CHAPTER 2

Analysis of Two-Dimensional Signals and Systems

Many physical phenomena are found experimentally to share the basic property that their response to several stimuli acting simultaneously is identically equal to the sum of the responses that each component stimulus would produce individually. Such phenomena are called *linear,* and the property they share is called *linearity.* Electrical networks composed of resistors, capacitors, and inductors are usually linear over a wide range of inputs. In addition, as we shall soon see, the wave equation describing the propagation of light through most media leads us naturally to regard optical imaging operations as linear mappings of "object" light distributions into "image" light distributions.

The single property of linearity leads to a vast simplification in the mathematical description of such phenomena and represents the foundation of a mathematical structure which we shall refer to here as *linear systems theory.* The great advantage afforded by linearity is the ability to express the response (be it voltage, current, light amplitude, or light intensity) to a complicated stimulus in terms of the responses to certain "elementary" stimuli. Thus if a stimulus is decomposed into a linear combination of elementary stimuli, each of which produces a known response of convenient form, then by virtue of linearity, the total response can be found as a corresponding linear combination of the responses to the elementary stimuli.

In this chapter we review some of the mathematical tools that are useful in describing linear phenomena, and discuss some of the mathematical decompositions that are often employed in their analysis. Throughout the later chapters we shall be concerned with stimuli (system inputs) and responses (system outputs) that may be either of two different physical quantities. If the illumination used in an optical system exhibits a property called *spatial coherence,* then we shall find that it is appropriate to describe the light as a spatial distribution of *complex-valued* field amplitude. When the illumination is totally lacking in spatial coherence, it is appropriate to describe the light as a spatial distribution of *real-valued* intensity. Attention will be focused here on the analysis of linear systems with complex-valued inputs; the results for real-valued inputs are thus included as special cases of the theory.

2.1 FOURIER ANALYSIS IN TWO DIMENSIONS

A mathematical tool of great utility in the analysis of both linear and nonlinear phenomena is *Fourier analysis.* This tool is widely used in the study of electrical networks and communication systems; it is assumed that the reader has encountered Fourier theory previously, and therefore that he or she is familiar with the analysis of functions of one independent variable (e.g. time). For a review of the fundamental mathematical concepts, see the books by Papoulis [226], Bracewell [32], and Gray and Goodman [131]. A particularly relevant treatment is by Bracewell [33]. Our purpose here is limited to extending the reader's familiarity to the analysis of functions of *two* independent variables. No attempt at great mathematical rigor will be made, but rather, an operational approach, characteristic of most engineering treatments of the subject, will be adopted.

2.1.1 Definition and Existence Conditions

The *Fourier transform* (alternatively the *Fourier spectrum* or *frequency spectrum*) of a (in general, complex-valued) function g of two independent variables x and y will be represented here by $\mathcal{F}\{g\}$ and is defined by[1]

$$\mathcal{F}\{g\} = \iint\limits_{-\infty}^{\infty} g(x, y) \exp[-j2\pi(f_X x + f_Y y)]\, dx\, dy. \tag{2-1}$$

The transform so defined is itself a complex-valued function of two independent variables f_X and f_Y, which we generally refer to as *frequencies*. Similarly, the *inverse Fourier transform* of a function $G(f_X, f_Y)$ will be represented by $\mathcal{F}^{-1}\{G\}$ and is defined as

$$\mathcal{F}^{-1}\{G\} = \iint\limits_{-\infty}^{\infty} G(f_X, f_Y) \exp[j2\pi(f_X x + f_Y y)]\, df_X\, df_Y. \tag{2-2}$$

Note that as mathematical operations the transform and inverse transform are very similar, differing only in the sign of the exponent appearing in the integrand. The inverse Fourier transform is sometimes referred to as the *Fourier integral* representation of a function $g(x, y)$.

Before discussing the properties of the Fourier transform and its inverse, we must first decide when (2-1) and (2-2) are in fact meaningful. For certain functions, these integrals may not exist in the usual mathematical sense, and therefore this discussion would be incomplete without at least a brief mention of "existence conditions". While a variety of sets of *sufficient* conditions for the existence of (2-1) are possible, perhaps the most common set is the following:

---

[1]When a single limit of integration appears above or below a double integral, then that limit applies to *both* integrations.

1. g must be absolutely integrable over the infinite (x, y) plane.
2. g must have only a finite number of discontinuities and a finite number of maxima and minima in any finite rectangle.
3. g must have no infinite discontinuities.

In general, any one of these conditions can be weakened at the price of strengthening one or both of the companion conditions, but such considerations lead us rather far afield from our purposes here.

As Bracewell [32] has pointed out, "physical possibility is a valid sufficient condition for the existence of a transform." However, it is often convenient in the analysis of systems to represent true physical waveforms by idealized mathematical functions, and for such functions one or more of the above existence conditions may be violated. For example, it is common to represent a strong, narrow time pulse by the so-called Dirac delta function[2] often represented by

$$\delta(t) = \lim_{N\to\infty} N \exp(-N^2\pi t^2), \tag{2-3}$$

where the limit operation provides a convenient mental construct but is not meant to be taken literally. See Appendix A for more details. Similarly, an idealized point source of light is often represented by the two-dimensional equivalent,

$$\delta(x, y) = \lim_{N\to\infty} N^2 \exp[-N^2\pi(x^2 + y^2)]. \tag{2-4}$$

Such "functions", being infinite at the origin and zero elsewhere, have an infinite discontinuity and therefore fail to satisfy existence condition 3. Other important examples are readily found; for example, the functions

$$f(x, y) = 1 \quad \text{and} \quad f(x, y) = \cos(2\pi f_X x) \tag{2-5}$$

both fail to satisfy existence condition 1.

If the majority of functions of interest are to be included within the framework of Fourier analysis, some generalization of the definition (2-1) is required. Fortunately, it is often possible to find a meaningful transform of functions that do not strictly satisfy the existence conditions, provided those functions can be defined as the limit of a sequence of functions that are transformable. By transforming each member function of the defining sequence, a corresponding sequence of transforms is generated, and we call the limit of this new sequence the *generalized Fourier transform* of the original function. Generalized transforms can be manipulated in the same manner as conventional transforms, and the distinction between the two cases can generally be ignored, it being understood that when a function fails to satisfy the existence conditions and yet is said to have a transform, then the generalized transform is actually meant. For a more detailed discussion of this generalization of Fourier analysis the reader is referred to the book by Lighthill [194].

To illustrate the calculation of a generalized transform, consider the Dirac delta function, which has been seen to violate existence condition 3. Note that each member function of the defining sequence (2-4) *does* satisfy the existence requirements and that each, in fact, has a Fourier transform given by (see Table 2.1)

---

[2]For a more detailed discussion of the delta function, including definitions, see Appendix A.

$$\mathcal{F}\{N^2 \exp[-N^2\pi(x^2 + y^2)]\} = \exp\left[-\frac{\pi(f_X^2 + f_Y^2)}{N^2}\right]. \tag{2-6}$$

Accordingly the generalized transform of $\delta(x, y)$ is found to be

$$\mathcal{F}\{\delta(x, y)\} = \lim_{N\to\infty}\left\{\exp\left[-\frac{\pi(f_X^2 + f_Y^2)}{N^2}\right]\right\} = 1. \tag{2-7}$$

Note that the spectrum of a delta function extends uniformly over the entire frequency domain.

For other examples of generalized transforms, see Table 2.1.

2.1.2 The Fourier Transform as a Decomposition

As mentioned previously, when dealing with linear systems it is often useful to decompose a complicated input into a number of more simple inputs, to calculate the response of the system to each of these "elementary" functions, and to superimpose the individual responses to find the total response. Fourier analysis provides the basic means of performing such a decomposition. Consider the familiar inverse transform relationship

$$g(t) = \int_{-\infty}^{\infty} G(f)\exp(j2\pi f t)\,df \tag{2-8}$$

expressing the time function g in terms of its frequency spectrum. We may regard this expression as a decomposition of the function $g(t)$ into a linear combination (in this case an integral) of elementary functions, each with a specific form $\exp(j2\pi f t)$. From this it is clear that the complex number $G(f)$ is simply a weighting factor that must be applied to the elementary function of frequency f in order to synthesize the desired $g(t)$.

In a similar fashion, we may regard the *two-dimensional* Fourier transform as a decomposition of a function $g(x, y)$ into a linear combination of elementary functions of the form $\exp[j2\pi(f_X x + f_Y y)]$. Such functions have a number of interesting properties. Note that for any particular frequency pair (f_X, f_Y) the corresponding elementary function has a phase that is zero or an integer multiple of 2π radians along lines described by the equation

$$y = -\frac{f_X}{f_Y}x + \frac{n}{f_Y}, \tag{2-9}$$

where n is an integer. Thus, as indicated in Fig. 2.1, this elementary function may be regarded as being "directed" in the (x,y) plane at an angle θ (with respect to the x axis) given by

$$\theta = \arctan\left(\frac{f_Y}{f_X}\right). \tag{2-10}$$

In addition, the spatial *period* (i.e. the distance between zero-phase lines) is given by

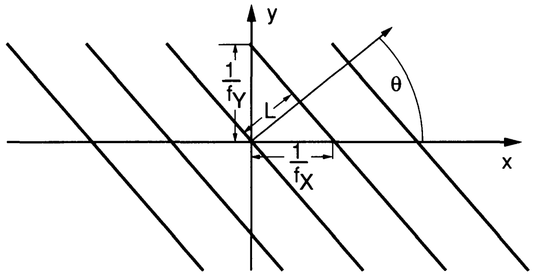

FIGURE 2.1
Lines of zero phase for the function $\exp[j2\pi(f_X x + f_Y y)]$.

$$L = \frac{1}{\sqrt{f_X^2 + f_Y^2}}. \tag{2-11}$$

In conclusion, then, we may again regard the inverse Fourier transform as providing a means for decomposing mathematical functions. The Fourier spectrum G of a function g is simply a description of the weighting factors that must be applied to each elementary function in order to synthesize the desired g. The real advantage obtained from using this decomposition will not be fully evident until our later discussion of invariant linear systems.

2.1.3 Fourier Transform Theorems

The basic definition (2-1) of the Fourier transform leads to a rich mathematical structure associated with the transform operation. We now consider a few of the basic mathematical properties of the transform, properties that will find wide use in later material. These properties are presented as mathematical theorems, followed by brief statements of their physical significance. Since these theorems are direct extensions of the analogous one-dimensional statements, the proofs are deferred to Appendix A.

1. **Linearity theorem.** $\mathcal{F}\{\alpha g + \beta h\} = \alpha\mathcal{F}\{g\} + \beta\mathcal{F}\{h\}$; that is, the transform of a weighted sum of two (or more) functions is simply the identically weighted sum of their individual transforms.
2. **Similarity theorem.** If $\mathcal{F}\{g(x, y)\} = G(f_X, f_Y)$, then
$$\mathcal{F}\{g(ax, by)\} = \frac{1}{|ab|} G\left(\frac{f_X}{a}, \frac{f_Y}{b}\right); \tag{2-12}$$
that is, a "stretch" of the coordinates in the space domain (x, y) results in a contraction of the coordinates in the frequency domain (f_X, f_Y), plus a change in the overall amplitude of the spectrum.
3. **Shift theorem.** If $\mathcal{F}\{g(x, y)\} = G(f_X, f_Y)$, then
$$\mathcal{F}\{g(x - a, y - b)\} = G(f_X, f_Y)\exp[-j2\pi(f_X a + f_Y b)]; \tag{2-13}$$
that is, translation in the space domain introduces a linear phase shift in the frequency domain.

4. **Rayleigh's theorem (Parseval's theorem).** If $\mathcal{F}\{g(x,y)\} = G(f_X, f_Y)$, then

$$\iint_{-\infty}^{\infty} |g(x,y)|^2 \, dx\, dy = \iint_{-\infty}^{\infty} |G(f_X, f_Y)|^2 \, df_X \, df_Y. \tag{2-14}$$

The integral on the left-hand side of this theorem can be interpreted as the energy contained in the waveform $g(x,y)$. This in turn leads us to the idea that the quantity $|G(f_X, f_Y)|^2$ can be interpreted as an energy density in the frequency domain.

5. **Convolution theorem.** If $\mathcal{F}\{g(x,y)\} = G(f_X, f_Y)$ and $\mathcal{F}\{h(x,y)\} = H(f_X, f_Y)$, then

$$\mathcal{F}\left\{\iint_{-\infty}^{\infty} g(\xi,\eta)\, h(x-\xi, y-\eta)\, d\xi\, d\eta\right\} = G(f_X, f_Y) H(f_X, f_Y). \tag{2-15}$$

The convolution of two functions in the space domain (an operation that will be found to arise frequently in the theory of linear systems) is entirely equivalent to the more simple operation of multiplying their individual transforms and inverse transforming.

6. **Autocorrelation theorem.** If $\mathcal{F}\{g(x,y)\} = G(f_X, f_Y)$, then

$$\mathcal{F}\left\{\iint_{-\infty}^{\infty} g(\xi,\eta)\, g^*(\xi - x, \eta - y)\, d\xi\, d\eta\right\} = |G(f_X, f_Y)|^2. \tag{2-16}$$

Similarly,

$$\mathcal{F}\{|g(x,y)|^2\} = \iint_{-\infty}^{\infty} G(\xi,\eta)\, G^*(\xi - f_X, \eta - f_Y)\, d\xi\, d\eta. \tag{2-17}$$

This theorem may be regarded as a special case of the convolution theorem in which we convolve $g(x,y)$ with $g^*(-x,-y)$.

7. **Fourier integral theorem.** At each point of continuity of g,

$$\mathcal{F}\mathcal{F}^{-1}\{g(x,y)\} = \mathcal{F}^{-1}\mathcal{F}\{g(x,y)\} = g(x,y). \tag{2-18}$$

At each point of discontinuity of g, the two successive transforms yield the angular average of the values of g in a small neighborhood of that point. That is, the successive transformation and inverse transformation of a function yields that function again, except at points of discontinuity.

The above transform theorems are of far more than just theoretical interest. They will be used frequently, since they provide the basic tools for the manipulation of Fourier transforms and can save enormous amounts of work in the solution of Fourier analysis problems.

2.1.4 Separable Functions

A function of two independent variables is called *separable* with respect to a specific coordinate system if it can be written as a product of two functions, each of which depends on only one of the independent variables. Thus the function g is separable in rectangular coordinates (x, y) if

$$g(x, y) = g_X(x)\, g_Y(y), \tag{2-19}$$

while it is separable in polar coordinates (r, θ) if

$$g(r, \theta) = g_R(r)\, g_\Theta(\theta). \tag{2-20}$$

Separable functions are often more convenient to deal with than more general functions, for separability often allows complicated two-dimensional manipulations to be reduced to more simple one-dimensional manipulations. For example, a function separable in rectangular coordinates has the particularly simple property that its two-dimensional Fourier transform can be found as a product of one-dimensional Fourier transforms, as evidenced by the following relation:

$$\begin{aligned}
\mathcal{F}\{g(x, y)\} &= \iint_{-\infty}^{\infty} g(x, y) \exp[-j2\pi(f_X x + f_Y y)]\, dx\, dy \\
&= \int_{-\infty}^{\infty} g_X(x) \exp[-j2\pi f_X x]\, dx \int_{-\infty}^{\infty} g_Y(y) \exp[-j2\pi f_Y y]\, dy \\
&= \mathcal{F}_X\{g_X\}\mathcal{F}_Y\{g_Y\}.
\end{aligned} \tag{2-21}$$

Thus the transform of g is itself separable into a product of two factors, one a function of f_X only and the second a function of f_Y only, and the process of two-dimensional transformation simplifies to a succession of more familiar one-dimensional manipulations.

Functions separable in polar coordinates are not so easily handled as those separable in rectangular coordinates, but it is still generally possible to demonstrate that two-dimensional manipulations can be performed by a series of one-dimensional manipulations. For example, the reader is asked to verify in the problems that the Fourier transform of a general function separable in polar coordinates can be expressed as an infinite sum of weighted *Hankel* transforms

$$\mathcal{F}\{g(r, \theta)\} = \sum_{k=-\infty}^{\infty} c_k (-j)^k \exp(jk\phi)\, \mathcal{H}_k\{g_R(r)\} \tag{2-22}$$

where

$$c_k = \frac{1}{2\pi} \int_0^{2\pi} g_\Theta(\theta) \exp(-jk\theta)\, d\theta$$

and $\mathcal{H}_k\{\}$ is the Hankel transform operator of order k, defined by

$$\mathcal{H}_k\{g_R(r)\} = 2\pi \int_0^\infty r\, g_R(r)\, J_k(2\pi r\rho)\, dr. \tag{2-23}$$

Here the function J_k is the kth-order Bessel function of the first kind.

2.1.5 Functions with Circular Symmetry: Fourier-Bessel Transforms

Perhaps the simplest class of functions separable in polar coordinates is composed of those possessing *circular symmetry.* The function g is said to be circularly symmetric if it can be written as a function of r alone, that is,

$$g(r, \theta) = g_R(r). \tag{2-24}$$

Such functions play an important role in the problems of interest here, since most optical systems have precisely this type of symmetry. We accordingly devote special attention to the problem of Fourier transforming a circularly symmetric function.

The Fourier transform of g in a system of rectangular coordinates is, of course, given by

$$G(f_X, f_Y) = \iint_{-\infty}^{\infty} g(x, y) \exp[-j2\pi(f_X x + f_Y y)]\, dx\, dy. \tag{2-25}$$

To fully exploit the circular symmetry of g, we make a transformation to polar coordinates in both the (x, y) and the (f_X, f_Y) planes as follows:

$$\begin{aligned} r &= \sqrt{x^2 + y^2} & x &= r\cos\theta \\ \theta &= \arctan\left(\frac{y}{x}\right) & y &= r\sin\theta \\ \rho &= \sqrt{f_X^2 + f_Y^2} & f_X &= \rho\cos\phi \\ \phi &= \arctan\left(\frac{f_Y}{f_X}\right) & f_Y &= \rho\sin\phi. \end{aligned} \tag{2-26}$$

For the present we write the transform as a function of both radius and angle,[3]

$$\mathcal{F}\{g\} = G_o(\rho, \phi). \tag{2-27}$$

Applying the coordinate transformations (2-26) to Eq. (2-25), the Fourier transform of g can be written

$$G_o(\rho, \phi) = \int_0^{2\pi} d\theta \int_0^\infty dr\, r g_R(r) \exp[-j2\pi r\rho(\cos\theta\cos\phi + \sin\theta\sin\phi)] \tag{2-28}$$

or equivalently,

$$G_o(\rho, \phi) = \int_0^\infty dr\, r g_R(r) \int_0^{2\pi} d\theta \exp[-j2\pi r\rho\cos(\theta - \phi)]. \tag{2-29}$$

[3]Note the subscript in G_o is added simply because the functional form of the expression for the transform in polar coordinates is in general different than the functional form for the same transform in rectangular coordinates.

Finally, we use the Bessel function identity

$$J_0(a) = \frac{1}{2\pi}\int_0^{2\pi} \exp[-ja\cos(\theta - \phi)]\,d\theta, \tag{2-30}$$

where J_0 is a Bessel function of the first kind, zero order, to simplify the expression for the transform. Substituting (2-30) in (2-29), the dependence of the transform on angle ϕ is seen to disappear, leaving G_0 as the following function of radius ρ,

$$G_o(\rho, \phi) = G_o(\rho) = 2\pi \int_0^\infty r g_R(r) J_0(2\pi r\rho)\,dr. \tag{2-31}$$

Thus the Fourier transform of a circularly symmetric function is itself circularly symmetric and can be found by performing the one-dimensional manipulation of (2-31). This particular form of the Fourier transform occurs frequently enough to warrant a special designation; it is accordingly referred to as the *Fourier-Bessel transform,* or alternatively as the *Hankel transform of zero order* (cf. Eq. (2-23)). For brevity, we adopt the former terminology.

By means of arguments identical with those used above, the *inverse* Fourier transform of a circularly symmetric spectrum $G_o(\rho)$ can be expressed as

$$g_R(r) = 2\pi \int_0^\infty \rho G_o(\rho) J_0(2\pi r\rho)\,d\rho. \tag{2-32}$$

Thus for circularly symmetric functions there is no difference between the transform and the inverse-transform operations.

Using the notation $\mathcal{B}\{\}$ to represent the Fourier-Bessel transform operation, it follows directly from the Fourier integral theorem that

$$\mathcal{B}\mathcal{B}^{-1}\{g_R(r)\} = \mathcal{B}^{-1}\mathcal{B}\{g_R(r)\} = \mathcal{B}\mathcal{B}\{g_R(r)\} = g_R(r) \tag{2-33}$$

at each value of r where $g_R(r)$ is continuous. In addition, the *similarity* theorem can be straightforwardly applied (see Prob. 2-6c) to show that

$$\mathcal{B}\{g_R(ar)\} = \frac{1}{a^2} G_o\left(\frac{\rho}{a}\right). \tag{2-34}$$

When using the expression (2-31) for the Fourier-Bessel transform, the reader should remember that it is no more than a special case of the two-dimensional Fourier transform, and therefore any familiar property of the Fourier transform has an entirely equivalent counterpart in the terminology of Fourier-Bessel transforms.

2.1.6 Some Frequently Used Functions and Some Useful Fourier Transform Pairs

A number of mathematical functions will find such extensive use in later material that considerable time and effort can be saved by assigning them special notations of their own. Accordingly, we adopt the following definitions of some frequently used functions:

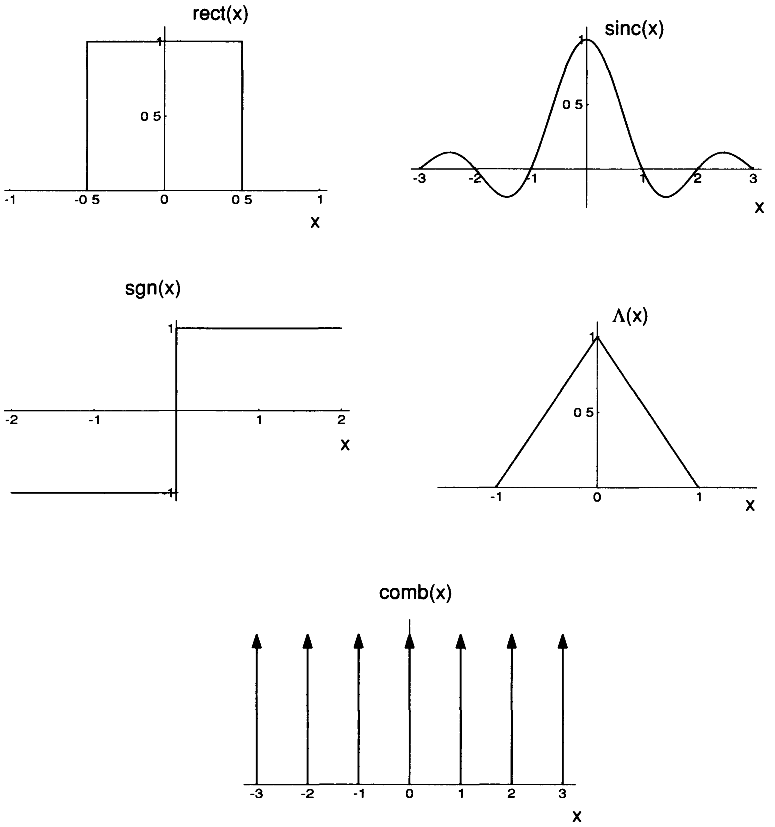

FIGURE 2.2
Special functions.

Rectangle function
$$\mathrm{rect}(x) = \begin{cases} 1 & |x| < \frac{1}{2} \\ \frac{1}{2} & |x| = \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$$

Sinc function
$$\mathrm{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$$

Signum function
$$\mathrm{sgn}(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}$$

Triangle function
$$\Lambda(x) = \begin{cases} 1 - |x| & |x| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Comb function
$$\mathrm{comb}(x) = \sum_{n=-\infty}^{\infty} \delta(x - n)$$

TABLE 2.1
Transform pairs for some functions separable in rectangular coordinates.

Function	Transform
$\exp[-\pi(a^2x^2 + b^2y^2)]$	$\frac{1}{\lvert ab\rvert} \exp\left[-\pi\left(\frac{f_X^2}{a^2} + \frac{f_Y^2}{b^2}\right)\right]$
$\mathrm{rect}(ax)\,\mathrm{rect}(by)$	$\frac{1}{\lvert ab\rvert} \mathrm{sinc}(f_X/a)\,\mathrm{sinc}(f_Y/b)$
$\Lambda(ax)\,\Lambda(by)$	$\frac{1}{\lvert ab\rvert} \mathrm{sinc}^2(f_X/a)\,\mathrm{sinc}^2(f_Y/b)$
$\delta(ax, by)$	$\frac{1}{\lvert ab\rvert}$
$\exp[j\pi(ax + by)]$	$\delta(f_X - a/2, f_Y - b/2)$
$\mathrm{sgn}(ax)\,\mathrm{sgn}(by)$	$\frac{ab}{\lvert ab\rvert} \frac{1}{j\pi f_X} \frac{1}{j\pi f_Y}$
$\mathrm{comb}(ax)\,\mathrm{comb}(by)$	$\frac{1}{\lvert ab\rvert} \mathrm{comb}(f_X/a)\,\mathrm{comb}(f_Y/b)$
$\exp[j\pi(a^2x^2 + b^2y^2)]$	$\frac{j}{\lvert ab\rvert} \exp\left[-j\pi\left(\frac{f_X^2}{a^2} + \frac{f_Y^2}{b^2}\right)\right]$
$\exp[-(a\lvert x\rvert + b\lvert y\rvert)]$	$\frac{1}{\lvert ab\rvert} \frac{2}{1 + (2\pi f_X/a)^2} \frac{2}{1 + (2\pi f_Y/b)^2}$

Circle function
$$\mathrm{circ}(\sqrt{x^2 + y^2}) = \begin{cases} 1 & \sqrt{x^2 + y^2} < 1 \\ \frac{1}{2} & \sqrt{x^2 + y^2} = 1 \\ 0 & \text{otherwise.} \end{cases}$$

The first five of these functions, depicted in Fig. 2.2, are all functions of only one independent variable; however, a variety of separable functions can be formed in two dimensions by means of products of these functions. The circle function is, of course, unique to the case of two-dimensional variables; see Fig. 2.3 for an illustration of its structure.

We conclude our discussion of Fourier analysis by presenting some specific two-dimensional transform pairs. Table 2.1 lists a number of transforms of functions separable in rectangular coordinates. For the convenience of the reader, the functions are presented with arbitrary scaling constants. Since the transforms of such functions can be found directly from products of familiar one-dimensional transforms, the proofs of these relations are left to the reader (cf. Prob. 2-2).

On the other hand, with a few exceptions (e.g. $\exp[-\pi(x^2 + y^2)]$, which is *both* separable in rectangular coordinates *and* circularly symmetric), transforms of most circularly symmetric functions cannot be found simply from a knowledge of one-dimensional transforms. The most frequently encountered function with circular symmetry is:

$$\mathrm{circ}(r) = \begin{cases} 1 & r < 1 \\ \frac{1}{2} & r = 1 \\ 0 & \text{otherwise} \end{cases}.$$

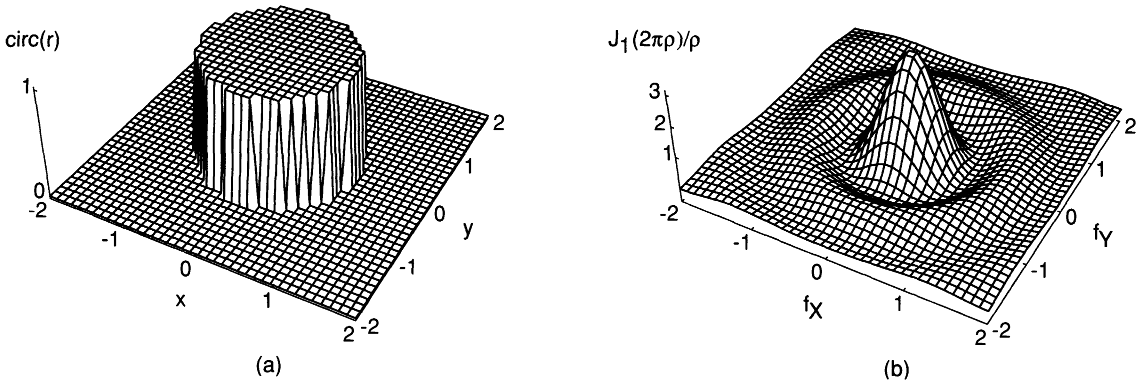

FIGURE 2.3
(a) The circle function and (b) its transform.

Accordingly, some effort is now devoted to finding the transform of this function. Using the Fourier-Bessel transform expression (2-31), the transform of the circle function can be written

$$\mathcal{B}\{\text{circ}(r)\} = 2\pi \int_0^1 r J_0(2\pi r\rho)\, dr.$$

Using a change of variables $r' = 2\pi r\rho$ and the identity

$$\int_0^x \xi J_0(\xi)\, d\xi = x J_1(x),$$

we rewrite the transform as

$$\mathcal{B}\{\text{circ}(r)\} = \frac{1}{2\pi\rho^2} \int_0^{2\pi\rho} r' J_0(r')\, dr' = \frac{J_1(2\pi\rho)}{\rho} \qquad \text{(2-35)}$$

where J_1 is a Bessel function of the first kind, order 1. Figure 2.3 illustrates the circle function and its transform. Note that the transform is circularly symmetric, as expected, and consists of a central lobe and a series of concentric rings of diminishing amplitude. Its value at the origin is π. As a matter of curiosity we note that the zeros of this transform are not equally spaced in radius. A convenient normalized version of this function, with value unity at the origin, is $2\frac{J_1(2\pi\rho)}{2\pi\rho}$. This particular function is called the "besinc" function, or the "jinc" function.

For a number of additional Fourier-Bessel transform pairs, the reader is referred to the problems (see Prob. 2-6).

2.2 LOCAL SPATIAL FREQUENCY AND SPACE-FREQUENCY LOCALIZATION

Each Fourier component of a function is a complex exponential of a unique spatial frequency. As such, every frequency component extends over the entire (x, y) domain. Therefore it is not possible to associate a spatial location with a particular spatial frequency. Nonetheless, we know that in practice certain portions of an image could contain parallel grid lines at a certain fixed spacing, and we are tempted to say that the particular frequency or frequencies represented by these grid lines are localized to certain spatial regions of the image. In this section we introduce the idea of local spatial frequencies and their relation to Fourier components.

For the purpose of this discussion, we consider the general case of complex-valued functions, which we will later see represent the amplitude and phase distributions of monochromatic optical waves. For now, they are just complex functions. Any such function can be represented in the form

$$g(x, y) = a(x, y) \exp[j\phi(x, y)] \qquad \text{(2-36)}$$

where $a(x, y)$ is a real and nonnegative amplitude distribution, while $\phi(x, y)$ is a real phase distribution. For this discussion we assume that the amplitude distribution $a(x, y)$

is a slowly varying function of (x, y), so that we can concentrate on the behavior of the phase function $\phi(x, y)$.

We define the *local spatial frequency* of the function g as a frequency pair (f_{lX}, f_{lY}) given by

$$f_{lX} = \frac{1}{2\pi}\frac{\partial}{\partial x}\phi(x, y) \qquad f_{lY} = \frac{1}{2\pi}\frac{\partial}{\partial y}\phi(x, y). \tag{2-37}$$

In addition, both f_{lX} and f_{lY} are defined to be zero in regions where the function $g(x, y)$ vanishes.

Consider the result of applying these definitions to the particular complex function

$$g(x, y) = \exp[j2\pi(f_X x + f_Y y)]$$

representing a simple linear-phase exponential of frequencies (f_X, f_Y). We obtain

$$f_{lX} = \frac{1}{2\pi}\frac{\partial}{\partial x}\left[2\pi(f_X x + f_Y y)\right] = f_X \qquad f_{lY} = \frac{1}{2\pi}\frac{\partial}{\partial y}\left[2\pi(f_X x + f_Y y)\right] = f_Y.$$

Thus we see that for the case of a single Fourier component, the local frequencies do indeed reduce to the frequencies of that component, and those frequencies are constant over the entire (x, y) plane.

Next consider a space-limited version of a quadratic-phase exponential function,[4] which we call a "finite chirp" function,[5]

$$g(x, y) = \exp[j\pi\beta(x^2 + y^2)]\,\mathrm{rect}\left(\frac{x}{2L_X}\right)\mathrm{rect}\left(\frac{y}{2L_Y}\right). \tag{2-38}$$

Performing the differentiations called for by the definitions of local frequencies, we find that they can be expressed as

$$f_{lX} = \beta x\,\mathrm{rect}\left(\frac{x}{2L_X}\right) \qquad f_{lY} = \beta y\,\mathrm{rect}\left(\frac{y}{2L_Y}\right). \tag{2-39}$$

We see that in this case the local spatial frequencies *do* depend on location in the (x, y) plane; within a rectangle of dimensions $2L_X \times 2L_Y$, f_{lX} varies linearly with the x-coordinate while f_{lY} varies linearly with the y-coordinate. Thus for this function (and for most others) there is a dependence of local spatial frequency on position in the (x, y) plane.[6]

Since the local spatial frequencies are bounded to covering a rectangle of dimensions $2L_X \times 2L_Y$, it would be tempting to conclude that the Fourier spectrum of $g(x, y)$ is also limited to the same rectangular region. In fact this is approximately true, but not exactly so. The Fourier transform of this function is given by the expression

---

[4]For a tutorial discussion of the importance of quadratic-phase functions in various fields of optics, see [229].

[5]The name "chirp function", without the finite length qualifier, will be used for the infinite-length quadratic phase exponential, $\exp[j\pi\beta(x^2 + y^2)]$.

[6]From the definition (2-37) the dimensions of f_{lX} and f_{lY} are both *cycles per meter*, in spite of what might appear to be a contrary implication of Eq. (2-39). The dimensions of β are meters^{-2}.

$$G(f_X, f_Y) = \int_{-L_X}^{L_X}\int_{-L_Y}^{L_Y} e^{j\pi\beta(x^2+y^2)} e^{-j2\pi(f_X x + f_Y y)}\, dx\, dy.$$

This expression is separable in rectangular coordinates, so it suffices to find the one-dimensional spectrum

$$G_X(f_X) = \int_{-L_X}^{L_X} e^{j\pi\beta x^2} e^{j2\pi f_X x}\, dx.$$

Completing the square in the exponent and making a change of variables of integration from x to $t = \sqrt{2\beta}\left(x - \frac{f_X}{\beta}\right)$ yields

$$G_X(f_X) = \frac{1}{\sqrt{2\beta}} e^{-j\pi\frac{f_X^2}{\beta}} \int_{-\sqrt{2\beta}\left(L_X+\frac{f_X}{\beta}\right)}^{\sqrt{2\beta}\left(L_X-\frac{f_X}{\beta}\right)} \exp\left[j\frac{\pi t^2}{2}\right] dt.$$

This integral can be expressed in terms of tabulated functions, the Fresnel integrals, which are defined by

$$C(z) = \int_0^z \cos\left(\frac{\pi t^2}{2}\right) dt \qquad S(z) = \int_0^z \sin\left(\frac{\pi t^2}{2}\right) dt. \tag{2-40}$$

The spectrum G_X can then be expressed as

$$G_X(f_X) = \frac{e^{-j\pi\frac{f_X^2}{\beta}}}{\sqrt{2\beta}} \left\{ C\left[\sqrt{2\beta}\left(L_X - \frac{f_X}{\beta}\right)\right] - C\left[\sqrt{2\beta}\left(-L_X - \frac{f_X}{\beta}\right)\right] \right.$$
$$\left. + jS\left[\sqrt{2\beta}\left(L_X - \frac{f_X}{\beta}\right)\right] - jS\left[\sqrt{2\beta}\left(-L_X - \frac{f_X}{\beta}\right)\right] \right\}.$$

The expression for G_Y is of course identical, except the Y subscript replaces the X subscript. Figure 2.4 shows a plot of $|G_X(f_X)|$ vs. f_X for the particular case of $L_X = 10$ and $\beta = 1$. As can be seen, the spectrum is almost flat over the region $(-L_X, L_X)$ and

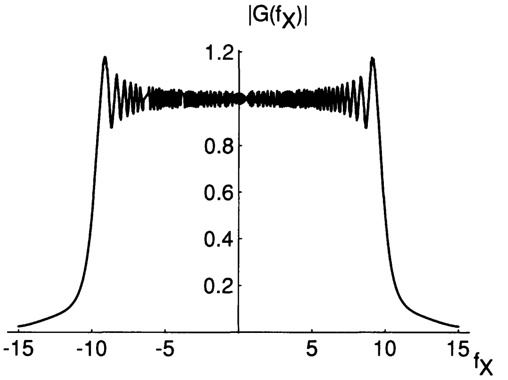

FIGURE 2.4
The spectrum of the finite chirp function, $L_X = 10, \beta = 1$.

almost zero outside that region. We conclude that local spatial frequency has provided a good (but not exact) indication of where the significant values of the Fourier spectrum will occur. However, local spatial frequencies are not the same entity as the frequency components of the Fourier spectrum. Examples can be found for which the local spatial frequency distribution and the Fourier spectrum are not in as good agreement as found in the above example. Good agreement can be expected only when the variations of $\phi(x, y)$ are sufficiently "slow" in the (x, y) plane to allow $\phi(x, y)$ to be well approximated by only three terms of its Taylor series expansion about any point (x, y), i.e. a constant term and two first-partial-derivative terms.

Local spatial frequencies are of special physical significance in optics. When the local spatial frequencies of the complex amplitude of a coherent optical wavefront are found, they correspond to the *ray directions* of the geometrical optics description of that wavefront. However, we are getting ahead of ourselves; we will return to this idea in later chapters and particularly in Appendix B.

2.3 LINEAR SYSTEMS

For the purposes of discussion here, we seek to define the word *system* in a way sufficiently general to include both the familiar case of electrical networks and the less-familiar case of optical imaging systems. Accordingly, a system is defined to be a mapping of a set of input functions into a set of output functions. For the case of electrical networks, the inputs and outputs are real-valued functions (voltages or currents) of a one-dimensional independent variable (time); for the case of imaging systems, the inputs and outputs can be real-valued functions (intensity) or complex-valued functions (field amplitude) of a two-dimensional independent variable (space). As mentioned previously, the question of whether intensity or field amplitude should be considered the relevant quantity will be treated at a later time.

If attention is restricted to deterministic (nonrandom) systems, then a specified input must map to a unique output. It is not necessary, however, that each output correspond to a unique input, for as we shall see, a variety of input functions can produce *no* output. Thus we restrict attention at the outset to systems characterized by many-to-one mappings.

A convenient representation of a system is a mathematical operator, $\mathcal{S}\{\}$, which we imagine to operate on input functions to produce output functions. Thus if the function $g_1(x_1, y_1)$ represents the input to a system, and $g_2(x_2, y_2)$ represents the corresponding output, then by the definition of $\mathcal{S}\{\}$, the two functions are related through

$$g_2(x_2, y_2) = \mathcal{S}\{g_1(x_1, y_1)\}. \tag{2-41}$$

Without specifying more detailed properties of the operator $\mathcal{S}\{\}$, it is difficult to state more specific properties of the general system than those expressed by Eq. (2-41). In the material that follows, we shall be concerned primarily, though not exclusively, with a restricted class of systems that are said to be *linear*. The assumption of linearity will be found to yield simple and physically meaningful representations of such systems; it will also allow useful relations between inputs and outputs to be developed.

2.3.1 Linearity and the Superposition Integral

A system is said to be *linear* if the following superposition property is obeyed for all input functions p and q and all complex constants a and b:

$$\mathcal{S}\{ap(x_1, y_1) + bq(x_1, y_1)\} = a\mathcal{S}\{p(x_1, y_1)\} + b\mathcal{S}\{q(x_1, y_1)\}. \tag{2-42}$$

As mentioned previously, the great advantage afforded by linearity is the ability to express the response of a system to an arbitrary input in terms of the responses to certain "elementary" functions into which the input has been decomposed. It is most important, then, to find a simple and convenient means of decomposing the input. Such a decomposition is offered by the so-called *sifting property* of the δ function (cf. Section 1 of Appendix A), which states that

$$g_1(x_1, y_1) = \iint\limits_{-\infty}^{\infty} g_1(\xi, \eta)\, \delta(x_1 - \xi, y_1 - \eta)\, d\xi\, d\eta. \tag{2-43}$$

This equation may be regarded as expressing g_1 as a linear combination of weighted and displaced δ functions; the elementary functions of the decomposition are, of course, just these δ functions.

To find the response of the system to the input g_1, substitute (2-43) in (2-41):

$$g_2(x_2, y_2) = \mathcal{S}\left\{ \iint\limits_{-\infty}^{\infty} g_1(\xi, \eta)\, \delta(x_1 - \xi, y_1 - \eta)\, d\xi\, d\eta \right\}. \tag{2-44}$$

Now, regarding the number $g_1(\xi, \eta)$ as simply a weighting factor applied to the elementary function $\delta(x_1 - \xi, y_1 - \eta)$, the linearity property (2-42) is invoked to allow $\mathcal{S}\{\}$ to operate on the individual elementary functions; thus the operator $\mathcal{S}\{\}$ is brought within the integral, yielding

$$g_2(x_2, y_2) = \iint\limits_{-\infty}^{\infty} g_1(\xi, \eta)\, \mathcal{S}\{\delta(x_1 - \xi, y_1 - \eta)\}\, d\xi\, d\eta. \tag{2-45}$$

As a final step we let the symbol $h(x_2, y_2; \xi, \eta)$ denote the response of the system at point (x_2, y_2) of the output space to a δ function input at coordinates (ξ, η) of the input space; that is,

$$h(x_2, y_2; \xi, \eta) = \mathcal{S}\{\delta(x_1 - \xi, y_1 - \eta)\}. \tag{2-46}$$

The function h is called the *impulse response* (or in optics, the *point-spread function*) of the system. The system input and output can now be related by the simple equation

$$g_2(x_2, y_2) = \iint\limits_{-\infty}^{\infty} g_1(\xi, \eta)\, h(x_2, y_2; \xi, \eta)\, d\xi\, d\eta. \tag{2-47}$$

This fundamental expression, known as the *superposition integral,* demonstrates the very important fact that a linear system is completely characterized by its responses to unit impulses. To completely specify the output, the responses must in general be known for impulses located at all possible points in the input plane. For the case of a linear *imaging* system, this result has the interesting physical interpretation that the effects of imaging elements (lenses, stops, etc.) can be fully described by specifying the (possibly complex-valued) images of *point sources* located throughout the object field.

2.3.2 Invariant Linear Systems: Transfer Functions

Having examined the input-output relations for a general linear system, we turn now to an important subclass of linear systems, namely *invariant* linear systems. An electrical network is said to be *time-invariant* if its impulse response $h(t;\tau)$ (that is, its response at time t to a unit impulse excitation applied at time τ) depends only on the time difference $(t-\tau)$. Electrical networks composed of fixed resistors, capacitors, and inductors are time-invariant since their characteristics do not change with time.

In a similar fashion, a linear imaging system is *space-invariant* (or equivalently, *isoplanatic*) if its impulse response $h(x_2, y_2; \xi, \eta)$ depends only on the distances $(x_2-\xi)$ and $(y_2-\eta)$ (i.e. the x and y distances between the excitation point and the response point). For such a system we can, of course, write

$$h(x_2, y_2; \xi, \eta) = h(x_2-\xi, y_2-\eta). \qquad (2\text{-}48)$$

Thus an imaging system is space-invariant if the image of a point source object changes only in location, not in functional form, as the point source explores the object field. In practice, imaging systems are seldom isoplanatic over their entire object field, but it is usually possible to divide that field into small regions (*isoplanatic patches*), within which the system is approximately invariant. To completely describe the imaging system, the impulse response appropriate for each isoplanatic patch should be specified; but if the particular portion of the object field of interest is sufficiently small, it often suffices to consider only the isoplanatic patch on the optical axis of the system. Note that for an invariant system the superposition integral (2-47) takes on the particularly simple form

$$g_2(x_2, y_2) = \iint_{-\infty}^{\infty} g_1(\xi, \eta)\, h(x_2-\xi, y_2-\eta)\, d\xi\, d\eta \qquad (2\text{-}49)$$

which we recognize as a two-dimensional *convolution* of the object function with the impulse response of the system. In the future it will be convenient to have a shorthand notation for a convolution relation such as (2-49), and accordingly this equation is written symbolically as

$$g_2 = g_1 \otimes h$$

where a $\otimes$ symbol between any two functions indicates that those functions are to be convolved.

The class of invariant linear systems has associated with it a far more detailed mathematical structure than the more general class of all linear systems, and it is precisely because of this structure that invariant systems are so easily dealt with. The simplicity of invariant systems begins to be evident when we note that the convolution relation (2-49) takes a particularly simple form after Fourier transformation. Specifically, transforming both sides of (2-49) and invoking the convolution theorem, the spectra $G_2(f_X, f_Y)$ and $G_1(f_X, f_Y)$ of the system output and input are seen to be related by the simple equation

$$G_2(f_X, f_Y) = H(f_X, f_Y)\, G_1(f_X, f_Y), \tag{2-50}$$

where H is the Fourier transform of the impulse response

$$H(f_X, f_Y) = \iint_{-\infty}^{\infty} h(\xi, \eta) \exp[-j2\pi(f_X\xi + f_Y\eta)]\, d\xi\, d\eta. \tag{2-51}$$

The function H, called the *transfer function* of the system, indicates the effects of the system in the "frequency domain". Note that the relatively tedious convolution operation of (2-49) required to find the system output is replaced in (2-50) by the often more simple sequence of Fourier transformation, multiplication of transforms, and inverse Fourier transformation.

From another point of view, we may regard the relations (2-50) and (2-51) as indicating that, for a linear invariant system, the input can be decomposed into elementary functions that are more convenient than the δ functions of Eq. (2-43). These alternative elementary functions are, of course, the complex-exponential functions of the Fourier integral representation. By transforming g_1 we are simply decomposing the input into complex-exponential functions of various spatial frequencies (f_X, f_Y). Multiplication of the input spectrum G_1 by the transfer function H then takes into account the effects of the system on each elementary function. Note that these effects are limited to an amplitude change and a phase shift, as evidenced by the fact that we simply multiply the input spectrum by a complex number $H(f_X, f_Y)$ at each (f_X, f_Y). Inverse transformation of the output spectrum G_2 synthesizes the output g_2 by adding up the modified elementary functions.

The mathematical term *eigenfunction* is used for a function that retains its original form (up to a multiplicative complex constant) after passage through a system. Thus we see that the *complex-exponential functions are the eigenfunctions of linear, invariant systems.* The weighting applied by the system to an eigenfunction input is called the *eigenvalue* corresponding to that input. Hence the transfer function describes the continuum of eigenvalues of the system.

Finally, it should be strongly emphasized that the simplifications afforded by transfer-function theory are only applicable for *invariant* linear systems. For applications of Fourier theory in the analysis of time-varying electrical networks, the reader may consult Ref. [158]; applications of Fourier analysis to space-variant imaging systems can be found in Ref. [199].

2.4 TWO-DIMENSIONAL SAMPLING THEORY

It is often convenient, both for data processing and for mathematical analysis purposes, to represent a function $g(x, y)$ by an array of its sampled values taken on a discrete set of points in the (x, y) plane. Intuitively, it is clear that if these samples are taken sufficiently close to each other, the sampled data are an accurate representation of the original function, in the sense that g can be reconstructed with considerable accuracy by simple interpolation. It is a less obvious fact that for a particular class of functions (known as *bandlimited* functions) the reconstruction can be accomplished *exactly,* provided only that the interval between samples is not greater than a certain limit. This result was originally pointed out by Whittaker [298] and was later popularized by Shannon [259] in his studies of information theory.

The sampling theorem applies to the class of bandlimited functions, by which we mean functions with Fourier transforms that are nonzero over only a finite region $\mathcal{R}$ of the frequency space. We consider first a form of this theorem that is directly analogous to the one-dimensional theorem used by Shannon. Later we very briefly indicate improvements of the theorem that can be made in some two-dimensional cases.

2.4.1 The Whittaker-Shannon Sampling Theorem

To derive what is perhaps the simplest version of the sampling theorem, we consider a rectangular lattice of samples of the function g, as defined by

$$g_s(x, y) = \operatorname{comb}\left(\frac{x}{X}\right)\operatorname{comb}\left(\frac{y}{Y}\right) g(x, y). \tag{2-52}$$

The sampled function g_s thus consists of an array of δ functions, spaced at intervals of width X in the x direction and width Y in the y direction, as illustrated in Fig. 2.5. The area under each δ function is proportional to the value of the function g at that particular point in the rectangular sampling lattice. As implied by the convolution theorem, the spectrum G_s of g_s can be found by convolving the transform of $\operatorname{comb}(x/X)\operatorname{comb}(y/Y)$ with the transform of g, or

$$G_s(f_X, f_Y) = \mathcal{F}\left\{\operatorname{comb}\left(\frac{x}{X}\right)\operatorname{comb}\left(\frac{y}{Y}\right)\right\} \otimes G(f_X, f_Y)$$

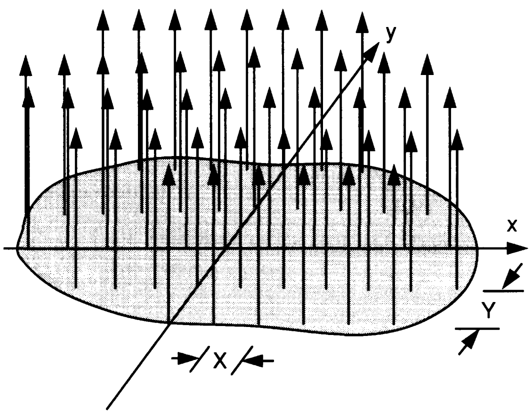

FIGURE 2.5
The sampled function.

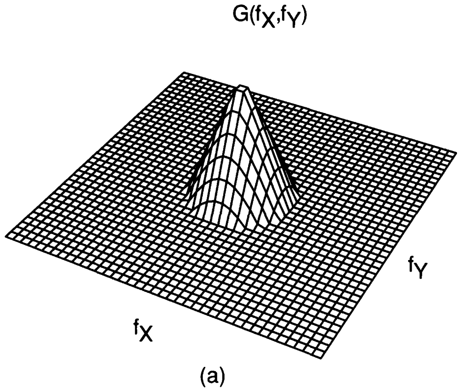

FIGURE 2.6a
Spectrum of the original function.

where the $\otimes$ again indicates that a two-dimensional convolution is to be performed. Now using Table 2.1 we have

$$\mathcal{F}\left\{\operatorname{comb}\left(\frac{x}{X}\right)\operatorname{comb}\left(\frac{y}{Y}\right)\right\} = XY\operatorname{comb}(Xf_X)\operatorname{comb}(Yf_Y)$$

while from the results of Prob. 2-1b,

$$XY\operatorname{comb}(Xf_X)\operatorname{comb}(Yf_Y) = \sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\delta\left(f_X - \frac{n}{X}, f_Y - \frac{m}{Y}\right).$$

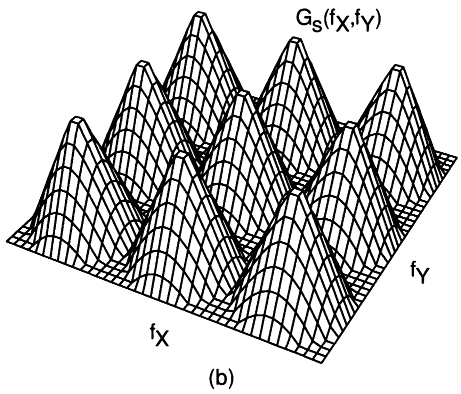

FIGURE 2.6b
Spectrum of the sampled data (only three periods are shown in each direction for this infinitely periodic function).

It follows that

$$G_s(f_X, f_Y) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} G\left(f_X - \frac{n}{X}, f_Y - \frac{m}{Y}\right). \tag{2-53}$$

Evidently the spectrum of g_s can be found simply by erecting the spectrum of g about each point $(n/X, m/Y)$ in the (f_X, f_Y) plane as shown in Fig. 2.6b.

Since the function g is assumed to be bandlimited, its spectrum G is nonzero over only a finite region $\mathcal{R}$ of the frequency space. As implied by Eq. (2-53), the region over which the spectrum of the *sampled* function is nonzero can be found by constructing the region $\mathcal{R}$ about each point $(n/X, m/Y)$ in the frequency plane. Now it becomes clear that if X and Y are sufficiently small (i.e. the samples are sufficiently close together), then the separations $1/X$ and $1/Y$ of the various spectral islands will be great enough to assure that the adjacent regions do not overlap (see Fig. 2.6b). Thus the recovery of the original spectrum G from G_s can be accomplished *exactly* by passing the sampled function g_s through a linear invariant filter that transmits the term $(n = 0, m = 0)$ of Eq. (2-53) without distortion, while perfectly excluding all other terms. Thus, at the output of this filter we find an exact replica of the original data $g(x, y)$.

As stated in the above discussion, to successfully recover the original data it is necessary to take samples close enough together to enable separation of the various spectral regions of G_s. To determine the maximum allowable separation between samples, let $2B_X$ and $2B_Y$ represent the widths in the f_X and f_Y directions, respectively, of the *smallest* rectangle[7] that completely encloses the region $\mathcal{R}$. Since the various terms in the spectrum (2-53) of the sampled data are separated by distances $1/X$ and $1/Y$ in the f_X and f_Y directions, respectively, separation of the spectral regions is assured if

$$X \leq \frac{1}{2B_X} \quad \text{and} \quad Y \leq \frac{1}{2B_Y}. \tag{2-54}$$

The *maximum* spacings of the sampling lattice for exact recovery of the original function are thus $(2B_X)^{-1}$ and $(2B_Y)^{-1}$.

Having determined the maximum allowable distances between samples, it remains to specify the exact transfer function of the filter through which the data should be passed. In many cases there is considerable latitude of choice here, since for many possible shapes of the region $\mathcal{R}$ there are a multitude of transfer functions that will pass the $(n = 0, m = 0)$ term of G_s and exclude all other terms. For our purposes, however, it suffices to note that if the relations (2-54) are satisfied, there is one transfer function that will always yield the desired result regardless of the shape of $\mathcal{R}$, namely

$$H(f_X, f_Y) = \text{rect}\left(\frac{f_X}{2B_X}\right)\text{rect}\left(\frac{f_Y}{2B_Y}\right). \tag{2-55}$$

The exact recovery of G from G_s is seen by noting that the spectrum of the output of such a filter is

[7]For simplicity we assume that this rectangle is centered on the origin. If this is not the case, the arguments can be modified in a straightforward manner to yield a somewhat more efficient sampling theorem.

$$G_s(f_X, f_Y)\,\mathrm{rect}\left(\frac{f_X}{2B_X}\right)\mathrm{rect}\left(\frac{f_Y}{2B_Y}\right) = G(f_X, f_Y).$$

The equivalent identity in the space domain is

$$\left[\mathrm{comb}\left(\frac{x}{X}\right)\mathrm{comb}\left(\frac{y}{Y}\right)g(x, y)\right]\otimes h(x, y) = g(x, y) \qquad (2\text{-}56)$$

where h is the impulse response of the filter,

$$h(x, y) = \mathcal{F}^{-1}\left\{\mathrm{rect}\left(\frac{f_X}{2B_X}\right)\mathrm{rect}\left(\frac{f_Y}{2B_Y}\right)\right\} = 4B_X B_Y\,\mathrm{sinc}(2B_X x)\,\mathrm{sinc}(2B_Y y).$$

Noting that

$$\mathrm{comb}\left(\frac{x}{X}\right)\mathrm{comb}\left(\frac{y}{Y}\right)g(x, y) = XY\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} g(nX, mY)\,\delta(x - nX, y - mY),$$

Eq. (2-56) becomes

$$g(x, y) = 4B_X B_Y XY\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} g(nX, mY)\,\mathrm{sinc}[2B_X(x - nX)]\,\mathrm{sinc}[2B_Y(y - mY)].$$

Finally, when the sampling intervals X and Y are taken to have their maximum allowable values, the identity becomes

$$g(x, y) = \sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} g\left(\frac{n}{2B_X}, \frac{m}{2B_Y}\right)\mathrm{sinc}\left[2B_X\left(x - \frac{n}{2B_X}\right)\right]\mathrm{sinc}\left[2B_Y\left(y - \frac{m}{2B_Y}\right)\right]. \qquad (2\text{-}57)$$

Equation (2-57) represents a fundamental result which we shall refer to as the *Whittaker-Shannon sampling theorem.* It implies that exact recovery of a bandlimited function can be achieved from an appropriately spaced rectangular array of its sampled values; the recovery is accomplished by injecting, at each sampling point, an interpolation function consisting of a product of sinc functions, where each interpolation function is weighted according to the sampled value of g at the corresponding point.

The above result is by no means the only possible sampling theorem. Two rather arbitrary choices were made in the analysis, and alternative choices at these two points will yield alternative sampling theorems. The first arbitrary choice, appearing early in the analysis, was the use of a *rectangular* sampling lattice. The second, somewhat later in the analysis, was the choice of the particular filter transfer function (2-55). Alternative theorems derived by making different choices at these two points are no less valid than Eq. (2-57); in fact, in some cases alternative theorems are more "efficient" in the sense that fewer samples per unit area are required to assure complete recovery. The reader interested in pursuing this extra richness of multidimensional sampling theory is referred to the works of Bracewell [31] and of Peterson and Middleton [230]. A more modern treatment of multidimensional sampling theory is found in Dudgeon and Mersereau [85].

2.4.2 Space-Bandwidth Product

It is possible to show that no function that is bandlimited can be perfectly space-limited as well. That is, if the spectrum G of a function g is nonzero over only a limited region $\mathcal{R}$ in the (f_X, f_Y) plane, then it is not possible for g to be nonzero over only a finite region in the (x, y) plane simultaneously. Nonetheless, in practice most functions do eventually fall to very small values, and therefore from a practical point-of-view it is usually possible to say that g has *significant* values only in some finite region. Exceptions are functions that do not have Fourier transforms in the usual sense, and have to be dealt with in terms of generalized Fourier transforms (e.g. $g(x, y) = 1$, $g(x, y) = \cos[2\pi(f_X x + f_Y y)]$, etc.).

If $g(x, y)$ is bandlimited and indeed has significant value over only a finite region of the (x, y) plane, then it is possible to represent g with good accuracy by a *finite number* of samples. If g is of significant value only in the region $-L_X \leq x < L_X$, $-L_Y \leq y < L_Y$, and if g is sampled, in accord with the sampling theorem, on a rectangular lattice with spacings $(2B_X)^{-1}$, $(2B_Y)^{-1}$ in the x and y directions, respectively, then the total number of significant samples required to represent $g(x, y)$ is seen to be

$$M = 16 L_X L_Y B_X B_Y, \tag{2-58}$$

which we call the *space-bandwidth product* of the function g. The space-bandwidth product can be regarded as the number of degrees of freedom of the given function.

The concept of space-bandwidth product is also useful for many functions that are not strictly bandlimited. If the function is approximately space-limited and approximately bandlimited, then a rectangle (size $2B_X \times 2B_Y$) within which most of the spectrum is contained can be defined in the frequency domain, and a rectangle (size $2L_X \times 2L_Y$) within which most of the function is contained can be defined in the space domain. The space-bandwidth product of the function is then approximately given by Eq. (2-58).

The space-bandwidth product of a function is a measure of its complexity. The ability of an optical system to accurately handle inputs and outputs having large space-bandwidth products is a measure of performance, and is directly related to the quality of the system.

PROBLEMS–CHAPTER 2

2-1. Prove the following properties of δ functions:

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

(b) $\mathrm{comb}(ax)\,\mathrm{comb}(by) = \frac{1}{|ab|}\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\delta\left(x - \frac{n}{a}, y - \frac{m}{b}\right)$.

2-2. Prove the following Fourier transform relations:

(a) $\mathcal{F}\{\mathrm{rect}(x)\,\mathrm{rect}(y)\} = \mathrm{sinc}(f_X)\,\mathrm{sinc}(f_Y)$.

(b) $\mathcal{F}\{\Lambda(x)\Lambda(y)\} = \text{sinc}^2(f_X)\,\text{sinc}^2(f_Y)$.

Prove the following generalized Fourier transform relations:

(c) $\mathcal{F}\{1\} = \delta(f_X, f_Y)$.

(d) $\mathcal{F}\{\text{sgn}(x)\,\text{sgn}(y)\} = \left(\frac{1}{j\pi f_X}\right)\left(\frac{1}{j\pi f_Y}\right)$.

2-3. Prove the following Fourier transform theorems:

(a) $\mathcal{F}\mathcal{F}\{g(x, y)\} = \mathcal{F}^{-1}\mathcal{F}^{-1}\{g(x, y)\} = g(-x, -y)$ at all points of continuity of g.

(b) $\mathcal{F}\{g(x, y)\,h(x, y)\} = \mathcal{F}\{g(x, y)\} \otimes \mathcal{F}\{h(x, y)\}$.

(c) $\mathcal{F}\{\nabla^2 g(x, y)\} = -4\pi^2(f_X^2 + f_Y^2)\mathcal{F}\{g(x, y)\}$ where ∇^2 is the Laplacian operator

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}.$$

2-4. Let the transform operators $\mathcal{F}_A\{\ \}$ and $\mathcal{F}_B\{\ \}$ be defined by

$$\mathcal{F}_A\{g\} = \frac{1}{a}\iint_{-\infty}^{\infty} g(\xi, \eta)\exp\left[-j\frac{2\pi}{a}(f_X\xi + f_Y\eta)\right] d\xi\, d\eta$$

$$\mathcal{F}_B\{g\} = \frac{1}{b}\iint_{-\infty}^{\infty} g(\xi, \eta)\exp\left[-j\frac{2\pi}{b}(x\xi + y\eta)\right] d\xi\, d\eta$$

(a) Find a simple interpretation for

$$\mathcal{F}_B\{\mathcal{F}_A\{g(x, y)\}\}.$$

(b) Interpret the result for $a > b$ and $a < b$.

2-5. The "equivalent area" Δ_{XY} of a function $g(x, y)$ can be defined by

$$\Delta_{XY} = \frac{\iint_{-\infty}^{\infty} g(x, y)\,dx\,dy}{g(0, 0)},$$

while the "equivalent bandwidth" $\Delta_{f_X f_Y}$ of g is defined in terms of its transform G by

$$\Delta_{f_X f_Y} = \frac{\iint_{-\infty}^{\infty} G(f_X, f_Y)\,df_X\,df_Y}{G(0, 0)}.$$

Show that $\quad \Delta_{XY}\,\Delta_{f_X f_Y} = 1.$

2-6. Prove the following Fourier-Bessel transform relations:

(a) If $g_R(r) = \delta(r - r_0)$, then

$$\mathcal{B}\{g_R(r)\} = 2\pi r_0 J_0(2\pi r_0 \rho).$$

(b) If $g_R(r) = 1$ for $a \le r \le 1$ and zero otherwise, then

$$\mathcal{B}\{g_R(r)\} = \frac{J_1(2\pi\rho) - aJ_1(2\pi a\rho)}{\rho}.$$

(c) If $\mathcal{B}\{g_R(r)\} = G(\rho)$, then

$$\mathcal{B}\{g_R(ar)\} = \frac{1}{a^2}G\left(\frac{\rho}{a}\right).$$

(d) $\mathcal{B}\{\exp(-\pi r^2)\} = \exp(-\pi\rho^2)$.

2-7. Let $g(r,\theta)$ be separable in polar coordinates.

(a) Show that if $g(r,\theta) = g_R(r)e^{jm\theta}$, then

$$\mathcal{F}\{g(r,\theta)\} = (-j)^m e^{jm\phi}\mathcal{H}_m\{g_R(r)\}$$

where $\mathcal{H}_m\{\}$ is the Hankel transform of order m,

$$\mathcal{H}_m\{g_R(r)\} = 2\pi\int_0^\infty r g_R(r)\, J_m(2\pi r\rho)\, dr$$

and (ρ,ϕ) are polar coordinates in the frequency space. (Hint: $\exp(ja\sin x) = \sum_{k=-\infty}^{\infty} J_k(a)\exp(jkx)$)

(b) With the help of part (a), prove the general relation presented in Eq. (2-22) for functions separable in polar coordinates.

2-8. Suppose that a sinusoidal input

$$g(x,y) = \cos[2\pi(f_X x + f_Y y)]$$

is applied to a linear system. Under what (sufficient) conditions is the output a real sinusoidal function of the same spatial frequency as the input? Express the amplitude and phase of that output in terms of an appropriate characteristic of the system.

2-9. Show that the zero-order Bessel function $J_0(2\pi\rho_o r)$ is an eigenfunction of any invariant linear system with a circularly symmetric impulse response. What is the corresponding eigenvalue?

2-10. The Fourier transform operator may be regarded as a mapping of functions into their transforms and therefore satisfies the definition of a system as presented in this chapter.

(a) Is this system *linear?*

(b) Can you specify a *transfer function* for this system? If yes, what is it? If no, why not?

2-11. The expression

$$p(x,y) = g(x,y) \otimes \left[\mathrm{comb}\left(\frac{x}{X}\right)\mathrm{comb}\left(\frac{y}{Y}\right)\right]$$

defines a periodic function, with period X in the x direction and period Y in the y direction.

(a) Show that the Fourier transform of p can be written

$$P(f_X, f_Y) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} G\left(\frac{n}{X}, \frac{m}{Y}\right) \delta\left(f_X - \frac{n}{X}, f_Y - \frac{m}{Y}\right)$$

where G is the Fourier transform of g.

(b) Sketch the function $p(x, y)$ when

$$g(x, y) = \operatorname{rect}\left(2\frac{x}{X}\right) \operatorname{rect}\left(2\frac{y}{Y}\right)$$

and find the corresponding Fourier transform $P(f_X, f_Y)$.

2-12. Show that a function with no nonzero spectral components outside a circle of radius B in the frequency plane obeys the following sampling theorem:

$$g(x, y) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} g\left(\frac{n}{2B}, \frac{m}{2B}\right) \frac{\pi}{4} \left\{ 2 \frac{J_1\left[2\pi B \sqrt{\left(x - \frac{n}{2B}\right)^2 + \left(y - \frac{m}{2B}\right)^2}\right]}{2\pi B \sqrt{\left(x - \frac{n}{2B}\right)^2 + \left(y - \frac{m}{2B}\right)^2}} \right\}.$$

2-13. The input to a certain imaging system is an *object* complex field distribution $U_o(x, y)$ of unlimited spatial frequency content, while the output of the system is an *image* field distribution $U_i(x, y)$. The imaging system can be assumed to act as a linear, invariant lowpass filter with a transfer function that is identically zero outside the region $|f_X| \leq B_X$, $|f_Y| \leq B_Y$ in the frequency domain. Show that there exists an "equivalent" object $U_0'(x, y)$ consisting of a rectangular array of point sources that produces exactly the same image U_i as does the true object U_o, and that the field distribution across the equivalent object can be written

$$U_o'(x, y) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \left[\iint_{-\infty}^{\infty} U_o(\xi, \eta) \operatorname{sinc}(n - 2B_X\xi) \operatorname{sinc}(m - 2B_Y\eta) \, d\xi \, d\eta \right] \times \delta\left(x - \frac{n}{2B_X}, y - \frac{m}{2B_Y}\right).$$

2-14. The *Wigner distribution function* of a one-dimensional function $g(x)$ is defined by

$$W(f, x) = \int_{-\infty}^{\infty} g(\xi + x/2) g^*(\xi - x/2) \exp(-j2\pi f \xi) \, d\xi$$

and is a description of the simultaneous (one-dimensional) space and spatial-frequency occupancy of a signal.

(a) Find the Wigner distribution function of the infinite-length chirp function by inserting $g(x) = \exp(j\pi\beta x^2)$ in the definition of $W(f, x)$.

(b) Show that the Wigner distribution function for the one-dimensional finite chirp

$$g(x) = \exp(j\pi\beta x^2) \operatorname{rect}\left(\frac{x}{2L}\right)$$

is given by

$$W(f, x) = (2L - |x|)\,\mathrm{sinc}\big[(2L - |x|)(\beta x - f)\big]$$

for $|x| < 2L$ and zero otherwise.

(c) If you have access to a computer and appropriate software, plot the Wigner distribution function of the finite-length chirp for $L = 10$ and $\beta = 1$, with x ranging from -10 to 10 and f ranging from -10 to 10. To make the nature of this function clearer, also plot $W(0, x)$ for $|x| \le 1$.