \Lambda} f[\mathbf{s}] \exp(-j2\pi\mathbf{u} \cdot \mathbf{s}) \right) \exp(j2\pi\mathbf{u} \cdot \mathbf{x}) d\mathbf{u} \\
&= d(\Lambda) \sum_{\mathbf{s} \in \Lambda} f[\mathbf{s}] \int_{\mathcal{P}^*} \exp(j2\pi\mathbf{u} \cdot (\mathbf{x} - \mathbf{s})) d\mathbf{u} \\
&= \sum_{\mathbf{s} \in \Lambda} f[\mathbf{s}] t(\mathbf{x} - \mathbf{s}), \quad \mathbf{x} \in \mathbb{R}^D
\end{aligned} \tag{4.9}$$

where

$$t(\mathbf{x}) = d(\Lambda) \int_{\mathcal{P}^*} \exp(j2\pi\mathbf{u} \cdot \mathbf{x}) d\mathbf{u}. \tag{4.10}$$

We identify $t(\mathbf{x})$ as the unit-sample response of a hybrid discrete-input, continuous-output filter. It is an ideal lowpass filter that passes the band \mathcal{P}^* . This important result is the *multidimensional Nyquist sampling theorem*: a multidimensional continuous-domain signal $f_c(\mathbf{x})$ can be exactly reconstructed from its samples on a lattice Λ if the Fourier transform $F_c(\mathbf{u})$ is zero everywhere outside some unit cell of the reciprocal lattice Λ^* .

On the other hand, if $f_c(\mathbf{x})$ is not bandlimited to a unit cell of Λ^* , the replicas $F_c(\mathbf{u} + \mathbf{r})$ will overlap and perfect reconstruction of $f_c(\mathbf{x})$ cannot be achieved from its samples on Λ . High frequencies in the input are mapped into lower frequencies in the sampled signal, a phenomenon called *aliasing*.

4.2 Practical Sampling

The situation discussed in the previous section does not correspond to sampling and reconstruction in real applications for several reasons:

- We cannot measure the intensity of an image at a single point in space-time. We must collect light over some spatial neighborhood of the desired point, and for some period of time.
- We don't want to measure the intensity at a single point anyway since that would inevitably result in aliasing.
- We cannot reconstruct the continuous space-time image using an ideal low-pass filter with a physical system.

In practice, light is collected over some spatiotemporal neighborhood of the point to be sampled using a *sampling aperture*:

$$f[\mathbf{x}] = \int_{\mathbb{R}^D} f_c(\mathbf{x} + \mathbf{s}) a(\mathbf{x}, \mathbf{s}) d\mathbf{s}, \quad \mathbf{x} \in \Lambda. \quad (4.11)$$

If we assume that the aperture is space-time invariant, $a(\mathbf{x}, \mathbf{s}) = a(\mathbf{s})$, then

$$f[\mathbf{x}] = \int_{\mathbb{R}^D} f_c(\mathbf{x} + \mathbf{s}) a(\mathbf{s}) d\mathbf{s}, \quad \mathbf{x} \in \Lambda. \quad (4.12)$$

Let $h_a(\mathbf{s}) = a(-\mathbf{s})$, and let $\mathbf{t} = -\mathbf{s}$. Then

$$f[\mathbf{x}] = \int_{\mathbb{R}^D} h_a(\mathbf{t}) f_c(\mathbf{x} - \mathbf{t}) d\mathbf{t}, \quad \mathbf{x} \in \Lambda. \quad (4.13)$$

This can be modeled as a two-step process. First, $f_c(\mathbf{x})$ is filtered with a continuous space-time filter with impulse response $h_a(\mathbf{x})$ to give

$$f_p(\mathbf{x}) = \int_{\mathbb{R}^D} h_a(\mathbf{t}) f_c(\mathbf{x} - \mathbf{t}) d\mathbf{t}, \quad \mathbf{x} \in \mathbb{R}^D. \quad (4.14)$$

Then, $f_p(\mathbf{x})$ undergoes *ideal* sampling on the lattice Λ ,

$$f[\mathbf{x}] = f_p(\mathbf{x}), \quad \mathbf{x} \in \Lambda. \quad (4.15)$$

In this case, we have

$$\begin{aligned} F(\mathbf{u}) &= \frac{1}{d(\Lambda)} \sum_{\mathbf{r} \in \Lambda^*} F_p(\mathbf{u} + \mathbf{r}) \\ &= \frac{1}{d(\Lambda)} \sum_{\mathbf{r} \in \Lambda^*} H_a(\mathbf{u} + \mathbf{r}) F_c(\mathbf{u} + \mathbf{r}) \end{aligned} \quad (4.16)$$

The ideal role of $H_a(\mathbf{u})$ is to bandlimit $F_c(\mathbf{u})$ to \mathcal{P}^* , while maintaining the frequency content within this band as much as possible.

In typical image acquisition systems, the spatial aperture is either a rect function or a Gaussian function. Of course, neither of these is an ideal low-pass filter, so typically $f_p(\mathbf{x})$ will not be bandlimited to a unit cell of Λ^* and some aliasing will occur, while desired signal components within Λ^* will be attenuated. The temporal aperture is typically a rect function, corresponding to integration over some period T s. Consider for example the sampling of the green component of a color image with the Bayer array of Fig. 3.1. Assume that $X = 1/512$ ph. Then, the sampling aperture is

$$a(x, y) = \frac{1}{X^2} \operatorname{rect}(x/X, y/X) = h_a(x, y) \quad (4.17)$$

since the aperture is symmetric about the origin. From Example 2.3,

$$H_a(u, v) = \frac{\sin(\pi u X) \sin(\pi v X)}{\pi^2 X^2 u v}. \quad (4.18)$$

Recall that a sampling matrix for the green-channel sampling lattice is $\mathbf{V} = \begin{bmatrix} 2X & X \\ 0 & X \end{bmatrix}$, and a sampling matrix for the reciprocal lattice is

$$\mathbf{V}^{-T} = \begin{bmatrix} \frac{1}{2X} & 0 \\ -\frac{1}{2X} & \frac{1}{X} \end{bmatrix}.$$

Fig. 4.2 shows contours of $|H_a(u, v)|$ along with the reciprocal lattice Λ^* and a possible unit cell. It is clear that if the image on the sensor has a lot of detail or high-frequency components, the sampling aperture will let much of that through and aliasing will be present. Compare with an ideal low-pass filter whose passband is the unit cell.

The case of a video camera with a Gaussian spatial aperture and a rect temporal aperture was considered in Example 2.6.

4.3 Practical reconstruction

We have seen that ideal reconstruction is achieved with an ideal low-pass filter whose passband is the unit cell. Practical display devices do not have such a characteristic.

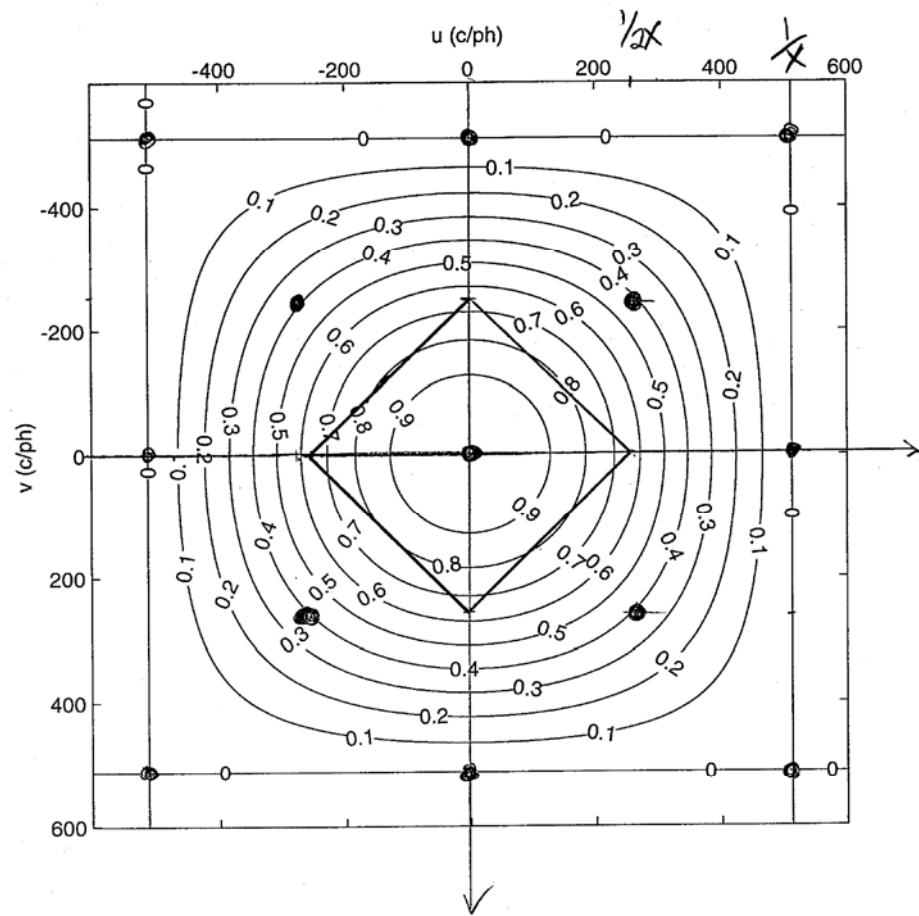


Figure 4.2: Frequency response of the rectangular aperture. Also shown is the reciprocal lattice and unit cell.

A suitable model for the reconstruction process is

$$f_r(\mathbf{x}) = \sum_{\mathbf{s} \in \Lambda} f[\mathbf{s}]d(\mathbf{x} - \mathbf{s}) \quad (4.19)$$

where $d(\mathbf{x})$ is called the display aperture. Thus the sample at position \mathbf{s} contributes the display aperture centered at position \mathbf{s} and weighted by the value $f[\mathbf{s}]$ to the reconstructed output. Combining practical reconstruction with practical sampling, we can relate the Fourier transform of the displayed image to that of the original image by

$$F_r(\mathbf{u}) = \frac{1}{d(\Lambda)} D(\mathbf{u}) \sum_{\mathbf{r} \in \Lambda^*} H_a(\mathbf{u} + \mathbf{r}) F_c(\mathbf{u} + \mathbf{r}), \quad (4.20)$$

where $D(\mathbf{u})$ is the continuous-domain Fourier transform of the display aperture.

In a cathode ray tube (CRT) display, we can model $d(\mathbf{x})$ by a Gaussian in the spatial domain and an exponential in the temporal domain. It is clear that this is not an ideal low-pass filter, and so a significant portion of the spectral replicas remain in the reconstructed signal. However, if the image is then viewed by a human observer, further filtering takes place in the human visual system.

Problems

1. A two-dimensional continuous-domain signal $f_c(x, y)$ has Fourier transform

$$F_c(u, v) = \begin{cases} ce^{-\alpha(|u|+|v|)} & u^2 + v^2 < W^2 \\ 0 & u^2 + v^2 \geq W^2 \end{cases}$$

for some real number W . The signal is sampled on a hexagonal lattice Λ with sampling matrix

$$\mathbf{V} = \begin{bmatrix} X & X/2 \\ 0 & \sqrt{3}X/2 \end{bmatrix}$$

to give the sampled signal $f[x, y]$, $(x, y) \in \Lambda$, with Fourier transform $F(u, v)$.

- What is the expression for $F(u, v)$ in terms of $F_c(u, v)$?

- (b) Find the largest possible value of X such that there is no aliasing? Sketch the region of support of the Fourier transform of the sampled signal in this case (including all replicas), and also indicate a unit cell of the reciprocal lattice Λ^* .
2. The face-centered cubic lattice is the most efficient lattice for the packing of spheres in three dimensions. A sampling matrix for this lattice is given by

$$\mathbf{V} = K \begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

where K is some real constant. Suppose that a bandlimited three-dimensional signal $f(\mathbf{x})$ satisfying $F(\mathbf{u}) = 0$ for $|\mathbf{u}| > W$ is sampled on a lattice whose reciprocal lattice is face-centered cubic. Find the least dense lattice such that there is no aliasing. Compare the resulting sampling density with the best orthogonal sampling for which there is no aliasing.

3. Fig. 4.3 illustrates the sensor in a hypothetical digital camera using a sensor element which is hexagonal in shape. There are $M = 740$ sensor elements in each horizontal row and there are $N = 480$ rows of sensor elements, for a total of 480×740 sensor elements. The centers of the sensor elements lie on a hexagonal lattice Λ , and each sensor element is a unit cell of this lattice. The output of each sensor element is the integral of light irradiance over the sensor element for some arbitrary exposure time, and it is associated with the lattice point at the center of the sensor element. Assume that the picture width is MX and the picture height (ph) is NY . We use the picture height as the unit of length. The sensor element is a regular hexagon with $Y = \sqrt{3}X/2$. (Note that Fig. 4.3 is just a sketch and is not drawn to scale.)

- (a) Give a sampling matrix for the lattice shown in Fig. 4.3 in units of ph.
 (b) What is the area of a sensor element, with correct units? What is the sampling density, with correct units?

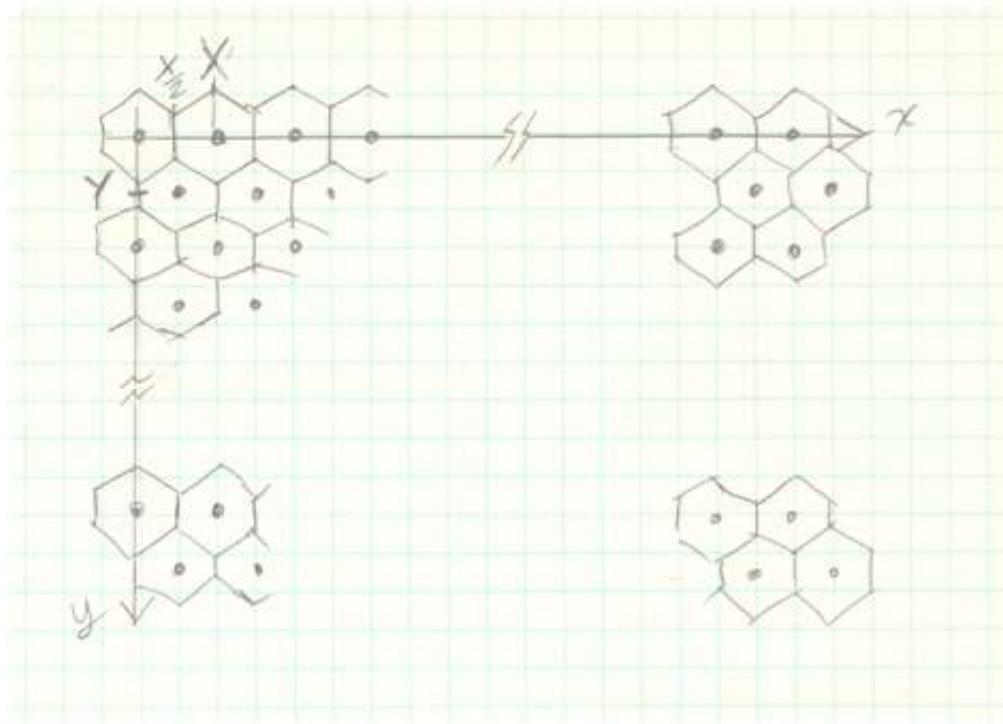


Figure 4.3: An image sensor with hexagonal sensor elements.

- (c) What is the aspect ratio of the sensor? Is it approximately $4/3$ or approximately $16/9$?
- (d) The sampling process carried out by this sensor can be modeled by a linear shift-invariant (LSI) continuous-space filter followed by ideal sampling on Λ . Give an expression for the impulse response $h_a(x, y)$ of this LSI filter with the correct gain. Assume that if the image irradiance is a constant value over a sensor element (in arbitrary normalized units), the sampled value is that same value, i.e., the DC gain of $h_a(x, y)$ is 1.0.
- (e) Give an expression for the frequency response $H_a(u, v)$ corresponding to the camera aperture impulse response $h_a(x, y)$.
- (f) Assume that the continuous-space input light irradiance $f_c(x, y)$ has a Fourier transform $F_c(u, v)$. Give an expression for the Fourier transform

of the sampled image $f[x, y], (x, y) \in \Lambda$ in terms of in terms of $F_c(u, v)$ and $H_a(u, v)$; you should explicitly evaluate the reciprocal lattice Λ^* .

Chapter 5

Analysis and design of multidimensional FIR filters

This chapter presents and analyzes a number of examples of multidimensional FIR filters that are often used, along with some issues of design. We first start with a few *ad hoc* designs.

5.1 Moving Average Filters

The moving average filter simply takes the average of the pixel values over the region \mathcal{B} centered on the output point \mathbf{x} ,

$$g[\mathbf{x}] = \frac{1}{|\mathcal{B}|} \sum_{\mathbf{s} \in \mathcal{B}} f[\mathbf{x} - \mathbf{s}], \quad (5.1)$$

where $|\mathcal{B}|$ denotes the number of elements in the set \mathcal{B} . This corresponds to an LSI filter with unit sample response

$$h_{MA}[\mathbf{x}] = \begin{cases} \frac{1}{|\mathcal{B}|} & \text{if } \mathbf{x} \in \mathcal{B}, \\ 0 & \text{otherwise.} \end{cases} \quad (5.2)$$

Its frequency response is

$$H_{MA}(\mathbf{u}) = \frac{1}{|\mathcal{B}|} \sum_{\mathbf{x} \in \mathcal{B}} \exp(-j2\pi \mathbf{u} \cdot \mathbf{x}). \quad (5.3)$$

It is clear that $H_{MA}(0) = 1$.

Consider the most common usage of this filter: Λ is a rectangular 2D lattice LAT ($\begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix}$) and \mathcal{B} is a square region centered at the origin,

$$\mathcal{B} = \{(n_1X, n_2Y) \mid -L \leq n_1 \leq L, -L \leq n_2 \leq L\}, \quad (5.4)$$

with $|\mathcal{B}| = (2L + 1)^2$. In this case we can write

$$g[x, y] = \frac{1}{(2L + 1)^2} \sum_{n_1=-L}^L \sum_{n_2=-L}^L f[x - n_1X, y - n_2Y]. \quad (5.5)$$

This filter is a crude low-pass filter. Its frequency response is

$$\begin{aligned} H(u, v) &= \frac{1}{(2L + 1)^2} \sum_{n_1=-L}^L \sum_{n_2=-L}^L \exp(-j2\pi(un_1X + vn_2Y)) \\ &= \frac{1}{2L + 1} \sum_{n_1=-L}^L \exp(-j2\pi un_1X) \frac{1}{2L + 1} \sum_{n_2=-L}^L \exp(-j2\pi vn_2Y). \end{aligned} \quad (5.6)$$

Each component sum has the same form; it is a geometric series familiar from 1D signals and systems. Evaluating gives

$$H(u, v) = \frac{1}{(2L + 1)^2} \frac{\sin(\pi u(2L + 1)X) \sin(\pi v(2L + 1)Y)}{\sin(\pi uX) \sin(\pi vX)}. \quad (5.7)$$

Fig. 5.1 shows the profile $|H(u, 0)|$ of this response for several values of L .

5.2 Gaussian Filters

Gaussian filters are very popular in image processing. In continuous space, the frequency response of a filter with a Gaussian impulse response is also Gaussian (Table 2.2). The Gaussian has the property of being simultaneously concentrated in space and frequency and having no oscillations. In discrete space, the unit sample response of a Gaussian filter has the form

$$h_G[\mathbf{x}] = c \exp(-\|\mathbf{x}\|^2/2r^2), \quad \mathbf{x} \in \mathcal{B}, \quad (5.8)$$

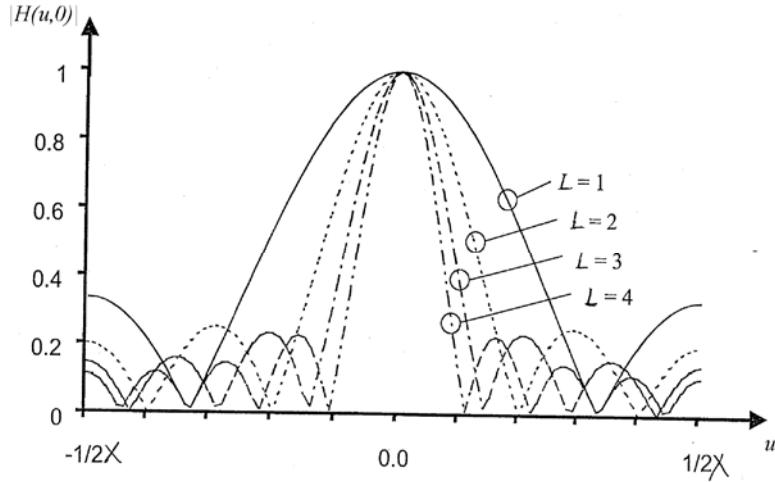


Figure 5.1: Profile $|H(u, 0)|$ of frequency response of moving average filter.

where r is a parameter that determines the spread of the unit sample response. The region \mathcal{B} is normally chosen large enough that samples outside it are negligibly small. In the common case of a square lattice

$$h_G[n_1 X, n_2 X] = c \exp(-(n_1^2 + n_2^2)X^2/2r^2), \quad -L \leq n_1, n_2 \leq L. \quad (5.9)$$

A Gaussian filter is a type of low-pass filter, and since we want a DC gain of 1.0, we choose the constant c so that $H(\mathbf{0}) = \sum_{\mathbf{x} \in \mathcal{B}} h_G[\mathbf{x}] = 1$, i.e.,

$$c = \frac{1}{\sum_{\mathbf{x} \in \mathcal{B}} \exp(-\|\mathbf{x}\|^2/2r^2)}. \quad (5.10)$$

Since the frequency response is periodic in the frequency domain, it cannot be a Gaussian as in the continuous case. However, if r is such that the Fourier transform of the continuous Gaussian is *mostly* confined to a unit cell of Λ^* , then using Eq. 4.8, we can conclude that the frequency response of $h_G[\mathbf{x}]$ within a unit cell will be approximately Gaussian:

$$H(u, v) \approx \frac{c}{d(\Lambda)} 2\pi r^2 \exp(-2\pi^2(u^2 + v^2)r^2), \quad (u, v) \in \mathcal{P}^*. \quad (5.11)$$

Note that a DC gain of 1 implies that $c \approx d(\Lambda)/2\pi r^2$. Let u_c be the (radial) frequency at which the frequency response has dropped to $1/\sqrt{2}$ of its DC value

(i.e., -3 dB). Then $\exp(-2\pi^2 u_c^2 r^2) = 2^{-1/2}$, or $2\pi^2 u_c^2 r^2 = 0.5 \ln 2$ and

$$u_c = \frac{\sqrt{\ln 2}}{2\pi r} \approx \frac{.1325}{r}. \quad (5.12)$$

If we are given u_c , then we should choose $r = .1325/u_c$.

Example 5.1. Design a low-pass filter on a square lattice with a 3 dB bandwidth of $1/8X$ c/ph in both the horizontal and the vertical directions (3 dB bandwidth corresponds to an attenuation of 0.707).

Solution. For a moving average filter, referring to Fig. 5.1, to meet the specification we need $L = 2$. For a Gaussian filter, according to (5.12), we should choose $r = 0.1325/(1/8X) = 1.06X$, and using (5.10) to have a DC gain of 1.0, we find $c = 0.1416$. Fig. 5.2 shows the contour and perspective plots of the resulting moving average filter, and Fig. 5.3 shows the response of the Gaussian filter. The desired 3dB passband is shown on both figures. Fig. 5.4 shows the result of applying each of these two filters to the Barbara image of Fig. 3.11(a). Note that although the filters have the same target 3dB bandwidth, the moving average filter has too much attenuation at the desired cutoff frequency and has introduced greater blurring into the image than the Gaussian filter. At the same time, the moving average filter has passed significantly more high frequency energy which could potentially cause problems. \square

5.3 Bandpass and Bandstop Filters

The moving-average and Gaussian filters we have just seen are lowpass filters. It is possible to generate bandpass and bandstop filters from these using the modulation property of the Fourier transform. Assume that $h_{LP}[\mathbf{x}]$ is the unit-sample response of a lowpass filter, which has a DC gain of about 1.0. Then

$$h_{BP}[\mathbf{x}] = h_{LP}[\mathbf{x}] \exp(j2\pi\mathbf{u}_0 \cdot \mathbf{x}) \quad (5.13)$$

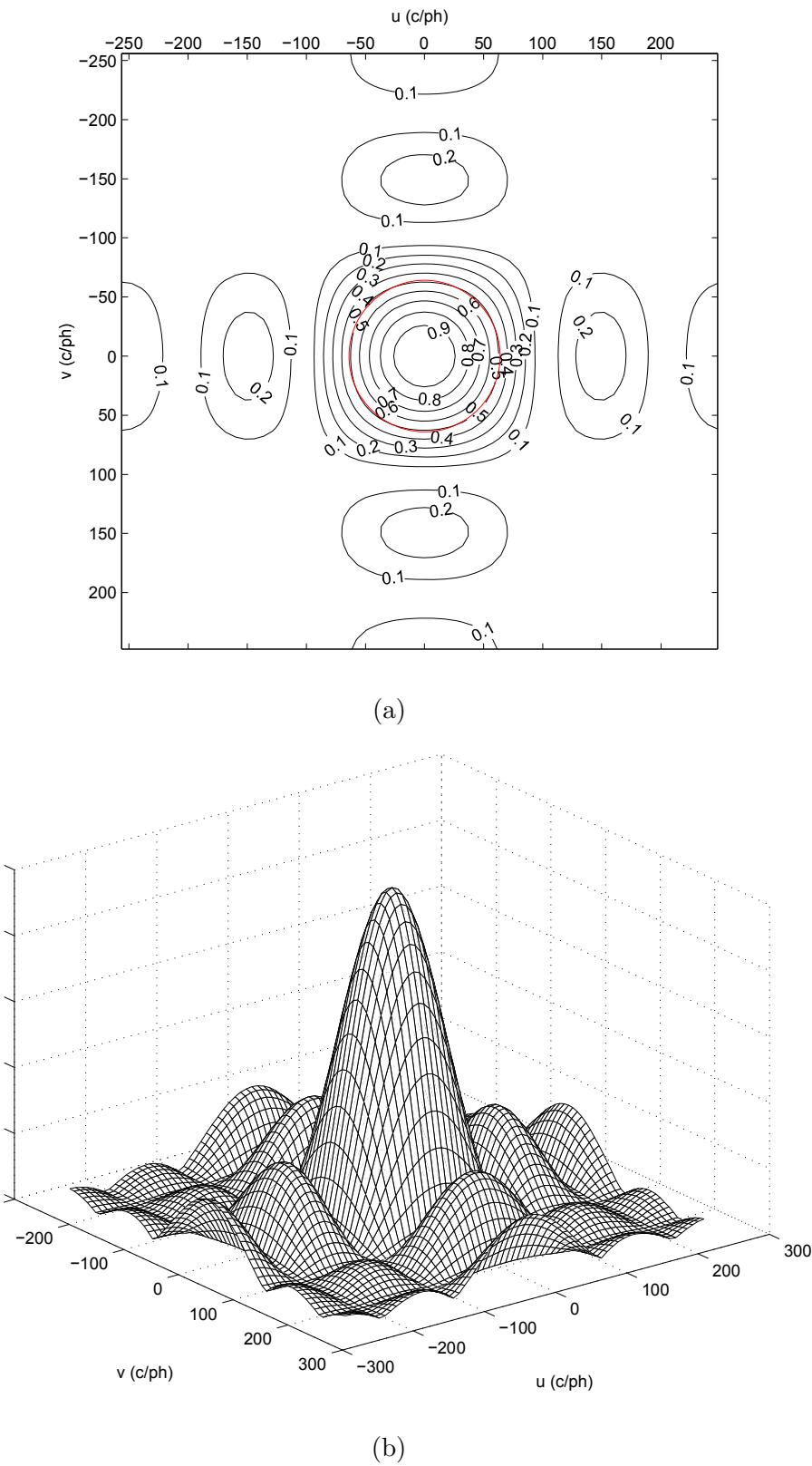
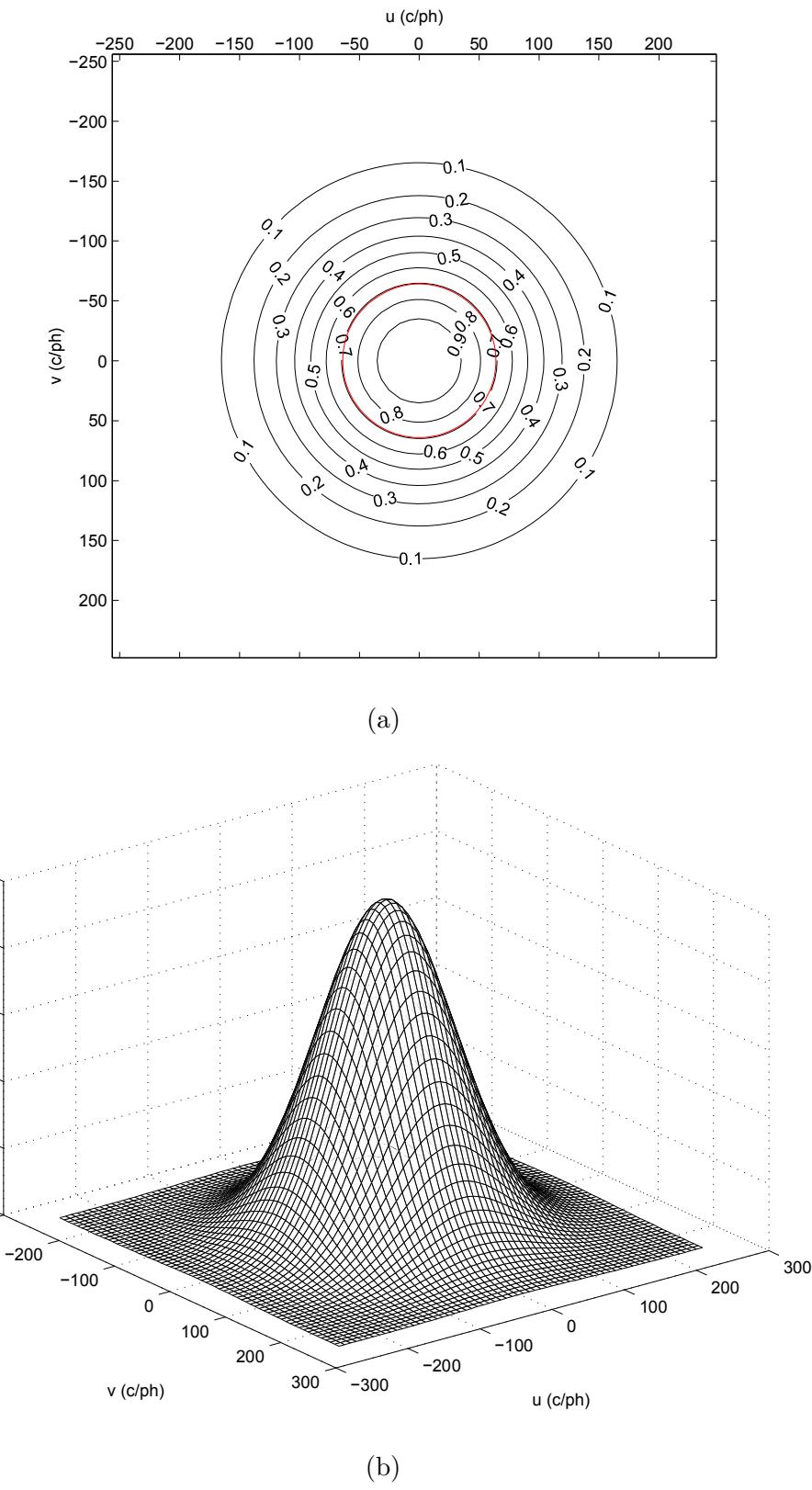


Figure 5.2: Frequency response of a moving average filter with $L = 2$. (a) Contour plot. (b) Perspective plot.

Figure 5.3: Frequency response of a Gaussian filter with $r = 1.06X$ and $c = 0.1416$.

(a) Contour plot. (b) Perspective plot.



(a)



(b)

Figure 5.4: Result of filtering Barbara image (a) moving average filter (b) Gaussian filter.

is the unit-sample response of a bandpass filter with center frequency \mathbf{u}_0 . From property (iii) in Table 3.1,

$$H_{BP}(\mathbf{u}) = H_{LP}(\mathbf{u} - \mathbf{u}_0). \quad (5.14)$$

Note that in general, $h_{BP}[\mathbf{x}]$ as defined above will be complex, except for special frequencies on the boundary of a unit cell of Λ^* , where $2\mathbf{u}_0 \in \Lambda^*$ (and so, $-\mathbf{u}_0$ and \mathbf{u}_0 represent the same discrete-domain frequency on the lattice Λ).

A real bandpass filter will have a frequency response that is symmetric about zero, with center frequencies $\pm\mathbf{u}_0$. Such a filter can be obtained by

$$\begin{aligned} h_{BP}[\mathbf{x}] &= 2h_{LP}[\mathbf{x}] \cos(2\pi\mathbf{u}_0 \cdot \mathbf{x}) \\ &= h_{LP}[\mathbf{x}] (\exp(j2\pi\mathbf{u}_0 \cdot \mathbf{x}) + \exp(-j2\pi\mathbf{u}_0 \cdot \mathbf{x})) \end{aligned} \quad (5.15)$$

$$H_{BP}(\mathbf{u}) = H_{LP}(\mathbf{u} - \mathbf{u}_0) + H_{LP}(\mathbf{u} + \mathbf{u}_0). \quad (5.16)$$

Note that in both cases discussed above, the frequency response retains its periodicity with respect to the reciprocal lattice.

A bandstop filter can be obtained from a bandpass filter by simply subtracting the unit-sample response of the bandpass filter from that of an all-pass filter $h_{AP}[\mathbf{x}] = \delta[\mathbf{x}]$. Thus

$$h_{BS}[\mathbf{x}] = \delta[\mathbf{x}] - h_{BP}[\mathbf{x}] \quad (5.17)$$

$$H_{BS}(\mathbf{u}) = 1 - H_{BP}(\mathbf{u}). \quad (5.18)$$

Example 5.2. Design a real Gaussian bandstop filter with center frequencies $\pm\mathbf{u}_0$ and 3 dB bandwidth $u_c = 15$ c/ph for a 512×512 image on a square sampling lattice. Use it to remove the specific frequency $\mathbf{u}_0 = (150, -86)$ c/ph from the image ‘barbara’.

Solution. From the specification, $r = 0.1325/u_c = 0.00883$ ph. Using the approximation $c = d(\lambda)/2\pi r^2 = 0.0078$, the desired unit-sample response is

$$\begin{aligned} h_{BS}[\mathbf{x}] &= \delta[\mathbf{x}] - 2c \cos(2\pi(u_0x + v_0y)) \exp\left(-\frac{(x^2 + y^2)}{2r^2}\right) \\ &= \delta[\mathbf{x}] - .0156 \cos(2\pi(150x - 86y)) \exp\left(\frac{-(x^2 + y^2)}{0.0001378}\right) \end{aligned}$$

A perspective plot of the frequency response is shown in Fig. 5.5. The filtered ‘barbara’ image is shown if Fig. 5.6. A close up of the original and filtered image are shown in Fig. 5.7. \square

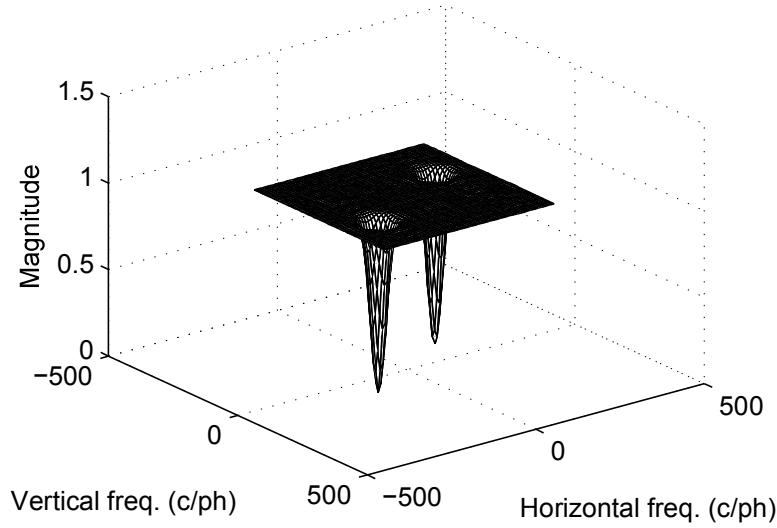


Figure 5.5: Frequency response of Gaussian notch filter.

5.4 Frequency-Domain Design of Multidimensional FIR Filters

FIR filter design involves the determination of the unit sample response of a filter such that its frequency response approximates a desired response as closely as possible. Many methods have been developed over the years. We present two of these methods here, namely the *window method* and the *least p^{th} optimization method*.

5.4.1 FIR filter design using windows

Let $H_I(\mathbf{u})$ be an ideal frequency response that will be approximated by the FIR filter with unit sample response $h[\mathbf{x}]$ defined on lattice Λ , confined to the finite set



Figure 5.6: Result of filtering Barbara image with Gaussian notch filter. Observe the area on the right knee.

\mathcal{B} . The ideal response is specified on a unit cell \mathcal{P}^* of Λ^* . We can determine the unit sample response corresponding to the ideal frequency response by taking the inverse Fourier transform to obtain

$$h_I[\mathbf{x}] = d(\Lambda) \int_{\mathcal{P}^*} H_I(\mathbf{u}) \exp(j2\pi\mathbf{u} \cdot \mathbf{x}) d\mathbf{u}. \quad (5.19)$$

However, this will in general result in a unit sample response that is not confined to \mathcal{B} , and is most likely of infinite extent. It can simply be truncated to \mathcal{B} by multiplying by the zero-one function $p_{\mathcal{B}}[\mathbf{x}]$:

$$h[\mathbf{x}] = h_I[\mathbf{x}]p_{\mathcal{B}}[\mathbf{x}]. \quad (5.20)$$

By property (v) of the Fourier transform

$$H(\mathbf{u}) = d(\Lambda) \int_{\mathcal{P}^*} H_I(\mathbf{r}) P_{\mathcal{B}}(\mathbf{u} - \mathbf{r}) d\mathbf{r} \quad (5.21)$$

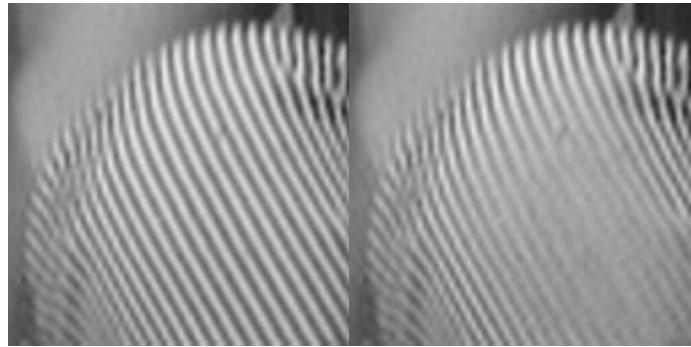


Figure 5.7: Result of filtering Barbara image with Gaussian notch filter: closeup of right knee. Left is original, right is filtered.

so that $H(\mathbf{u})$ may end up being quite different than $H_I(\mathbf{u})$, after convolving with $P_{\mathcal{B}}(\mathbf{u})$ in the frequency domain. Typically \mathcal{B} is a rectangular region, so that $p_{\mathcal{B}}$ is a discrete rect function whose Fourier transform can be computed in a straightforward fashion. In fact, we have already done it in computing the frequency response of the moving average filter; the only difference is that we don't have the $1/(2L + 1)^2$ term. Thus for a $(2L + 1) \times (2L + 1)$ filter,

$$P_{\mathcal{B}}(\mathbf{u}) = \frac{\sin(\pi u(2L + 1)X) \sin(\pi v(2L + 1)Y)}{\sin(\pi uX) \sin(\pi vX)}. \quad (5.22)$$

This function $p_{\mathcal{B}}$ is sometimes referred to as a “boxcar window.” Convolving the ideal response with the Fourier transform of the window will smear out the response and introduce ripples due to the sidelobes of $P_{\mathcal{B}}(\mathbf{u})$ (see Fig. 5.1).

Other windows $w[\mathbf{x}]$ have been developed that reduce the level of the sidelobes without widening the main lobe by much [Harr 78]. These windows taper off to a value of zero outside \mathcal{B} rather than simply truncating h_I . Windows have been studied extensively in one-dimensional signal processing. Many of these windows have a continuous analytical specification $w(t)$ where $w(t) = 0$ for $|t| > 0.5$. They can be extended to two-dimensional windows either separably or by forming a circularly symmetric window [Huan 72]. Examples of 1D windows are Hamming, Hanning,

Kaiser, Blackman and others. For example, the Hamming window is defined as

$$w_H(t) = \begin{cases} 0.54 + 0.46 \cos(2\pi t), & \text{if } |t| \leq 0.5; \\ 0, & \text{otherwise,} \end{cases} \quad (5.23)$$

and the boxcar window is given by $w_B(t) = \text{rect}(t)$.

A two-dimensional window with an approximately rectangular region of support

$$\mathcal{B} = \{(x, y) \in \Lambda \mid |x| \leq Q_1, |y| \leq Q_2\} \quad (5.24)$$

can be defined from a one-dimensional window $w_1(t)$ using a separable construction

$$w_R[x, y] = w_1\left(\frac{x}{2Q_1}\right) w_1\left(\frac{y}{2Q_2}\right), \quad (x, y) \in \Lambda. \quad (5.25)$$

Similarly, a two-dimensional window with an approximately circular region of support

$$\mathcal{B} = \{(x, y) \in \Lambda \mid \sqrt{x^2 + y^2} \leq Q\}$$

can be obtained using

$$w_C[x, y] = w_1\left(\frac{\sqrt{x^2 + y^2}}{2Q}\right), \quad (x, y) \in \Lambda.. \quad (5.26)$$

This construction could also be used with a rectangular region of support and applied to non-rectangular sampling structures. If w_1 is a good one-dimensional window, in general w_R and w_C will be good two-dimensional windows [Huan 72].

MATLAB provides functions to design 2D FIR filters on rectangular lattices using the window method. The following code designs a low-pass filter with a circular passband of radius $\frac{1}{8X}$ c/ph using a 2D window derived from the Hamming window using (5.26). The unit sample response is of size 51×51 .

```
%Window design
ph=512;
[f1,f2] = freqspace(51,'meshgrid');
Hd = ones(51);
```

```

rad = sqrt(f1.^2 + f2.^2);
Hd((rad>0.25)) = 0;
hw = fwind1(Hd,hamming(51));
[HW,fx,fy] = freqz2(hw,64,64,[1/ph 1/ph]);

```

The resulting filter response is shown in Fig. 5.8. We see that the filter has an approximately circular passband and a sharp transition band. The filtered Barbara image is shown in Fig. 5.9. It is clear that the sharp transition band in the filter's frequency response has introduced objectionable ringing in the filtered image.

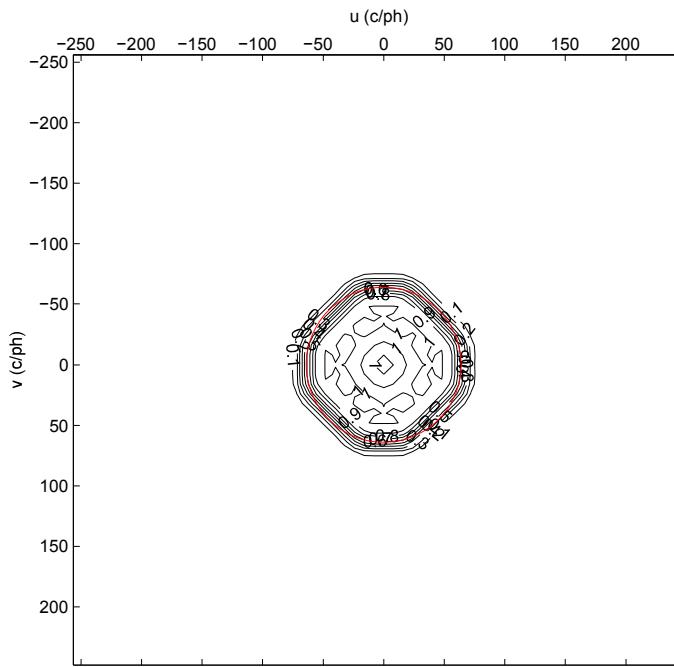
5.4.2 FIR filter design using least p^{th} optimization

Filter design using optimization is based on a numerical measure of how closely the filter response approximates the desired response $H_I(\mathbf{u})$. Such a measure is the p -error criterion

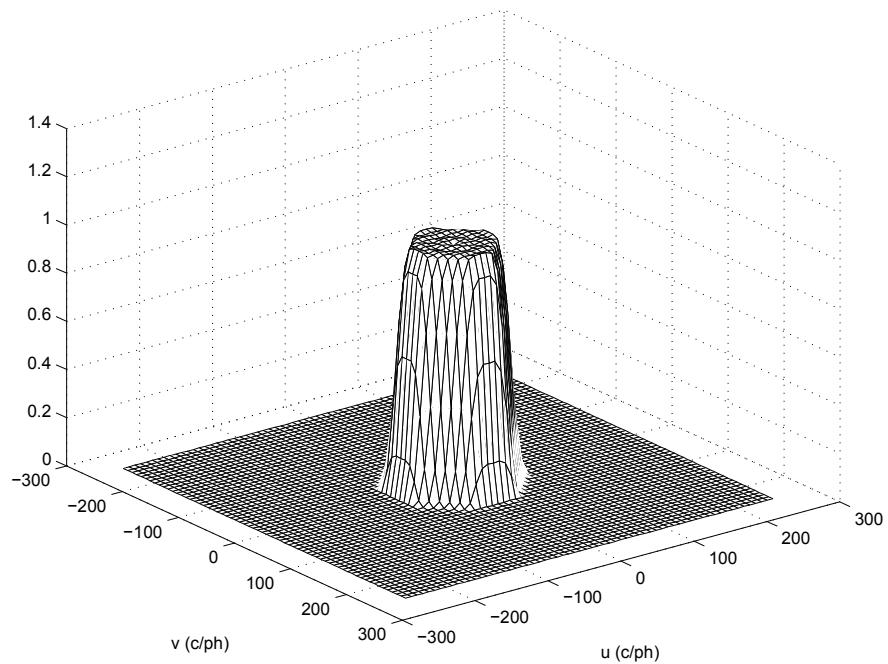
$$\mathcal{E} = \int_{\mathcal{P}^*} |H(\mathbf{u}) - H_I(\mathbf{u})|^p d\mathbf{u}. \quad (5.27)$$

Of course $\mathcal{E} = 0$ if $H(\mathbf{u}) = H_I(\mathbf{u})$ for all \mathbf{u} , and it increases as they become more different. An optimization method tries to minimize \mathcal{E} by varying the nonzero filter coefficients in the region of support \mathcal{B} . Typically, such optimization methods make use of the derivative of \mathcal{E} with respect to the filter coefficients.

An example of the application of this method to a practical filter design problem is taken from US patent 4,829,367, "Apparatus and method for encoding and decoding a NTSC color video signal" by E. Dubois and P. Faubert. Fig. 5.10(a)-(d) shows the ideal response of the four filters used in this system. The ideal response is 1 in the region marked P, 0 in the region marked S, and varies smoothly from 1 to 0 in the region marked T, for example according to a raised-cosine transition. The filters obtained with the least p^{th} method are shown in Fig. 5.11 for the first three of these cases.



(a)



(b)

Figure 5.8: Frequency response of a filter designed in MATLAB using a 2D separable 51×51 Hamming window. (a) Contour plot. (b) Perspective plot.

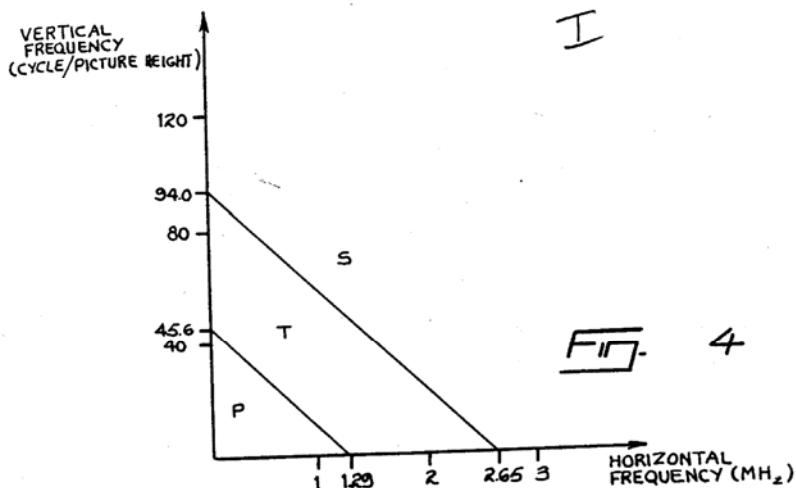


Figure 5.9: Result of filtering Barbara image using the Hamming window designed filter of Fig. 5.8.

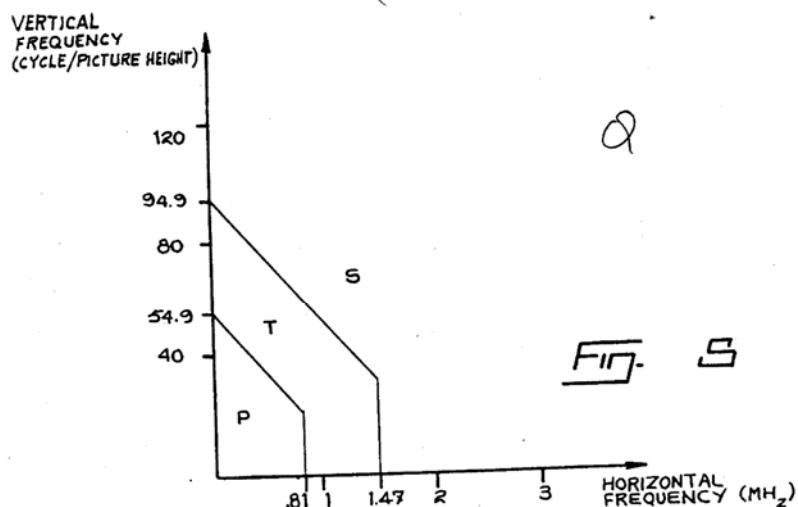
U.S. Patent May 9, 1989

Sheet 4 of 7

4,829,367



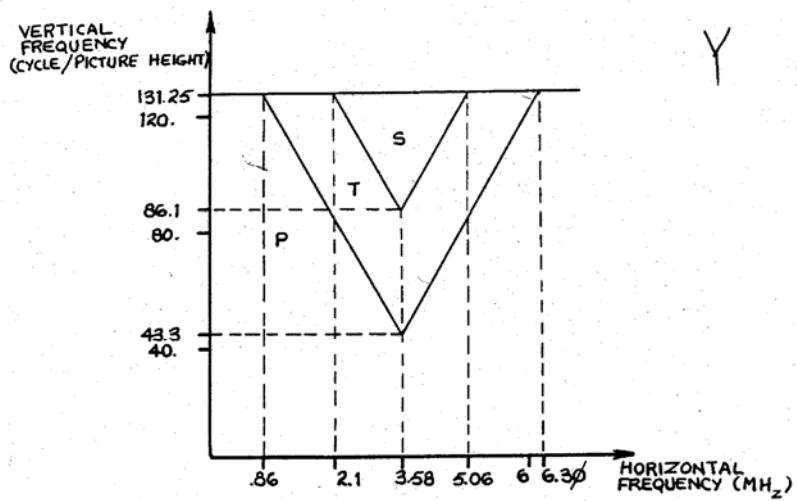
(a)



(b)

Figure 5.10: Specifications for (a) I filter and (b) Q filter (c) Y filter and (d) C filter in NTSC encoding system.

U.S. Patent May 9, 1989 Sheet 5 of 7 4,829,367



(c)

Fig.

6

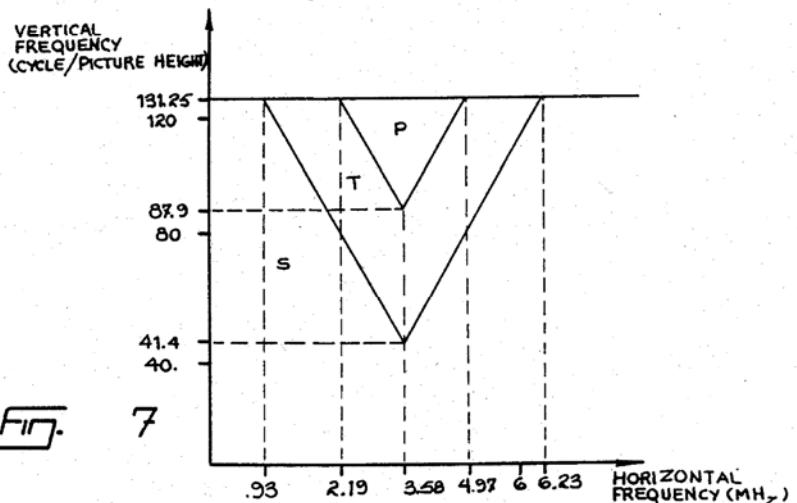


Fig.

7

(d)

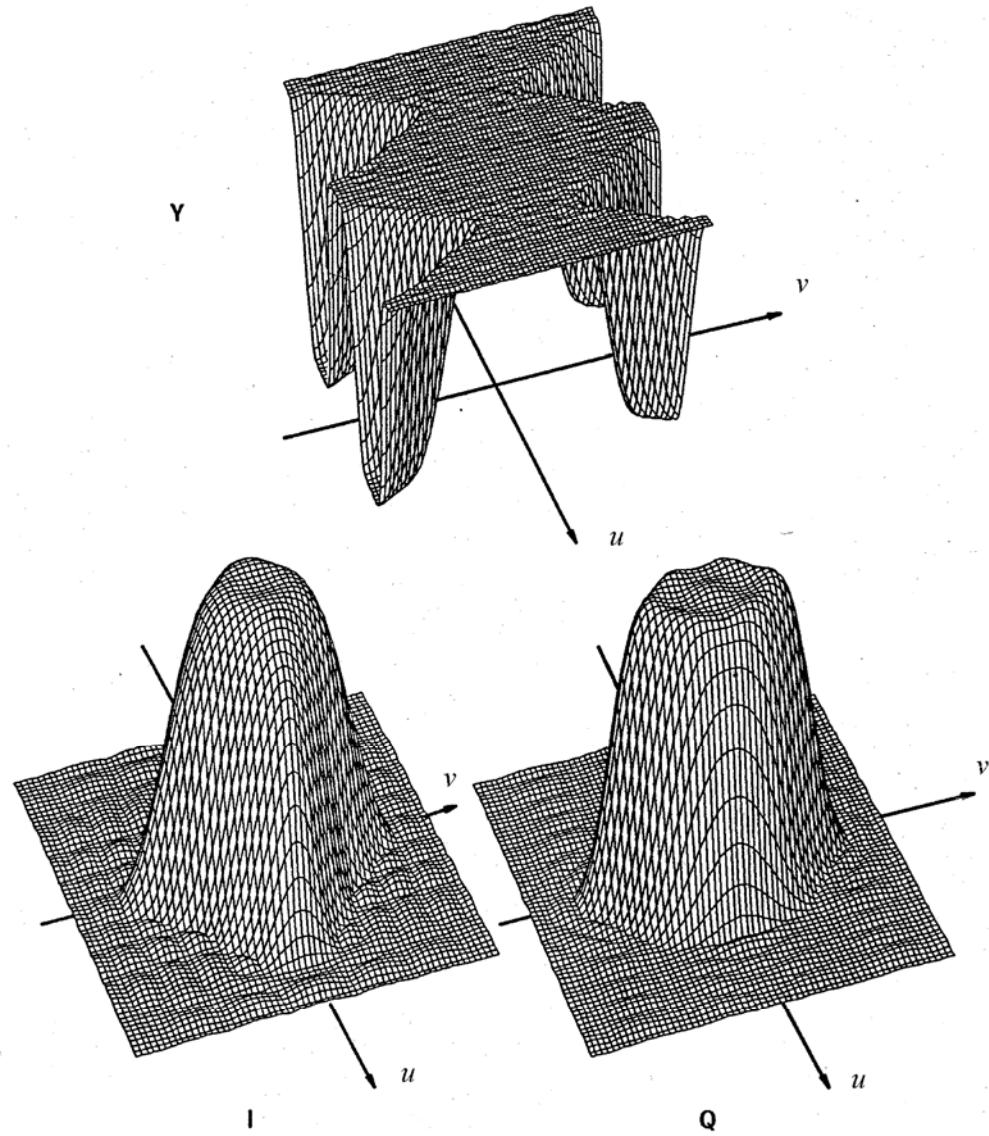


Figure 5.11: Response of the Y , I and Q filters obtained by least- p^{th} design from the specifications of Fig. 5.10

Problems

1. Derive in detail the expression for the frequency response of a moving average filter on a rectangular lattice, with a rectangular region of support, as given in Eq. 5.7.
2. Determine the frequency response of a moving average filter for a hexagonal lattice with a diamond-shaped region of support.
3. A square image ($pw = ph$) is sampled on the hexagonal lattice Λ generated by the sampling matrix

$$\mathbf{V} = \begin{bmatrix} X & X/2 \\ 0 & \sqrt{3}X/2 \end{bmatrix}$$

where $X = 1/512$ ph. Design a Gaussian FIR filter with unit sample response $h[\mathbf{x}] = c \exp(-\|\mathbf{x}\|^2/2r^2)$ for $\mathbf{x} \in \mathcal{B}$ having a 3dB bandwidth of $0.2/X$ c/ph. The region of support of the FIR filter is $\mathcal{B} = \{\mathbf{x} \in \Lambda \mid \|\mathbf{x}\| \leq 3X\}$. Having determined the correct values of c and r , give the coefficients of the filter. Give an analytical approximation for the frequency response of the filter. Make a contour plot and a perspective plot of the frequency response of the filter over the frequency range $-2/X \leq u \leq 2/X$, $-4/(\sqrt{3}X) \leq v \leq 4/(\sqrt{3}X)$. Sketch (by hand if you wish) on the contour plot the points of the reciprocal lattice Λ^* and a Voronoi unit cell of Λ^* , and comment on the periodicity of the frequency response. Recall that the Voronoi unit cell consists of all points in Λ^* closer to the origin than to any other point of Λ^* .

Chapter 6

Changing the Sampling Structure of an Image

There are two main applications for changing the sampling structure of an image: resizing and format conversion (or standards conversion). Resizing generally involves displaying the image on the same display but with a larger or smaller size, and so with a different number of samples per picture height. Format conversion involves switching between different formats such as European and North American television standards, which may be displayed at the same size, but again with a different number of samples per picture height and perhaps with different temporal sampling characteristics. This application is one where the picture height is a particularly convenient choice for the unit of length.

An image $f[\mathbf{x}]$ is initially sampled on a lattice Λ , and it is required to produce a new image $g[\mathbf{x}]$ sampled on a different lattice Γ . If Γ is less dense than Λ , we are downsampling; conversely if Γ is denser than Λ , we are upsampling. The *problem* we are addressing is to change the sampling structure of an image with minimal loss of information and minimal introduction of artifacts.

We will consider three cases:

1. upsampling to a superlattice

2. downsampling to a sublattice
3. arbitrary lattice conversion

Before addressing these topics, we need to introduce some additional results on lattices, specifically related to sublattices.

6.1 Sublattices

In order to study upsampling and downsampling, we introduce the notion of a sublattice and establish certain properties.

Definition: Λ is a *sublattice* of Γ if both Λ and Γ are lattices in \mathbb{R}^D , and every point of Λ is also a point of Γ . We write $\Lambda \subset \Gamma$. If Λ is a sublattice of Γ , then Γ is a *superlattice* of Λ .

It is generally easy to observe if a lattice is a sublattice of another lattice by sketching the points of the two lattices on a common set of axes. However, the following shows how to determine this numerically without having to sketch the lattices, which is particularly helpful in three or more dimensions.

Theorem 6.1. The lattice $\Lambda = \text{LAT}(\mathbf{V}_\Lambda)$ is a sublattice of $\Gamma = \text{LAT}(\mathbf{V}_\Gamma)$ if and only if $(\mathbf{V}_\Gamma)^{-1}\mathbf{V}_\Lambda$ is an integer matrix.

Proof. Assume that $\Lambda \subset \Gamma$. We can write $\Lambda = \{\mathbf{V}_\Lambda \mathbf{n} \mid \mathbf{n} \in \mathbb{Z}^D\}$. Then for any $\mathbf{n} \in \mathbb{Z}^D$, $\mathbf{V}_\Lambda \mathbf{n} = \mathbf{V}_\Gamma \mathbf{m}$ for some integer vector $\mathbf{m} \in \mathbb{Z}^D$. Thus $\mathbf{V}_\Gamma^{-1}\mathbf{V}_\Lambda \mathbf{n}$ is an integer vector for any $\mathbf{n} \in \mathbb{Z}^D$. In particular, if \mathbf{n} consists of 1 in position k and 0 elsewhere, then $\mathbf{V}_\Gamma^{-1}\mathbf{V}_\Lambda \mathbf{n}$ is the k^{th} column of $\mathbf{V}_\Gamma^{-1}\mathbf{V}_\Lambda$ which must be an integer vector. Since this holds for all k from 1 to D , all columns of $\mathbf{V}_\Gamma^{-1}\mathbf{V}_\Lambda$ are integer, and so $\mathbf{V}_\Gamma^{-1}\mathbf{V}_\Lambda$ is an integer matrix.

Conversely, if $\mathbf{V}_\Gamma^{-1}\mathbf{V}_\Lambda$ is an integer matrix \mathbf{M} , then $\mathbf{V}_\Lambda = \mathbf{V}_\Gamma \mathbf{M}$. It follows that $\mathbf{V}_\Lambda \mathbf{n} = \mathbf{V}_\Gamma \mathbf{M} \mathbf{n} = \mathbf{V}_\Gamma \mathbf{m} \in \Gamma$ for any integer vector \mathbf{n} . Thus $\Lambda \subset \Gamma$. \square

Corollary: If $\Lambda \subset \Gamma$, then $d(\Lambda)$ is an integer multiple of $d(\Gamma)$.

Proof. If $\Lambda \subset \Gamma$, then $\mathbf{V}_\Lambda = \mathbf{V}_\Gamma \mathbf{M}$ for some integer matrix \mathbf{M} . Then

$$\begin{aligned} d(\Lambda) &= |\det(\mathbf{V}_\Lambda)| = |\det(\mathbf{V}_\Gamma \mathbf{M})| \\ &= |\det(\mathbf{V}_\Gamma)| |\det(\mathbf{M})| = |\det(\mathbf{M})| d(\Gamma) \end{aligned} \quad (6.1)$$

and the determinant of any integer matrix is an integer. \square

Thus, by examining $d(\Lambda)$ and $d(\Gamma)$, we can immediately see which sublattice relation is possible, if any.

Example 6.1. Let

$$\mathbf{V}_\Lambda = \begin{bmatrix} 2X & 0 \\ 0 & 2Y \end{bmatrix} \quad \text{and} \quad \mathbf{V}_\Gamma = \begin{bmatrix} 2X & X \\ 0 & Y \end{bmatrix}.$$

Then, $d(\Lambda) = 4XY$ and $d(\Gamma) = 2XY$, so that $d(\Lambda) = 2d(\Gamma)$. Thus, $\Lambda \subset \Gamma$ is possible. Checking,

$$\mathbf{V}_\Gamma^{-1} \mathbf{V}_\Lambda = \begin{bmatrix} \frac{1}{2X} & -\frac{1}{2Y} \\ 0 & \frac{1}{Y} \end{bmatrix} \begin{bmatrix} 2X & 0 \\ 0 & 2Y \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}$$

which is an integer matrix, and thus $\Lambda \subset \Gamma$ is confirmed. These ideas are illustrated in Fig. 6.1, where the fact that $\Lambda \subset \Gamma$ is clear from the illustration.

Now, let's see what happens in the frequency domain.

Theorem 6.2. If $\Lambda \subset \Gamma$, then $\Gamma^* \subset \Lambda^*$.

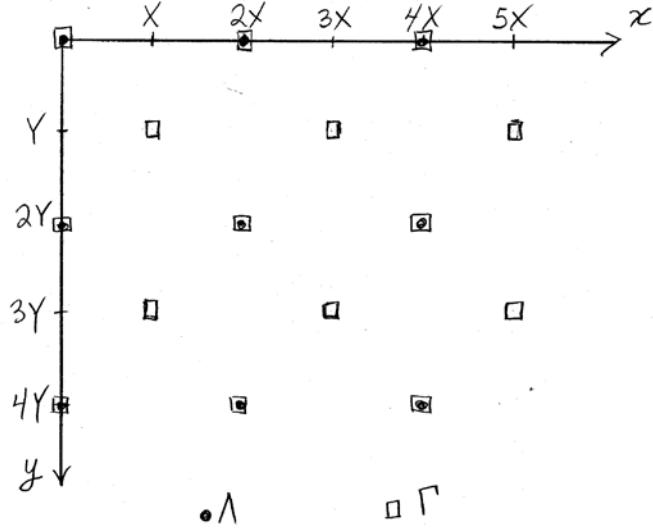
Proof. Suppose that $\mathbf{k} \in \Gamma^*$. Then, by definition, $\mathbf{k} \cdot \mathbf{x} \in \mathbb{Z}$ for all $\mathbf{x} \in \Gamma$. Since, $\Lambda \subset \Gamma$, it follows that $\mathbf{k} \cdot \mathbf{x} \in \mathbb{Z}$ for all $\mathbf{x} \in \Lambda$. Thus $\mathbf{k} \in \Lambda^*$ and so $\Gamma^* \subset \Lambda^*$. \square

Example 6.1 (continued). A sampling matrix for Λ^* is

$$\mathbf{V}_\Lambda^{-T} = \begin{bmatrix} \frac{1}{2X} & 0 \\ 0 & \frac{1}{2Y} \end{bmatrix} = \mathbf{V}_{\Lambda^*}$$

and a sampling matrix for Γ^* is

$$\mathbf{V}_\Gamma^{-T} = \begin{bmatrix} \frac{1}{2X} & 0 \\ -\frac{1}{2Y} & \frac{1}{Y} \end{bmatrix} = \mathbf{V}_{\Gamma^*}.$$

Figure 6.1: Illustration of lattices Λ and Γ of Example 6.1, where $\Lambda \subset \Gamma$.

We see that

$$(\mathbf{V}_{\Lambda^*})^{-1} \mathbf{V}_{\Gamma^*} = \begin{bmatrix} 2X & 0 \\ 0 & 2Y \end{bmatrix} \begin{bmatrix} \frac{1}{2X} & 0 \\ -\frac{1}{2Y} & \frac{1}{Y} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix}$$

which is an integer matrix, confirming that $\Gamma^* \subset \Lambda^*$. Note that $d(\Gamma^*) = \frac{1}{2XY}$ and $d(\Lambda^*) = \frac{1}{4XY}$, and so $d(\Gamma^*) = 2d(\Lambda^*)$. In fact, in general,

$$\frac{d(\Lambda)}{d(\Gamma)} = \frac{d(\Gamma^*)}{d(\Lambda^*)}. \quad (6.2)$$

6.2 Upsampling

Suppose that we want to upsample an image $f[\mathbf{x}]$ defined on the lattice Λ in Fig. 6.1 to the superlattice Γ , a 2:1 upsampling. In the frequency domain, $F(\mathbf{u})$ is periodic, with periodicity given by the reciprocal lattice Λ^* , as indicated in Fig. 6.2. The desired upsampled signal $g[\mathbf{x}]$ defined on Γ has a Fourier transform $G(\mathbf{u})$ that has periodicity given by Γ^* . Both $f[\mathbf{x}]$ and $g[\mathbf{x}]$ will correspond to the same continuous-space image if we obtain $G(\mathbf{u})$ from $F(\mathbf{u})$ by setting the shaded replicas in Fig. 6.2

to zero. How can this be done?

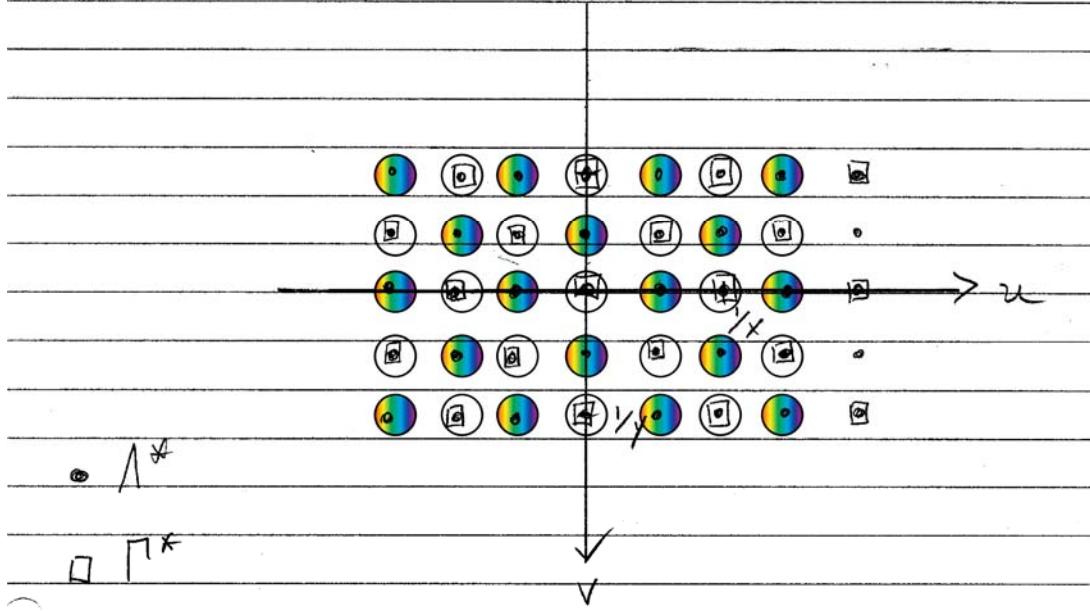


Figure 6.2: Illustration of reciprocal lattices Λ^* and Γ^* of Example 6.1, where $\Gamma^* \subset \Lambda^*$.

It is done in a two-step process. Let \mathcal{S}_Λ and \mathcal{S}_Γ be suitable spaces of signals defined on lattices Λ and Γ respectively. First $f[\mathbf{x}]$ is upsampled to Γ by inserting zeros at all the points in Γ that do not belong to Λ . We define the zero-insertion upsampler $\mathcal{U}_{\Lambda \uparrow \Gamma} : \mathcal{S}_\Lambda \rightarrow \mathcal{S}_\Gamma$ by

$$q = \mathcal{U}_{\Lambda \uparrow \Gamma} f : q[\mathbf{x}] = \begin{cases} f[\mathbf{x}], & \mathbf{x} \in \Lambda; \\ 0, & \mathbf{x} \in \Gamma \setminus \Lambda. \end{cases} \quad (6.3)$$

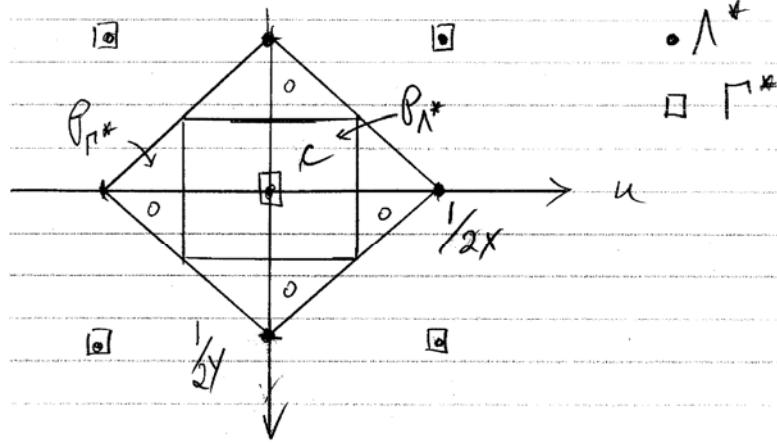


Figure 6.3: Illustration of unit cells of reciprocal lattices Λ^* and Γ^* of Example 6.1.

What does this do in the frequency domain?

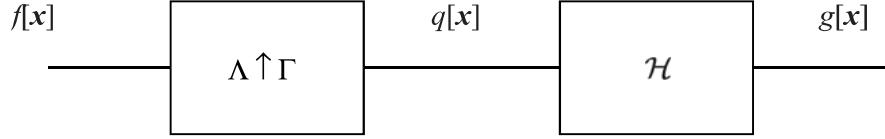
$$\begin{aligned}
 Q(\mathbf{u}) &= \sum_{\mathbf{x} \in \Gamma} q[\mathbf{x}] \exp(-j2\pi \mathbf{u} \cdot \mathbf{x}) \\
 &= \sum_{\mathbf{x} \in \Lambda} f[\mathbf{x}] \exp(-j2\pi \mathbf{u} \cdot \mathbf{x}) \\
 &= F(\mathbf{u}).
 \end{aligned} \tag{6.4}$$

Thus $q[\mathbf{x}]$ and $f[\mathbf{x}]$ have identical Fourier transforms — but $q[\mathbf{x}]$ is defined on Γ and so it can now be processed by a filter defined on Γ . If this filter is $h[\mathbf{x}]$, $\mathbf{x} \in \Gamma$, we know that its frequency response is specified by its values over one unit cell of Γ^* . In the ideal case, we want $H(\mathbf{u})$ to be equal to a constant c on the unit cell of Λ^* and zero elsewhere in a unit cell of Γ^* ,

$$H(\mathbf{u}) = \begin{cases} c & \mathbf{u} \in \mathcal{P}_{\Lambda^*}; \\ 0 & \mathbf{u} \in \mathcal{P}_{\Gamma^*} \setminus \mathcal{P}_{\Lambda^*}, \end{cases} \tag{6.5}$$

where \mathcal{P}_{Λ^*} and \mathcal{P}_{Γ^*} are unit cells of Λ^* and Γ^* respectively. This is illustrated in Fig. 6.3 for our example.

What should the constant c be? We want $f[\mathbf{x}]$ and $g[\mathbf{x}]$ to correspond to the

Figure 6.4: Block diagram of upsampler from Λ to superlattice Γ .

same continuous-domain signal $f_c(\mathbf{x})$. In this case,

$$F(\mathbf{u}) = \frac{1}{d(\Lambda)} F_c(\mathbf{u}) \quad \mathbf{u} \in \mathcal{P}_{\Lambda^*} \quad (6.6)$$

$$\text{and } G(\mathbf{u}) = \frac{1}{d(\Gamma)} F_c(\mathbf{u}) \quad \mathbf{u} \in \mathcal{P}_{\Gamma^*}, \quad (6.7)$$

where we assume that $F_c(\mathbf{u}) = 0$ for $\mathbf{u} \in \mathbb{R}^D \setminus \mathcal{P}_{\Lambda^*}$. This leads us to conclude that

$$G(\mathbf{u}) = \begin{cases} \frac{d(\Lambda)}{d(\Gamma)} F(\mathbf{u}), & \mathbf{u} \in \mathcal{P}_{\Lambda^*}; \\ 0, & \mathbf{u} \in \mathcal{P}_{\Gamma^*} \setminus \mathcal{P}_{\Lambda^*}, \end{cases} \quad (6.8)$$

and so $c = \frac{d(\Lambda)}{d(\Gamma)}$, which is the interpolation ratio.

An upsampler can thus be implemented using the arrangement shown in Fig. 6.4, where the zero-insertion operator $\mathcal{U}_{\Lambda \uparrow \Gamma}$ is defined in equation (6.3) and \mathcal{H} is the LSI filter defined in equation (6.5). The overall upsampler is $\mathcal{H}_u = \mathcal{H}\mathcal{U}_{\Lambda \uparrow \Gamma}$, which can easily be seen to be a linear system. It is also shift invariant, in the sense that $\mathcal{H}_u \mathcal{T}_{\mathbf{d}} f = \mathcal{T}_{\mathbf{d}} \mathcal{H}_u f$ for any $\mathbf{d} \in \Lambda$. Because $\Lambda \subset \Gamma$, the shift system $\mathcal{T}_{\mathbf{d}}$ is well defined on both \mathcal{S}_{Λ} and \mathcal{S}_{Γ} for $\mathbf{d} \in \Lambda$. Of course in practice $H(\mathbf{u})$ will not be an ideal low-pass filter, but it will be an approximation designed in some way.

Note that an image usually has a lot of energy concentrated at DC, due to a nonzero average brightness. Thus, there is a lot of energy concentrated at points of Λ^* , and consequently, referring to Fig. 6.3, we want

$$H(\mathbf{u}) = 0 \quad \text{for } \mathbf{u} \in \Lambda^* \setminus \Gamma^* \quad (6.9)$$

(the four vertices of the diamond). This should be enforced as a constraint in the filter design process.

Example 6.2. Design a system to upsample an image defined on a square lattice $\Lambda = \text{LAT}(\text{diag}(X, X))$ by a factor of two in each dimension to obtain a new image defined on the square lattice $\Gamma = \text{LAT}(\text{diag}(X/2, X/2))$.

Solution. We have $d(\Lambda) = X^2$ and $d(\Gamma) = X^2/4$, so that the interpolation ratio is $\frac{d(\Lambda)}{d(\Gamma)} = 4$. Thus, the ideal interpolation filter will be

$$H_I(\mathbf{u}) = \begin{cases} 4 & \mathbf{u} \in \mathcal{P}_{\Lambda^*} \\ 0 & \mathbf{u} \in \mathcal{P}_{\Gamma^*} \setminus \mathcal{P}_{\Lambda^*} \end{cases}$$

The situation is depicted in the frequency domain in Fig. 6.5. Since the pass-

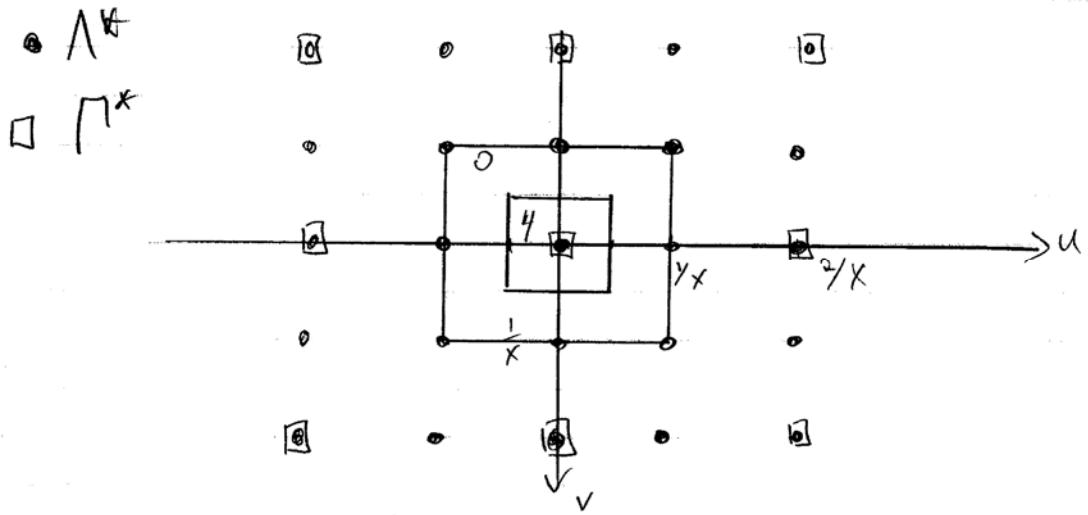


Figure 6.5: Upsampling of rectangulary sampled signal by a factor of 2 in each dimension; frequency domain view.

band \mathcal{P}_{Λ^*} is rectangular in shape, we can use a separable filter. The following one-dimensional filters are suitable to form a separable two-dimensional interpola-

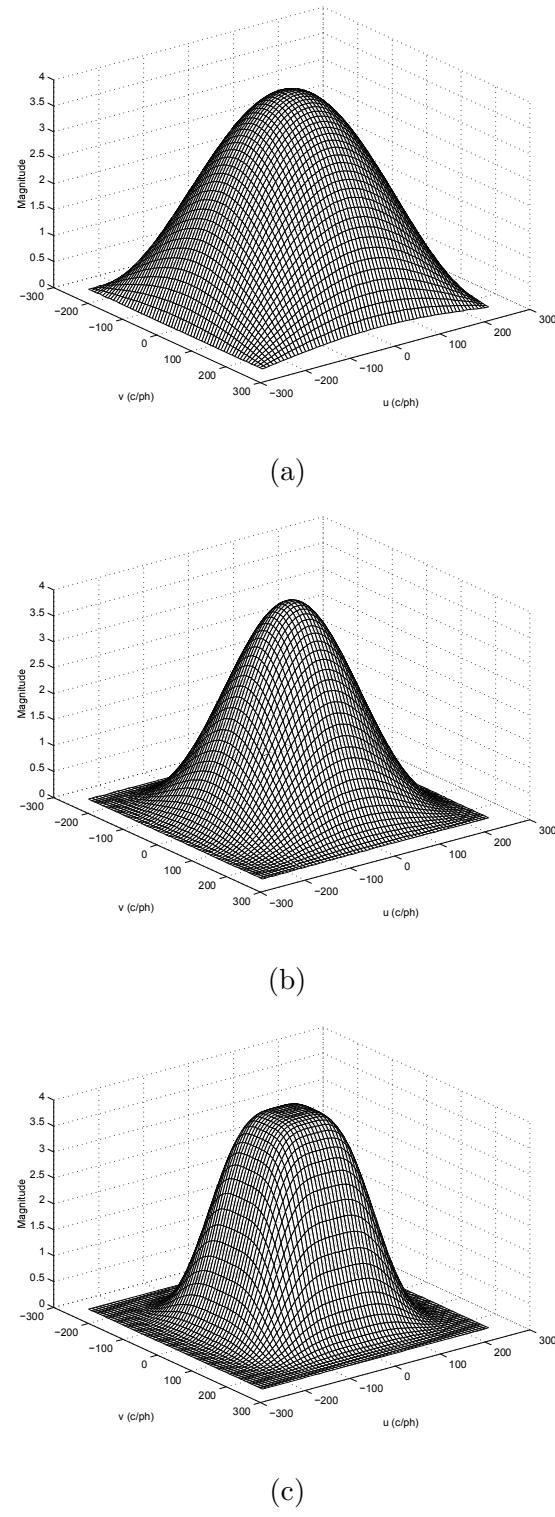
tor.

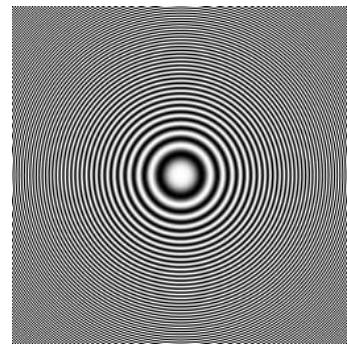
$$\mathbf{h}_{10} = \begin{bmatrix} 0 & \frac{X}{2} \\ 1 & 1 \end{bmatrix}$$

$$\mathbf{h}_{11} = \begin{bmatrix} -\frac{X}{2} & 0 & \frac{X}{2} \\ \frac{1}{2} & 1 & \frac{1}{2} \end{bmatrix}$$

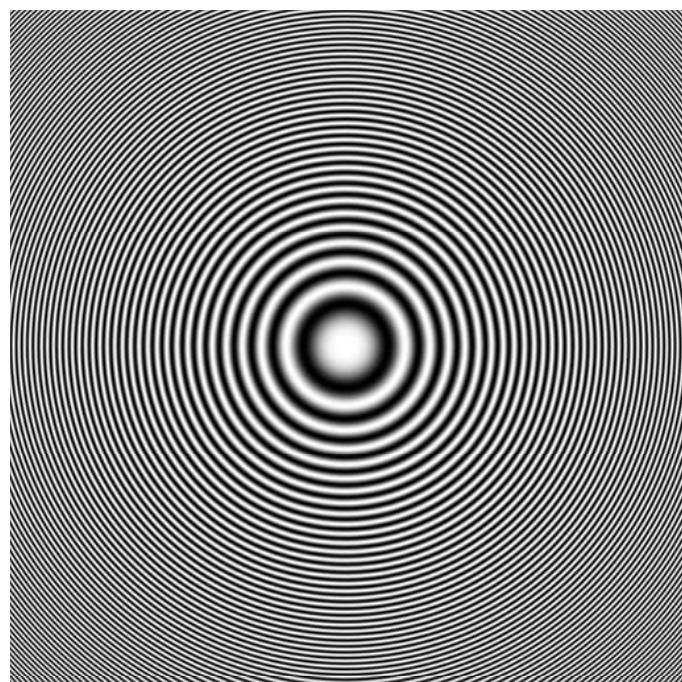
$$\mathbf{h}_{13} = \begin{bmatrix} -\frac{3X}{2} & -X & -\frac{X}{2} & 0 & \frac{X}{2} & X & \frac{3X}{2} \\ -\frac{1}{16} & 0 & \frac{9}{16} & 1 & \frac{9}{16} & 0 & -\frac{1}{16} \end{bmatrix}$$

Each of these 1D interpolators leaves non-zero samples of $q[\mathbf{x}]$ untouched, has a DC gain of 2.0, and has a null response at $u = 1/X$. The two-dimensional interpolation filters are defined by $h_i[x, y] = h_{1i}[x]h_{1i}[y]$ for $i = 0, 1, 3$. Fig. 6.6 shows a perspective view of the frequency response of these three separable filters. Fig. 6.7 shows the result of upsampling a zoneplate by a factor of 2 in each dimension using these three interpolation filters. Fig. 6.7(a) shows the original zoneplate, Fig. 6.7(b) shows the ideal result (which we know since we are dealing with an analytically defined function), and Fig. 6.7(c)-(e) shows the result of interpolating using h_0 , h_1 and h_3 respectively. The successive improvement in going from h_0 to h_1 to h_3 can be observed if these critical images are viewed correctly on a high-quality display. \square

Figure 6.6: Separable interpolation filters. (a) $H_0(u, v)$. (b) $H_1(u, v)$. (c) $H_3(u, v)$.

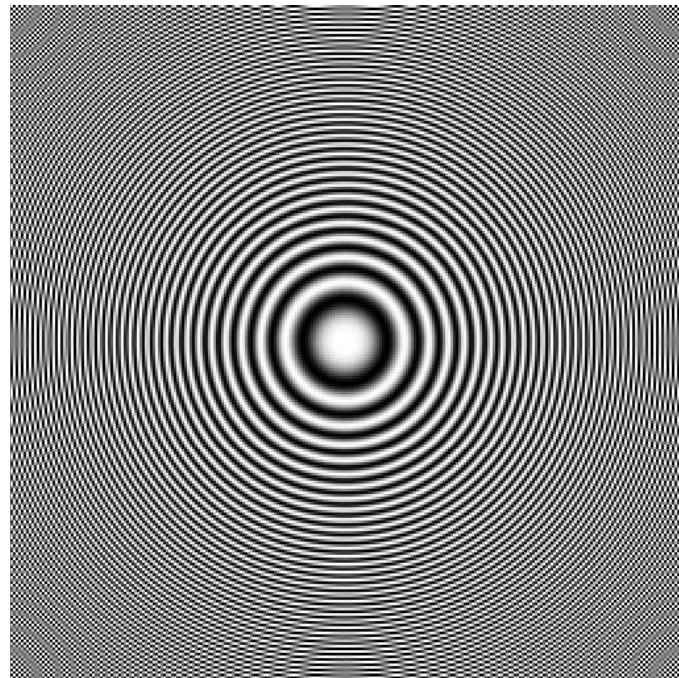


(a)

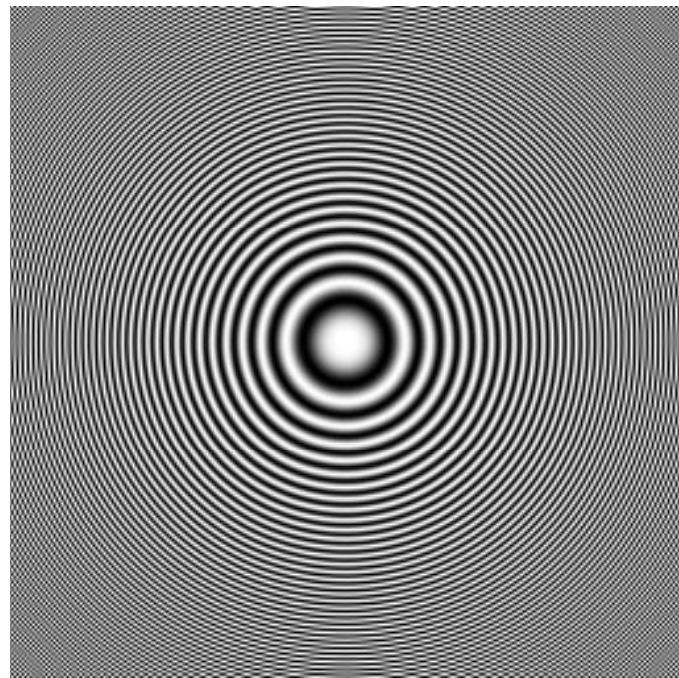


(b)

Figure 6.7: (a) Original 256×256 zoneplate image. (b) Ideal 512×512 zoneplate image.



(c)



(d)

Figure 6.7 (c) Zoneplate interpolated with h_0 . (d) Zoneplate interpolated with h_1 .

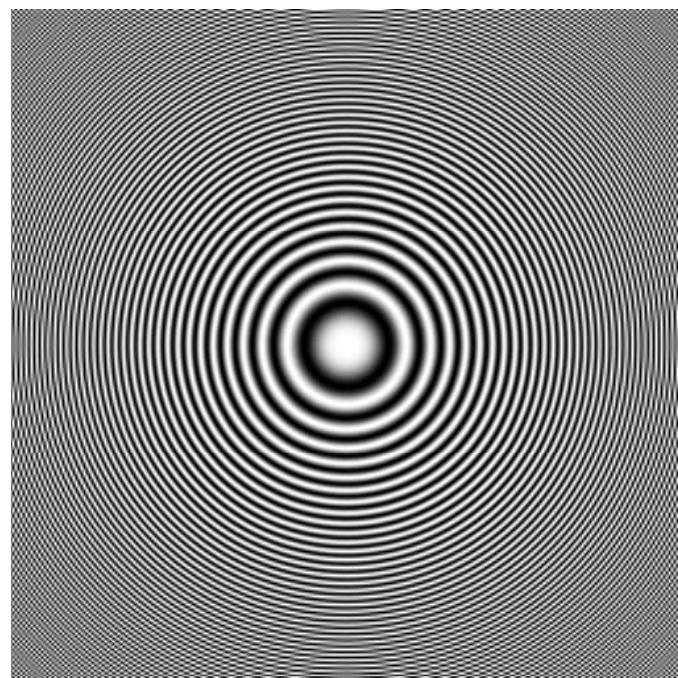


Figure 6.7 (e) Zoneplate interpolated with h_3 .

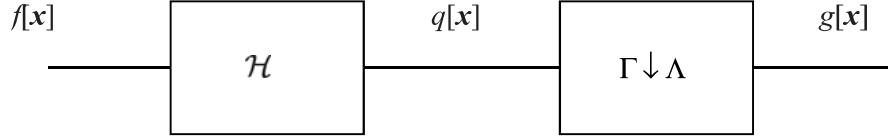


Figure 6.8: Block diagram of downampler.

6.3 Downsampling

The reverse operation to upsampling is downsampling. Assume that f is sampled on a lattice Γ and that $\Lambda \subset \Gamma$ is a sublattice of Γ . The downsampled image is

$$g[\mathbf{x}] = f[\mathbf{x}], \quad \mathbf{x} \in \Lambda \quad (6.10)$$

where the samples on $\Gamma \setminus \Lambda$ are discarded. Referring to Fig. 6.2, it is clear that the subsampling will introduce extra spectral replicas on the points of $\Lambda^* \setminus \Gamma^*$ (the shaded replicas). In order to avoid aliasing, we need to prefilter the image so that its Fourier transform is confined to a unit cell of Λ^* before subsampling. Thus, a subsampling system has the structure shown in Fig. 6.8.

It can be shown that we can write

$$\Lambda^* = \bigcup_{k=1}^K (\mathbf{d}_k + \Gamma^*) \quad (6.11)$$

where $K = d(\Lambda)/d(\Gamma)$, for a suitably chosen set of $\mathbf{d}_k \in \Lambda^*$. For example, in Fig. 6.5 we can use $\mathbf{d}_1 = [0 \ 0]^T$, $\mathbf{d}_2 = [\frac{1}{X} \ 0]^T$, $\mathbf{d}_3 = [0 \ \frac{1}{X}]^T$, $\mathbf{d}_4 = [\frac{1}{X} \ \frac{1}{X}]^T$. Then, with no prefiltering, it can be shown that

$$G(\mathbf{u}) = \frac{1}{K} \sum_{k=1}^K F(\mathbf{u} + \mathbf{d}_k) \quad (6.12)$$

and with prefiltering,

$$G(\mathbf{u}) = \frac{1}{K} \sum_{k=1}^K H(\mathbf{u} + \mathbf{d}_k)F(\mathbf{u} + \mathbf{d}_k). \quad (6.13)$$

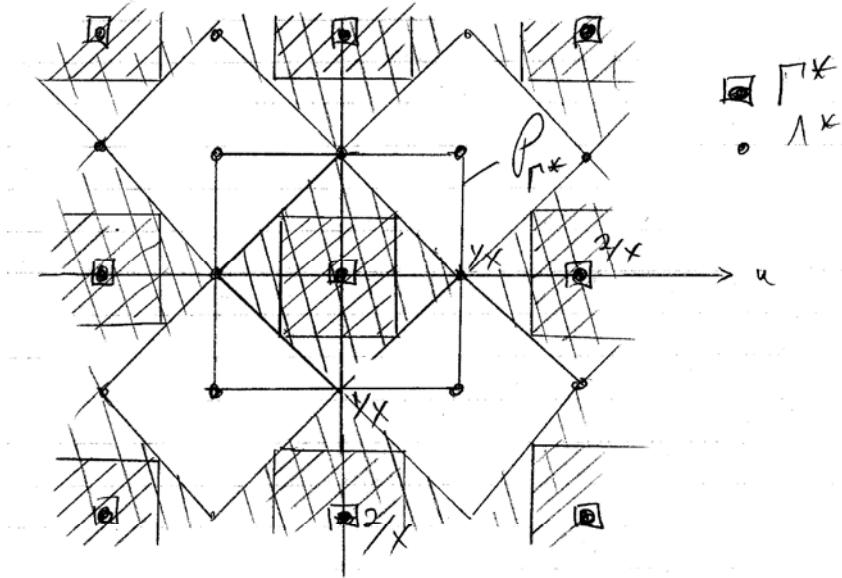


Figure 6.9: Illustration of downsampling in the frequency domain.

The ideal antialiasing prefilter has frequency response

$$H(\mathbf{u}) = \begin{cases} 1 & \mathbf{u} \in \mathcal{P}_{\Lambda^*} \\ 0 & \mathbf{u} \in \mathcal{P}_{\Gamma^*} \setminus \mathcal{P}_{\Lambda^*}. \end{cases} \quad (6.14)$$

We can illustrate again with the example of Fig. 6.5. Suppose that $f[\mathbf{x}]$ is defined on Γ and has Fourier transform confined to the diamond-shaped region in Fig. 6.9 (single hatch), which is well within the unit cell \mathcal{P}_{Γ^*} of Γ^* . However, if we directly subsample to Λ^* , there will be serious aliasing. We must prefilter $f[\mathbf{x}]$ so that its Fourier transform is limited to a unit cell of Λ^* , the double-crosshatched square region in Fig. 6.9.

6.4 Arbitrary Sampling Structure Conversion

What can we do if neither the input nor the output lattice is a sublattice of the other one? There are two common approaches that we present here. Suppose that the input signal is sampled on the lattice Λ_1 and the output is sampled on the lattice

Λ_2 , where $\Lambda_1 \not\subset \Lambda_2$ and $\Lambda_2 \not\subset \Lambda_1$. In the first approach, we try to find a third lattice Λ_3 that is a superlattice of both Λ_1 and Λ_2 . We can then upsample the input signal from Λ_1 to Λ_3 , and then downsample from Λ_3 to Λ_2 . The second approach is to use polynomial interpolation.

6.4.1 Sampling structure conversion using a common superlattice

Assume that we can find a common superlattice Λ_3 such that $\Lambda_1 \subset \Lambda_3$ and $\Lambda_2 \subset \Lambda_3$. Assuming that such a common superlattice exists, we want to find the least-dense common superlattice. In the two-step procedure, the input signal is first upsampled from Λ_1 to Λ_3 using the procedure of Section 6.2. This intermediate signal is then downsampled from Λ_3 to Λ_2 , with appropriate prefiltering if necessary, using the procedure of Section 6.3. Combining the block diagrams of Fig. 6.8 and Fig. 6.4, we get the system shown in Fig. 6.10. Of course, the two filters \mathcal{H}_1 and \mathcal{H}_2 can be combined into a single filter \mathcal{H} . Combining Eq. (6.8) and Eq. (6.14), we find that the ideal conversion filter is given by

$$H(\mathbf{u}) = H_1(\mathbf{u})H_2(\mathbf{u}) = \begin{cases} \frac{d(\Lambda_1)}{d(\Lambda_3)} & \mathbf{u} \in \mathcal{P}_{\Lambda_1^*} \cap \mathcal{P}_{\Lambda_2^*} \\ 0 & \mathbf{u} \in \mathcal{P}_{\Lambda_3^*} \setminus (\mathcal{P}_{\Lambda_1^*} \cap \mathcal{P}_{\Lambda_2^*}). \end{cases} \quad (6.15)$$

An example will serve to clarify the ideas.

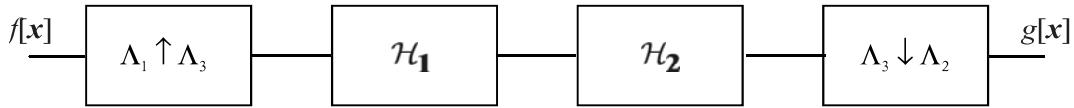


Figure 6.10: Block diagram of sampling structure conversion system.

Example 6.3. Design a system to convert a signal defined on the lattice $\Lambda_1 = \text{LAT} \left(\begin{bmatrix} X & 0 \\ Y & 2Y \end{bmatrix} \right) = \text{LAT}(\mathbf{V}_1)$ to a signal defined on the lattice $\Lambda_2 = \text{LAT} \left(\begin{bmatrix} 2X & X \\ 0 & 2Y \end{bmatrix} \right) = \text{LAT}(\mathbf{V}_2)$.

Solution. $d(\Lambda_1) = 2XY$ and $d(\Lambda_2) = 4XY$ so Λ_2 has a lower sampling density than

Λ_1 . Is $\Lambda_2 \subset \Lambda_1$? Performing the test, we see that

$$\mathbf{V}_1^{-1} \mathbf{V}_2 = \begin{bmatrix} \frac{1}{X} & 0 \\ -\frac{1}{2X} & \frac{1}{2Y} \end{bmatrix} \begin{bmatrix} 2X & X \\ 0 & 2Y \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ -1 & \frac{1}{2} \end{bmatrix}$$

which is *not* an integer matrix. Thus $\Lambda_2 \not\subset \Lambda_1$. In fact, the situation can be seen in Fig. 6.11.

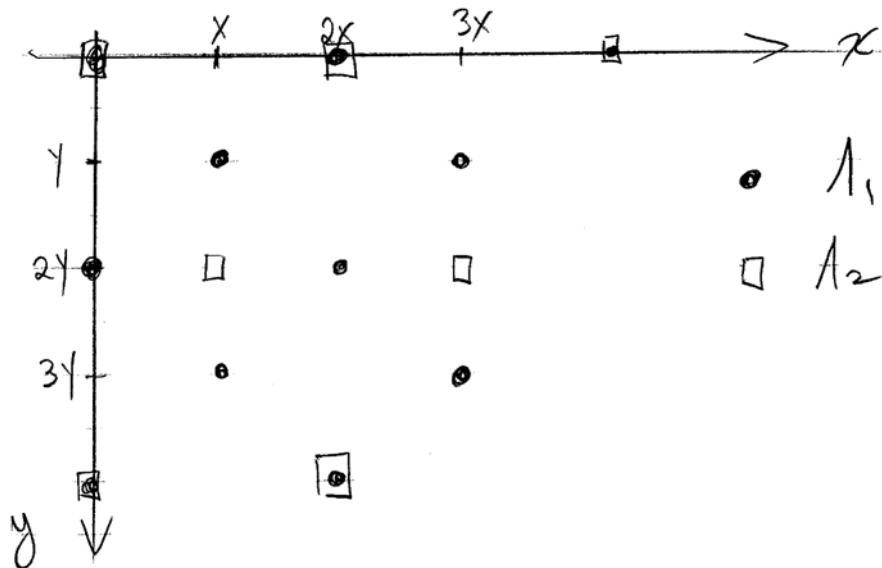


Figure 6.11: Lattices Λ_1 and Λ_2 for example 6.3.

By inspection of Fig. 6.11, we see that the least dense lattice that contains both Λ_1 and Λ_2 is $\Lambda_3 = \text{LAT} \left(\begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix} \right)$. Thus, the first step is to upsample the signal from Λ_1 to Λ_3 . The situation is illustrated in the frequency domain in Fig. 6.12. The spectral replicas centered on points of Λ_1^* that are not in Λ_3^* must be removed by a low-pass filter defined on Λ_3 . These are *not* hatched in Fig. 6.12.

In the second step, the signal is downsampled from Λ_3 to Λ_2 . The effect of this is illustrated in the frequency domain in Fig. 6.13. The subsampling introduces additional copies of the spectrum of the signal at points of Λ_2^* that are not in Λ_3^* . This will clearly introduce aliasing if prefiltering is not applied. Referring to Fig. 6.13, a suitable prefilter has passband as shown hatched.

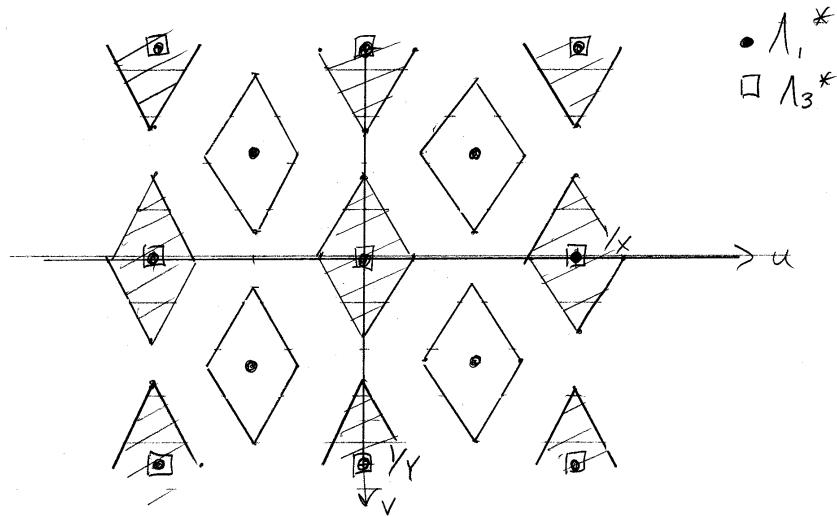


Figure 6.12: Frequency domain situation for upconversion from Λ_1 to Λ_3 in example 6.3.

□

6.4.2 Polynomial interpolation

For separable, rectangular sampling, sample structure conversion is often performed using polynomial interpolation. This can be used to compute values at arbitrary locations, not just on a sublattice or superlattice. It gives a rapid way to resize an image by an arbitrary factor. The most frequently used polynomial interpolators are zero-order hold, linear (straight-line) interpolation, and cubic interpolation. Since they are applied separately, they can be studied in one dimension. This approach can be thought of as producing a continuous version of the image and resampling.

Polynomial interpolation is done piecewise: a different polynomial is used to construct each segment of length X . This is evident from the illustrations of zero-order hold and linear, straight-line interpolation shown in Fig. 6.14.

We can write piecewise expressions for the interpolated signals for zero order

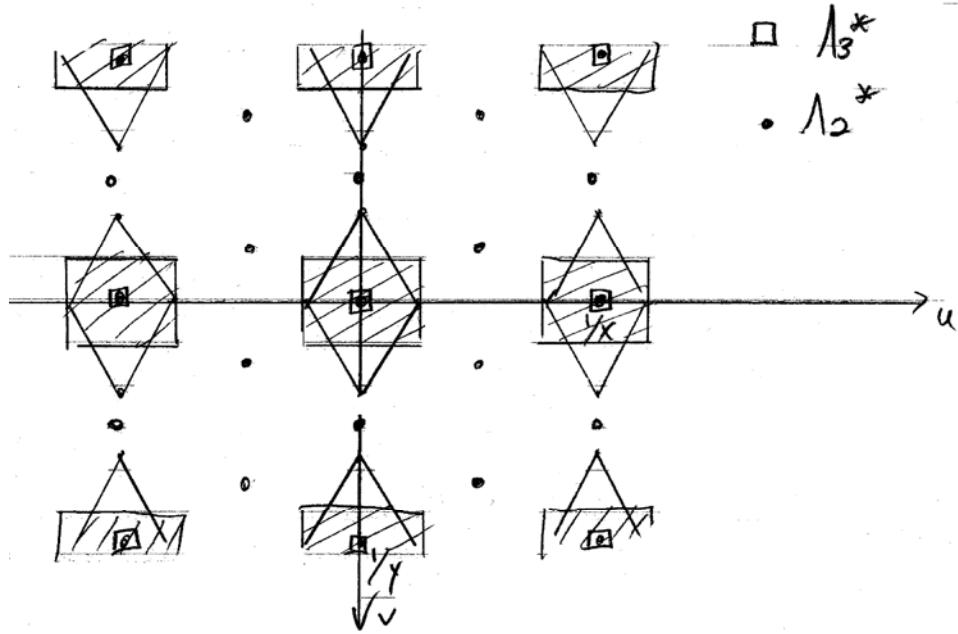


Figure 6.13: Frequency domain situation for downconversion from Λ_3 to Λ_2 in example 6.3.

hold:

$$f_c^0(x) = f[kX] \quad \text{if} \quad kX - \frac{X}{2} < x \leq kX + \frac{X}{2} \quad (6.16)$$

and for linear straight-line interpolation:

$$f_c^1(x) = \left(1 - \frac{x - kX}{X}\right) f[kX] + \left(\frac{x - kX}{X}\right) f[(k+1)X] \quad \text{if} \quad kX < x \leq (k+1)X. \quad (6.17)$$

We can write these formulas in the form of a convolution:

$$f_0^m(x) = \sum_{k=-\infty}^{\infty} f[kX] h_m(x - kX). \quad (6.18)$$

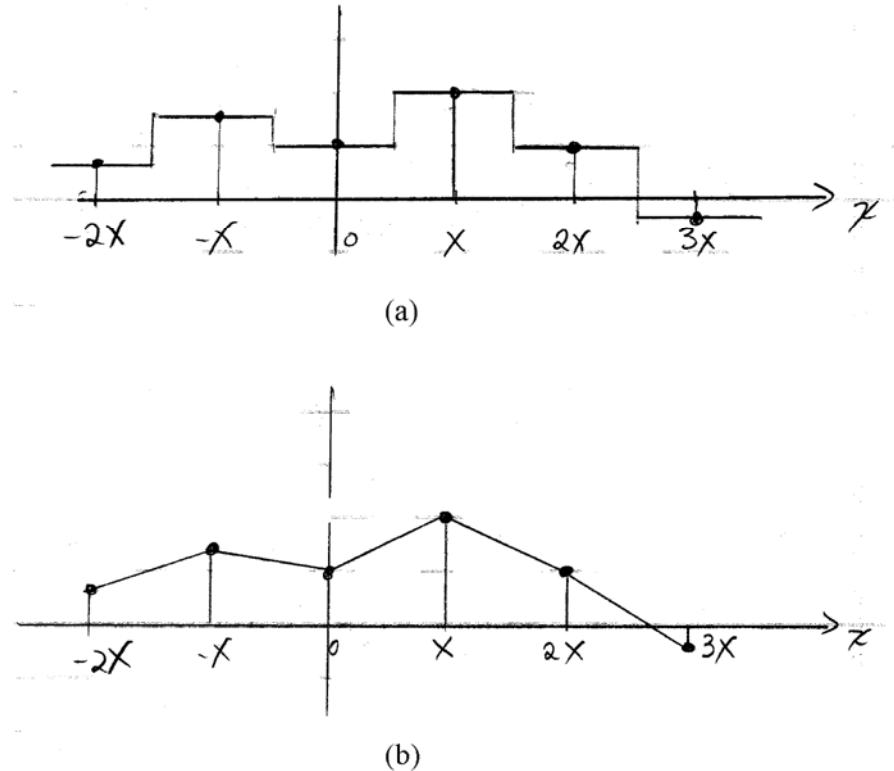


Figure 6.14: Illustration of (a) zero-order hold and (b) linear, straight-line interpolation.

For the two cases considered above, we have

$$h_0(x) = \begin{cases} 1 & \text{if } -\frac{X}{2} < x \leq \frac{X}{2} \\ 0 & \text{otherwise} \end{cases} \quad (6.19)$$

$$h_1(x) = \begin{cases} 1 - \frac{|x|}{X} & \text{if } |x| \leq X \\ 0 & \text{otherwise.} \end{cases} \quad (6.20)$$

This operation is illustrated graphically in Fig. 6.15.

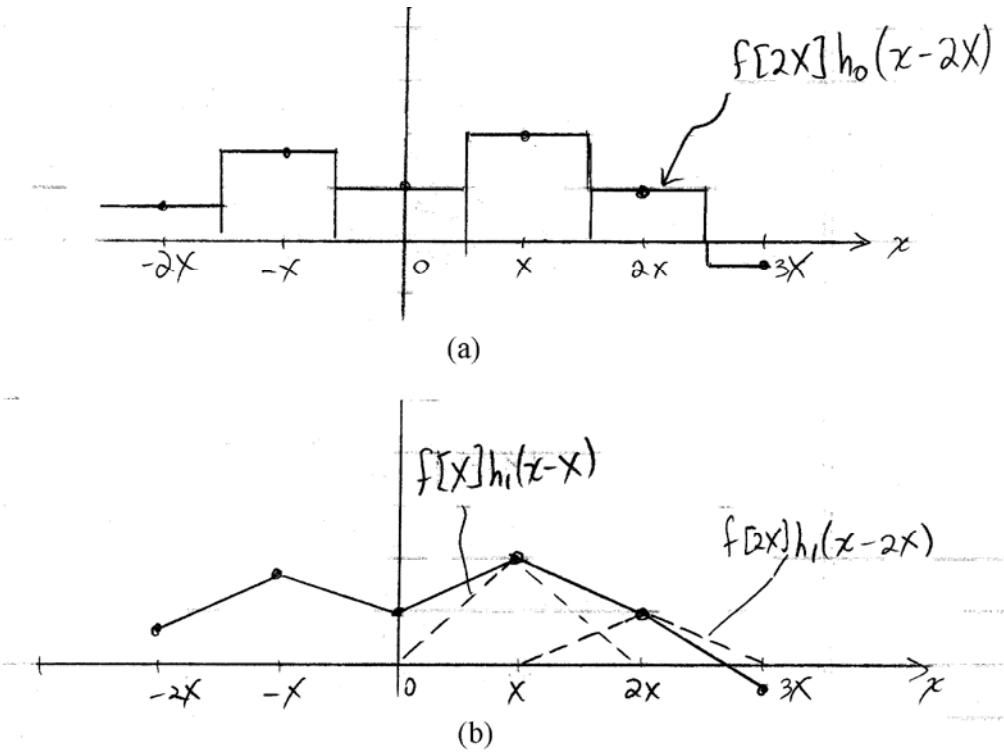


Figure 6.15: Details of interpolation using (a) zero-order hold and (b) linear, straight-line interpolation.

A straightforward analysis shows that

$$F_c^m(u) = F(u)H_m(u) \quad (6.21)$$

where $H_m(u)$ has the characteristics of a low-pass filter for these polynomial interpolators. We easily find that

$$H_0(u) = \frac{\sin(\pi u X)}{\pi u} \quad (6.22)$$

$$\text{and} \quad H_1(u) = X \frac{\sin^2(\pi u X)}{(\pi u X)^2} \quad (6.23)$$

where the latter is a better low-pass filter. Even better results can be obtained with

cubic interpolation:

$$h_3(x) = \begin{cases} 1.5 \left(\frac{|x|}{X}\right)^3 - 2.5 \left(\frac{|x|}{X}\right)^2 + 1 & \text{if } 0 \leq |x| < X \\ -0.5 \left(\frac{|x|}{X}\right)^3 + 2.5 \left(\frac{|x|}{X}\right)^2 - 4 \left(\frac{|x|}{X}\right) + 2 & \text{if } X \leq |x| < 2X \\ 0 & \text{otherwise.} \end{cases} \quad (6.24)$$

MATLAB implements these polynomial interpolators using the `imresize` function, using parameters ‘nearest’, ‘bilinear’ and ‘bicubic’ respectively.

Problems

- For each of the following pairs of lattices Λ_1 and Λ_2 , state whether $\Lambda_1 \subset \Lambda_2$, $\Lambda_2 \subset \Lambda_1$ or neither. If neither, find (by inspection) the least dense lattice Λ_3 such that $\Lambda_1 \subset \Lambda_3$ and $\Lambda_2 \subset \Lambda_3$. For each lattice Λ_1 , Λ_2 and Λ_3 (if required), determine and sketch the reciprocal lattice and a unit cell of the reciprocal lattice.

$$(a) \quad V_{\Lambda_1} = \begin{bmatrix} X & 0 \\ 0 & X \end{bmatrix} \quad V_{\Lambda_2} = \begin{bmatrix} 2X & 0 \\ 0 & 2X \end{bmatrix}$$

$$(b) \quad V_{\Lambda_1} = \begin{bmatrix} X & 0 \\ 0 & X \end{bmatrix} \quad V_{\Lambda_2} = \begin{bmatrix} 3X & X \\ 0 & X \end{bmatrix}$$

$$(c) \quad V_{\Lambda_1} = \begin{bmatrix} 2X & 0 \\ 0 & 2X \end{bmatrix} \quad V_{\Lambda_2} = \begin{bmatrix} X & X \\ X & -X \end{bmatrix}$$

$$(d) \quad V_{\Lambda_1} = \begin{bmatrix} 1.5X & 0 \\ 0 & 1.5X \end{bmatrix} \quad V_{\Lambda_2} = \begin{bmatrix} X & 0 \\ 0 & X \end{bmatrix}$$

$$(e) \quad V_{\Lambda_1} = \begin{bmatrix} X & X \\ X & -X \end{bmatrix} \quad V_{\Lambda_2} = \begin{bmatrix} 1.5X & 1.5X \\ 1.5X & -1.5X \end{bmatrix}$$