2 / ANALYSIS OF TWO-DIMENSIONAL LINEAR 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 of the component stimuli would produce individually. Such phenomena are called *linear*, and the property that 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 linearity of 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 lacks 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 in such applications and therefore that he 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 [Ref. 2-1] and Bracewell [Ref. 2-2]. 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.

Definition and existence conditions

The *Fourier transform* (alternatively the *Fourier spectrum* or *frequency spectrum*) of a complex function[1] $\mathbf{g}$ of two independent variables, x and y, will be represented here by $\mathcal{F}\{\mathbf{g}\}$ and is defined by[2]

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

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 $\mathbf{G}(f_X,f_Y)$ will be represented by $\mathcal{F}^{-1}\{\mathbf{G}\}$ and is defined as

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

Note that as mathematical operations the transform and inverse trans-

[1] Boldface sans serif type will be used throughout to indicate that a function is complex-valued.

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

form are very similar, differing only in the sign of the exponent appearing in the integrand.

Before discussing the properties of the Fourier transform and its inverse, we must first decide when the definitions (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. g must be absolutely integrable over the infinite xy 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 [Ref. 2-2] 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 δ function,[1] defined by

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

Similarly, an idealized point source of light is often represented by the two-dimensional equivalent,

$$\delta(x,y) = \lim_{N\to\infty} N^2 \exp[-N^2\pi(x^2 + y^2)] \tag{2-3}$$

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

$$f(x,y) = 1 \qquad \text{and} \qquad f(x,y) = \cos(2\pi f_X x)$$

both fail to satisfy existence condition 1.

Evidently, if the majority of functions of interest are to be included within the framework of Fourier analysis, some generalization of the

[1] For a more detailed discussion of the δ function, including alternative definitions, see Sec. A in the appendix.

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 [Ref. 2-3].

To illustrate the calculation of a generalized transform, consider the Dirac δ function, which has been seen to violate existence condition 3. Note that each member function of the defining sequence (2-3) *does* satisfy the existence requirements and that each, in fact, has a Fourier transform given by (see Table 2-1).

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

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

$$\mathfrak{F}\{\delta(x,y)\} = \lim_{N\to\infty}\left\{\exp\left[-\frac{\pi(f_X{}^2 + f_Y{}^2)}{N^2}\right]\right\} = 1 \qquad (2\text{-}4)$$

Evidently the spectrum of a δ function extends uniformly over the entire frequency domain.

For other examples of generalized transforms see Table 2-1.

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 a basic means of performing one such decomposition. Consider the familiar inverse-transform relationship

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

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

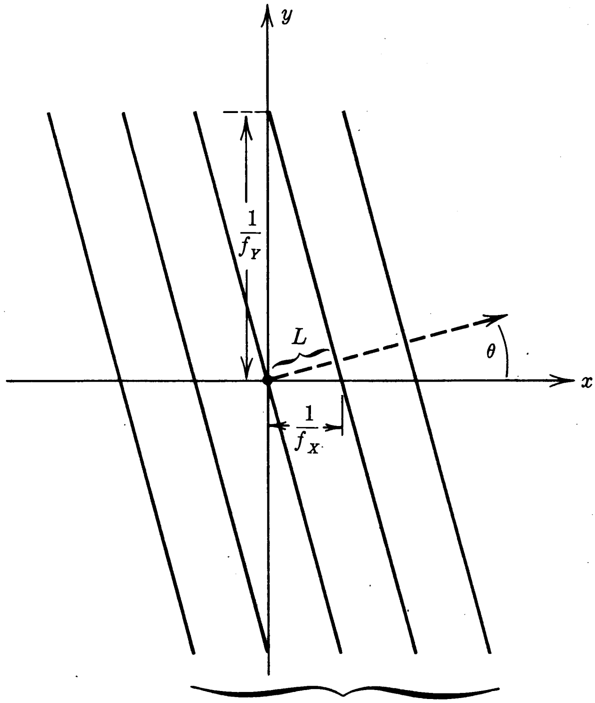

Figure 2-1 Lines of zero phase for the function $\exp [j2\pi(f_X x + f_Y y)]$.

linear combination (i.e., an integral) of elementary functions, each with the specific form $\exp (j2\pi ft)$. Evidently the complex number $\mathbf{G}(f)$ 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 $\mathbf{g}(x,y)$ 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 zero phase along lines described by

$$y = -\frac{f_X}{f_Y}x + \frac{n}{f_Y} \qquad (n \text{ an integer})$$

Thus, as indicated in Fig. 2-1, this elementary function may be regarded as being "directed" in the xy plane at an angle θ (with respect to the

x axis) given by

$$\theta = \tan^{-1} \frac{f_Y}{f_X} \tag{2-5}$$

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

$$L = \frac{1}{\sqrt{f_X^2 + f_Y^2}} \tag{2-6}$$

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

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 a brief statement of their physical significance. Since these theorems are direct extensions of the analogous one-dimensional statements, the proofs are deferred to the appendix.

1. ***Linearity theorem.*** $\mathfrak{F}\{\alpha \mathbf{g} + \beta \mathbf{h}\} = \alpha \mathfrak{F}\{\mathbf{g}\} + \beta \mathfrak{F}\{\mathbf{h}\}$; that is, the transform of a sum of two functions is simply the sum of their individual transforms.
2. ***Similarity theorem.*** If $\mathfrak{F}\{\mathbf{g}(x,y)\} = \mathbf{G}(f_X,f_Y)$, then

   $$\mathfrak{F}\{\mathbf{g}(ax,by)\} = \frac{1}{|ab|} \mathbf{G}\left(\frac{f_X}{a}, \frac{f_Y}{b}\right)$$

   that is, a "stretching" 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 $\mathfrak{F}\{\mathbf{g}(x,y)\} = \mathbf{G}(f_X,f_Y)$, then

   $$\mathfrak{F}\{\mathbf{g}(x - a, y - b)\} = \mathbf{G}(f_X,f_Y) \exp[-j2\pi(f_X a + f_Y b)]$$

   that is, translation of a function in the space domain introduces a linear phase shift in the frequency domain.

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

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

This theorem is generally interpretable as a statement of conservation of energy.

5. ***Convolution theorem.*** If $\mathfrak{F}\{\mathbf{g}(x,y)\} = \mathbf{G}(f_X,f_Y)$ and

$$\mathfrak{F}\{\mathbf{h}(x,y)\} = \mathbf{H}(f_X,f_Y)$$

then

$$\mathfrak{F}\left\{\iint_{-\infty}^{\infty} \mathbf{g}(\xi,\eta)\mathbf{h}(x-\xi,\, y-\eta)\, d\xi\, d\eta\right\} = \mathbf{G}(f_X,f_Y)\mathbf{H}(f_X,f_Y)$$

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.

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

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

Similarly,

$$\mathfrak{F}\{|\mathbf{g}(\xi,\eta)|^2\} = \iint_{-\infty}^{\infty} \mathbf{G}(\xi,\eta)\mathbf{G}^*(\xi-f_X,\, \eta-f_Y)\, d\xi\, d\eta$$

This theorem may be regarded as a special case of the convolution theorem.

7. ***Fourier integral theorem.*** At each point of continuity of $\mathbf{g}$

$$\mathfrak{F}\mathfrak{F}^{-1}\{\mathbf{g}(x,y)\} = \mathfrak{F}^{-1}\mathfrak{F}\{\mathbf{g}(x,y)\} = \mathbf{g}(x,y)$$

At each point of discontinuity of $\mathbf{g}$, the two successive transforms yield the angular average of the value of $\mathbf{g}$ 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 basic tools for the manipulation of Fourier transforms and can save enormous amounts of work in the solution of Fourier analysis problems.

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 independent variable. Thus a function $\mathbf{g}$ is separable in the rectangular coordinates (x,y) if

$$\mathbf{g}(x,y) = \mathbf{g}_X(x)\mathbf{g}_Y(y) \tag{2-7}$$

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

$$\mathbf{g}(r,\theta) = \mathbf{g}_R(r)\mathbf{g}_\theta(\theta) \tag{2-8}$$

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:

$$\begin{aligned}\mathfrak{F}\{\mathbf{g}(x,y)\} &= \iint_{-\infty}^{\infty} \mathbf{g}(x,y) \exp\left[-j2\pi(f_X x + f_Y y)\right] dx\, dy \\ &= \int_{-\infty}^{\infty} \mathbf{g}_X(x) \exp\left[-j2\pi f_X x\right] dx \int_{-\infty}^{\infty} \mathbf{g}_Y(y) \exp\left[-j2\pi f_Y y\right] dy \\ &= \mathfrak{F}_X\{\mathbf{g}_X\}\mathfrak{F}_Y\{\mathbf{g}_Y\}\end{aligned} \tag{2-9}$$

Thus the transform of $\mathbf{g}$ 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 means of a series of one-dimensional manipulations. For example, the process of Fourier transforming a function separable in polar coordinates is considered in the problems (see Prob. 2-7), where the reader is asked to verify that the two-dimensional spectrum can be found by performing a series of one-dimensional operations called *Hankel transforms*.

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 $\mathbf{g}$ is said to be circularly symmetric if it can be written as a function of radius r alone, that is,

$$\mathbf{g}(r,\theta) = \mathbf{g}_R(r) \tag{2-10}$$

Such functions play a particularly 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 $\mathbf{g}$ in a system of rectangular coordinates is, of course, given by

$$\mathbf{G}(f_X,f_Y) = \iint\limits_{-\infty}^{\infty} \mathbf{g}(x,y) \exp\,[-j2\pi(f_Xx + f_Yy)]\, dx\, dy \tag{2-11}$$

To fully exploit the circular symmetry of $\mathbf{g}$, we make a transformation to polar coordinates in both the xy and f_Xf_Y planes as follows:

$$\begin{aligned} r &= \sqrt{x^2 + y^2} & x &= r\cos\theta \\ \theta &= \tan^{-1}\left(\frac{y}{x}\right) & y &= r\sin\theta \\ \rho &= \sqrt{f_X{}^2 + f_Y{}^2} & f_X &= \rho\cos\phi \\ \phi &= \tan^{-1}\left(\frac{f_Y}{f_X}\right) & f_Y &= \rho\sin\phi \end{aligned} \tag{2-12}$$

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

$$\mathcal{F}\{\mathbf{g}\} = \mathbf{G}_0(\rho,\phi)$$

Applying the coordinate transformations (2-12) to Eq. (2-11), the Fourier transform of $\mathbf{g}$ can be written

$$\mathbf{G}_0(\rho,\phi) = \int_0^{2\pi} d\theta \int_0^{\infty} dr \cdot r\mathbf{g}_R(r) \exp\,[-j2\pi r\rho(\cos\theta\cos\phi + \sin\theta\sin\phi)]$$

or equivalently,

$$\mathbf{G}_0(\rho,\phi) = \int_0^{\infty} dr \cdot r\mathbf{g}_R(r) \int_0^{2\pi} d\theta \exp\,[-j2\pi r\rho\cos(\theta - \phi)] \tag{2-13}$$

Finally, we use the Bessel function identity

$$J_0(a) = \frac{1}{2\pi}\int_0^{2\pi} \exp\,[-ja\cos(\theta - \phi)]\, d\theta \tag{2-14}$$

where J_0 is a Bessel function of the first kind, zero order, to simplify the expression for the transform. Substituting (2-14) in (2-13), the dependence of the transform on angle ϕ is seen to disappear, leaving $\mathbf{G}_0$ as the following function of radius ρ,

$$\mathbf{G}_0(\rho) = 2\pi \int_0^\infty r\mathbf{g}_R(r)J_0(2\pi r\rho)\,dr \tag{2-15}$$

Thus the Fourier transform of a circularly symmetric function is itself circularly symmetric and can be found by performing the one-dimensional manipulation (2-15). This particular form of the Fourier transform occurs frequently enough to warrant a special designation; the expression (2-15) is accordingly referred to as the *Fourier-Bessel transform,* or alternatively, as the *Hankel transform of zero order.* For brevity we adopt the former terminology.

By means of arguments identical with those used above, the *inverse* Fourier transform of a circularly symmetric function $\mathbf{G}_0(\rho)$ can be expressed as

$$\mathbf{g}_R(r) = 2\pi \int_0^\infty \rho\mathbf{G}_0(\rho)J_0(2\pi r\rho)\,d\rho$$

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

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

$$\mathcal{B}\mathcal{B}^{-1}\{\mathbf{g}_R(r)\} = \mathcal{B}\mathcal{B}\{\mathbf{g}_R(r)\} = \mathbf{g}_R(r)$$

at each value of r where $\mathbf{g}_R(r)$ is continuous. In addition, the *similarity* theorem can be straightforwardly applied (see Prob. 2-4) to show that

$$\mathcal{B}\{\mathbf{g}_R(ar)\} = \frac{1}{a^2}\mathbf{G}_0\left(\frac{\rho}{a}\right)$$

When using the expression (2-15) 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.

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:

Rectangle function

$$\operatorname{rect}(x) = \begin{cases} 1 & |x| \le \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$$

Sinc function

$$\operatorname{sinc}(x) = \frac{\sin \pi x}{\pi x}$$

Sign function

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

Triangle function

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

Comb function

$$\operatorname{comb}(x) = \sum_{n=-\infty}^{\infty} \delta(x - n)$$

Circle function

$$\operatorname{circ}(\sqrt{x^2 + y^2}) = \begin{cases} 1 & \sqrt{x^2 + y^2} \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

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 on a two-dimensional space can be formed by means of products of these functions. The circle function is, of course, unique to the case of two independent 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. Since the transforms of such functions can be found directly from products of

Table 2-1 Transform pairs for some functions separable in rectangular coordinates

Function	Transform
$\exp[-\pi(x^2 + y^2)]$	$\exp[-\pi(f_X^2 + f_Y^2)]$
$\operatorname{rect}(x)\operatorname{rect}(y)$	$\operatorname{sinc}(f_X)\operatorname{sinc}(f_Y)$
$\Lambda(x)\Lambda(y)$	$\operatorname{sinc}^2(f_X)\operatorname{sinc}^2(f_Y)$
$\delta(x,y)$	1
$\exp[j\pi(x + y)]$	$\delta(f_X - \frac{1}{2}, f_Y - \frac{1}{2})$
$\operatorname{sgn}(x)\operatorname{sgn}(y)$	$\frac{1}{j\pi f_X}\frac{1}{j\pi f_Y}$
$\operatorname{comb}(x)\operatorname{comb}(y)$	$\operatorname{comb}(f_X)\operatorname{comb}(f_Y)$

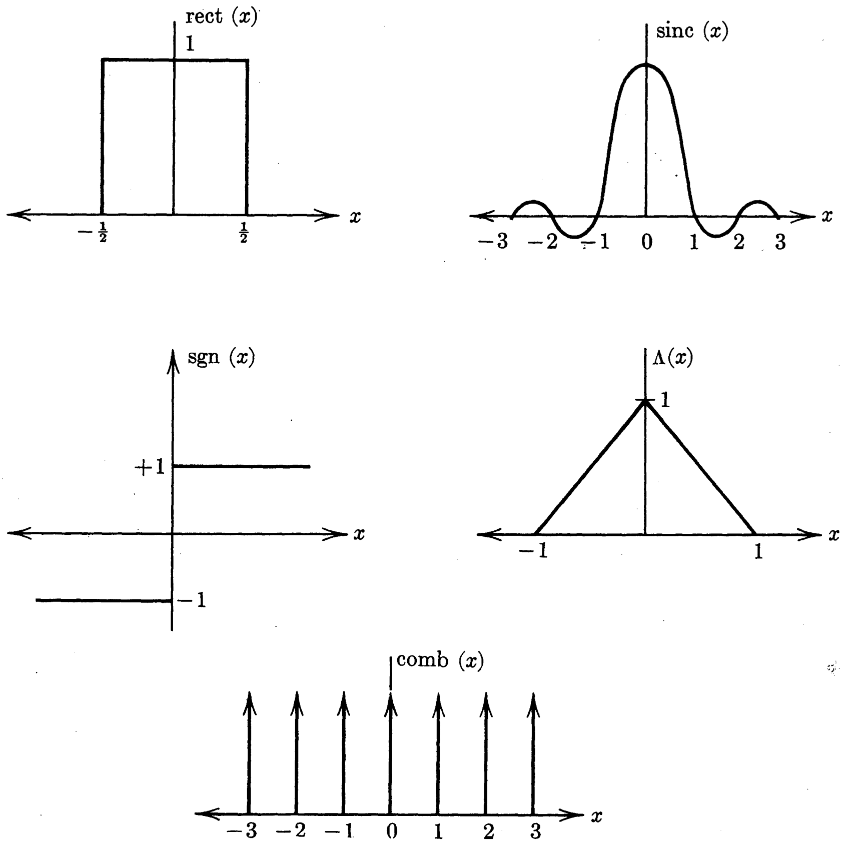

Figure 2-2 Special functions.

familiar one-dimensional transforms, the proofs of these relations are left to the reader (see Prob. 2-2).

On the other hand, 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:

$$\mathrm{circ}(r) = \begin{cases} 1 & r \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Accordingly, some effort is now devoted to finding the transform of this function. Using the Fourier-Bessel transform expression (2-15), the transform of the circle function can be written

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

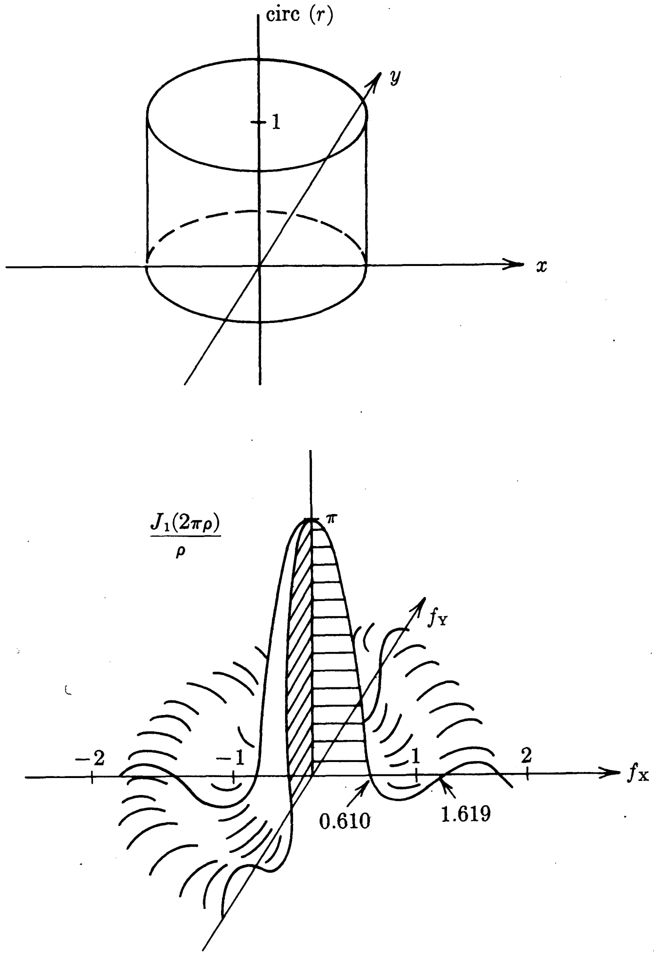

Figure 2-3 The circle function and its transform.

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

$$\int_0^x \xi J_0(\xi)\, d\xi = xJ_1(x)$$

we rewrite the transform as

$$\mathcal{B}\{\mathrm{circ}(r)\} = \frac{1}{2\pi\rho^2}\int_0^{2\pi\rho} r' J_0(r')\, dr' = \frac{J_1(2\pi\rho)}{\rho} \tag{2-16}$$

where J_1 is a Bessel function of the first kind, order one. Figure 2-3 illustrates the circle function and its transform. Note that the transform is circularly symmetric, as expected, and consists of a central spike and a series of concentric rings of diminishing amplitude. As a matter of curiosity we note that the zeros of this transform are not equally spaced in radius. For a number of additional Fourier-Bessel transform pairs, the reader is referred to the problems (see Prob. 2-4).

2-2 LINEAR SYSTEMS

For the purposes of our discussions 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 devices. 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 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 the intensity or the field amplitude should be considered the system variable will be treated at a later time.

If attention is restricted to deterministic (nonrandom) systems, then a specified input must map into 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-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 $\mathbf{g}_1(x_1,y_1)$ represents the input to a system, and $\mathbf{g}_2(x_2,y_2)$ represents the corresponding output, then by the definition of $\mathcal{S}\{\ \}$, the two functions are related through

$$\mathbf{g}_2(x_2,y_2) = \mathcal{S}\{\mathbf{g}_1(x_1,y_1)\} \tag{2-17}$$

Without specifying more detailed properties of the operator $\mathcal{S}\{\ \}$, it is difficult to state more specific properties of a general system than those expressed by Eq. (2-17). 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.

Linearity and the superposition integral

A system is said to be *linear* if the following superposition property is obeyed for all input functions $\mathbf{t}$ and $\mathbf{s}$ and all complex constants $\mathbf{a}$ and $\mathbf{b}$:

$$\mathcal{S}\{\mathbf{a}\mathbf{s}(x_1,y_1) + \mathbf{b}\mathbf{t}(x_1,y_1)\} = \mathbf{a}\mathcal{S}\{\mathbf{s}(x_1,y_1)\} + \mathbf{b}\mathcal{S}\{\mathbf{t}(x_1,y_1)\} \tag{2-18}$$

As mentioned previously, the great advantage afforded by linearity is the ability to express the response of the 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. Sec. A in the appendix), which states that

$$\mathbf{g}_1(x_1,y_1) = \iint\limits_{-\infty}^{\infty} \mathbf{g}_1(\xi,\eta)\delta(x_1 - \xi, y_1 - \eta)\, d\xi\, d\eta \tag{2-19}$$

This equation may be regarded as expressing $\mathbf{g}_1$ 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 $\mathbf{g}_1$, substitute (2-19) in (2-17):

$$\mathbf{g}_2(x_2,y_2) = \mathcal{S}\left\{\iint\limits_{-\infty}^{\infty} \mathbf{g}_1(\xi,\eta)\delta(x_1 - \xi, y_1 - \eta)\, d\xi\, d\eta\right\}$$

Now, regarding the number $\mathbf{g}_1(\xi,\eta)$ as simply a weighting factor applied to the elementary function $\delta(x_1 - \xi, y_1 - \eta)$, the linearity property (2-18) is invoked to allow $\mathcal{S}\{\cdot\}$ to operate on the individual elementary functions; thus the operator $\mathcal{S}\{\ \}$ is brought within the integral, yielding

$$\mathbf{g}_2(x_2,y_2) = \iint\limits_{-\infty}^{\infty} \mathbf{g}_1(\xi,\eta)\mathcal{S}\{\delta(x_1 - \xi, y_1 - \eta)\}\, d\xi\, d\eta$$

As a final step we let the symbol $\mathbf{h}(x_2,y_2;\xi,\eta)$ 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,

$$\mathbf{h}(x_2,y_2;\xi,\eta) = \mathcal{S}\{\delta(x_1 - \xi, y_1 - \eta)\} \tag{2-20}$$

The function $\mathbf{h}$ is called the *impulse response* of the system. The system input and output can now be related by the simple equation

$$\mathbf{g}_2(x_2,y_2) = \iint\limits_{-\infty}^{\infty} \mathbf{g}_1(\xi,\eta)\mathbf{h}(x_2,y_2;\xi,\eta)\, d\xi\, d\eta \tag{2-21}$$

This fundamental expression, known as the *superposition integral,* demonstrates the very important fact that a linear system is completely characterized by its response 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.

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 said to be *space-invariant* (or equivalently, *isoplanatic*) if its impulse response $\mathsf{h}(x_2,y_2;\xi,\eta)$ depends only on the distances $(x_2 - \xi)$ and $(y_2 - \eta)$. For such a system we can, of course, write

$$\mathsf{h}(x_2,y_2;\xi,\eta) = \mathsf{h}(x_2 - \xi,\, y_2 - \eta) \tag{2-22}$$

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 object field, but it is usually possible to divide the object field into small regions (*isoplanatic patches*) within which the system is approximately invariant. To completely describe the imaging system, the impulse response appropriate to 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 axis of the system. Note that for an invariant system the superposition integral (2-21) takes on the particularly simple form

$$\mathsf{g}_2(x_2,y_2) = \iint\limits_{-\infty}^{\infty} \mathsf{g}_1(\xi,\eta)\mathsf{h}(x_2 - \xi,\, y_2 - \eta)\, d\xi\, d\eta \tag{2-23}$$

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 (2-23), and accordingly this equation is rewritten

$$\mathbf{g}_2 = \mathbf{g}_1 * \mathbf{h}$$

where an asterisk between any two functions is a convenient symbol indicating 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 (2-23) takes on a particularly simple form after Fourier transformation. Specifically, transforming both sides of (2-23) and invoking the convolution theorem, the spectra $\mathbf{G}_2(f_X,f_Y)$ and $\mathbf{G}_1(f_X,f_Y)$ of the system output and input are seen to be related by the simple equation

$$\mathbf{G}_2(f_X,f_Y) = \mathbf{H}(f_X,f_Y)\mathbf{G}_1(f_X,f_Y) \tag{2-24}$$

where $\mathbf{H}$ is the Fourier transform of the impulse response

$$\mathbf{H}(f_X,f_Y) = \iint_{-\infty}^{\infty} \mathbf{h}(\xi,\eta) \exp\left[-j2\pi(f_X\xi + f_Y\eta)\right] d\xi\, d\eta \tag{2-25}$$

The function $\mathbf{H}$, called the *transfer function* of the system, indicates the effects of the system in the "frequency domain." Note that the relatively tedious convolution operation (2-23) required to find the system output is replaced in (2-24) 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 (2-24) and (2-25) as indicating that, for linear invariant systems, the input can be decomposed into elementary functions that are more convenient than the δ functions of Eq. (2-19). These alternative elementary functions are, of course, the complex-exponential functions. By transforming $\mathbf{g}_1$ we are simply decomposing the input into complex-exponential functions of various spatial frequencies (f_X,f_Y). Multiplication of the input spectrum $\mathbf{G}_1$ by the transfer function $\mathbf{H}$ 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 $\mathbf{H}(f_X,f_Y)$ at each (f_X,f_Y). Inverse transformation of the output spectrum $\mathbf{G}_2$ simply synthesizes the output $\mathbf{g}_2$ by adding up all the modified elementary functions.

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. 2-4; applications of Fourier analysis to space-variant imaging systems can be found in Ref. 2-5.

2-3 TWO-DIMENSIONAL SAMPLING THEORY

It is often convenient, both for data processing and for mathematical analysis purposes, to represent a function $\mathbf{g}(x,y)$ by an array of its sampled values taken on a discrete set of points in the xy 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 $\mathbf{g}$ 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*, providing only that the interval between samples is not greater than a certain limit. This result was originally pointed out by Whittaker [Ref. 2-6] and was later popularized by Shannon [Ref. 2-7] 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 this theorem that can be made in some two-dimensional cases.

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 $\mathbf{g}$, as defined by

$$\mathbf{g}_s(x,y) = \text{comb}\left(\frac{x}{X}\right)\text{comb}\left(\frac{y}{Y}\right)\mathbf{g}(x,y) \tag{2-26}$$

The sampled function $\mathbf{g}_s$ 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. 2-4. The area under each δ function is proportional to the value of the function $\mathbf{g}$ at that particular point in the rectangular sampling lattice. As implied by the convolution theorem, the spectrum $\mathbf{G}_s$ of $\mathbf{g}_s$ can be found by convolving the transform of $\text{comb}(x/X)\ \text{comb}(y/Y)$

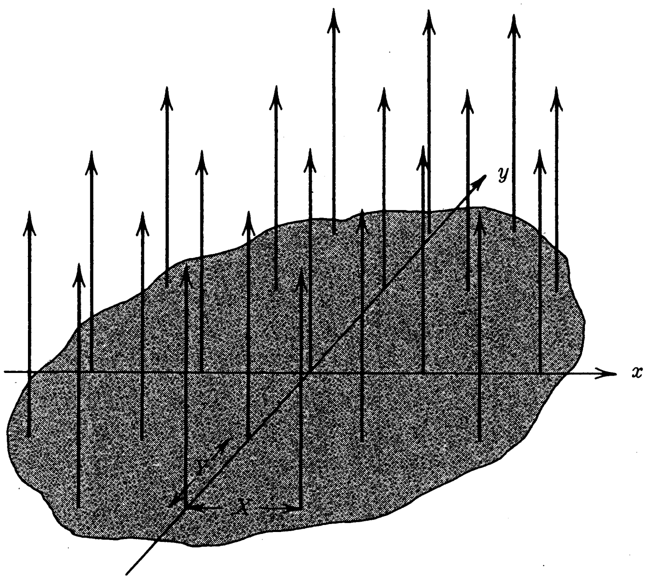

Figure 2-4 The sampled function.

with the transform of $\mathbf{g}$, or

$$\mathbf{G}_s(f_X,f_Y) = \mathfrak{F}\left\{\text{comb}\left(\frac{x}{X}\right)\text{comb}\left(\frac{y}{Y}\right)\right\} * \mathbf{G}(f_X,f_Y)$$

where the asterisk again indicates that a convolution is to be performed. Now using Table 2-1 and the similarity theorem, we have

$$\mathfrak{F}\left\{\text{comb}\left(\frac{x}{X}\right)\text{comb}\left(\frac{y}{Y}\right)\right\} = XY\,\text{comb}(Xf_X)\,\text{comb}(Yf_Y)$$

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

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

It follows that the spectrum of the sampled function is given by

$$\mathbf{G}_s(f_X,f_Y) = \sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\mathbf{G}\left(f_X - \frac{n}{X}, f_Y - \frac{m}{Y}\right) \qquad (2\text{-}27)$$

Evidently the spectrum of $\mathbf{g}_s$ can be found simply by erecting the spectrum of $\mathbf{g}$ about each point $(n/X,m/Y)$ in the f_Xf_Y plane as shown in Fig. 2-5.

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. (2-27), 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 regions will be great enough to assure that adjacent regions do not overlap (see Fig. 2-5). Thus recovery of the original spectrum **G** from $\mathbf{G}_s$ can be accomplished *exactly* by passing the sampled function $\mathbf{g}_s$ through a linear filter that transmits the term $(n = 0,\ m = 0)$ of Eq. (2-27) without distortion, while perfectly excluding all other terms. Thus, at the output of this filter we find an exact replica of the original data $\mathbf{g}(x,y)$.

As stated in the above discussion, to successfully recover the original data it is necessary to take the samples close enough together to enable separation of the various spectral regions of $\mathbf{G}_s$. 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[1] that completely encloses the region $\mathcal{R}$. Since the various terms in the spectrum (2-27) 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

$$X \leq \frac{1}{2B_X} \quad \text{and} \quad Y \leq \frac{1}{2B_Y} \tag{2-28}$$

[1] 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.

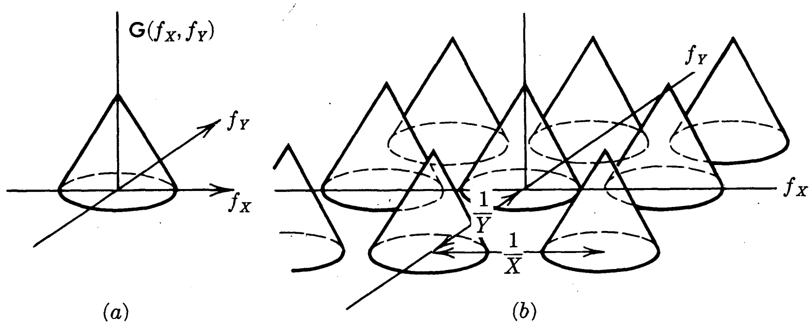

Figure 2-5 Spectra of (a) the original function and (b) the sampled data.

The *maximum* spacings of the sample 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 sampled 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 $\mathbf{G}_s$ and exclude all other terms. For our purposes, however, it suffices to note that if the relations (2-28) are satisfied, there is one transfer function that will always yield the desired result regardless of the specific shape of $\mathcal{R}$, namely,

$$\mathbf{H}(f_X,f_Y) = \text{rect}\left(\frac{f_X}{2B_X}\right) \text{rect}\left(\frac{f_Y}{2B_Y}\right) \tag{2-29}$$

The exact recovery of $\mathbf{G}$ from $\mathbf{G}_s$ is seen by noting that the spectrum of the output of such a filter is

$$\mathbf{G}_s(f_X,f_Y) \,\text{rect}\left(\frac{f_X}{2B_X}\right) \text{rect}\left(\frac{f_Y}{2B_Y}\right) \equiv \mathbf{G}(f_X,f_Y)$$

The equivalent identity in the space domain is

$$\left[\text{comb}\left(\frac{x}{X}\right) \text{comb}\left(\frac{y}{Y}\right) \mathbf{g}(x,y)\right] * \mathbf{h}(x,y) = \mathbf{g}(x,y) \tag{2-30}$$

where $\mathbf{h}$ is the impulse response of the filter

$$\begin{aligned}\mathbf{h}(x,y) &= \iint_{-\infty}^{\infty} \text{rect}\left(\frac{f_X}{2B_X}\right) \text{rect}\left(\frac{f_Y}{2B_Y}\right) \exp\left[j2\pi(f_X x + f_Y y)\right] df_X\, df_Y \\ &= 4B_X B_Y \,\text{sinc}(2B_X x)\, \text{sinc}(2B_Y y)\end{aligned}$$

Noting that

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

Eq. (2-30) becomes

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

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

$$\mathbf{g}(x,y) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \mathbf{g}\left(\frac{n}{2B_X}, \frac{m}{2B_Y}\right) \operatorname{sinc}\left[2B_X\left(x - \frac{n}{2B_X}\right)\right] \operatorname{sinc}\left[2B_Y\left(y - \frac{m}{2B_Y}\right)\right] \quad (2\text{-}31)$$

Equation (2-31) 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 sample point, an interpolation function consisting of a product of sinc functions.

The above result is by no means the only possible sampling theorem. Two rather arbitrary choices were made in the analysis, and alternative assumptions 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 transfer function (2-29). Alternative theorems derived by making different choices at these two points are, of course, no less valid than Eq. (2-31); in fact, in some cases alternative theorems can be more "efficient" in the sense that fewer samples per unit area are required to assure complete recovery. The reader interested in further pursuing this extra richness of the multidimensional sampling theory is referred to the works of Bracewell [Ref. 2-8] and of Peterson and Middleton [Ref. 2-9]. In addition, sampling theorems involving the values of derivatives of the function as well as the function itself have been discussed by Linden [Ref. 2-10].

PROBLEMS

2-1 Prove the following properties of δ functions:

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

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

2-2 Prove the following Fourier transform relations:

(*a*) $\mathcal{F}\{\operatorname{rect}(x)\operatorname{rect}(y)\} = \operatorname{sinc}(f_X)\operatorname{sinc}(f_Y)$

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

Prove the following generalized Fourier transform relations:
(c) $\mathcal{F}\{1\} = \delta(f_x, f_y)$
(d) $\mathcal{F}\{\operatorname{sgn}(x)\operatorname{sgn}(y)\} = \left(\frac{1}{j\pi f_X}\right)\left(\frac{1}{j\pi f_Y}\right)$

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) $\mathcal{F}\{g(x,y)h(x,y)\} = \mathcal{F}\{g(x,y)\} * \mathcal{F}\{h(x,y)\}$
(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 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) $\mathcal{B}\{\exp(-\pi r^2)\} = \exp(-\pi\rho^2)$

2-5 The expression

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

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 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 transform $P(f_X, f_Y)$.

2-6 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 expression for

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

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

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

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

where $\mathcal{H}_m\{\quad\}$ is a 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*) Show that for a more general case of an arbitrary angular dependence $g_\theta(\theta)$, the Fourier transform can be expressed by the following infinite series of Hankel transforms:

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

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

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) condition 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 a function with no nonzero spectral components outside a circle of radius B in the frequency plane obeys the following sampling theorem:

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

2-10 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* that characterizes this system? If so, what is it? If not, why not?

2-11 The "equivalent area" Δ_{XY} of a function $g(x,y)$ can be defined by

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

while the "equivalent bandwidth" of g is defined in terms of its transform G by

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

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

2-12 A certain complex-valued function of two independent variables (x,y) has a spatial Fourier transform that is identically zero outside the region $|f_X| \leq B_X$, $|f_Y| \leq B_Y$ in the frequency domain. Show that the portion of this function extending over the region $|x| \leq X$, $|y| \leq Y$ in the space domain can be specified (approximately) by $32B_XB_YXY$ real numbers. Why is this only an approximation, and when will it be a good one? (The number $16B_XB_YXY$ is commonly called the *space-bandwidth product* of the portion of the function considered.)

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 space-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 $U_o'(x,y)$, 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

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

REFERENCES

2-1 Papoulis, A.: *The Fourier Integral and Its Applications*, McGraw-Hill Book Company, New York, 1962.

2-2 Bracewell, R. N.: *The Fourier Transform and Its Applications*, McGraw-Hill Book Company, New York, 1965.

2-3 Lighthill, M. J.: *Introduction to Fourier Analysis and Generalized Functions*, Cambridge University Press, New York, 1960.

2-4 Kailath, T. in E. J. Baghdady (ed.): Channel Characterization: Time-variant Dispersive Channels, *Lectures on Communication System Theory*, McGraw-Hill Book Company, New York, 1960.

2-5 Lohmann, A. W., and D. P. Paris: Space-Variant Image Formation, *J. Opt. Soc. Am.*, **55**:1007 (1965).

2-6 Whittaker, E. T.: On the Functions Which Are Represented by the Expansions of the Interpolation Theory, *Proc. Roy. Soc. Edinburgh, Sect. A*, **35**:181 (1915).

2-7 Shannon, C. E.: Communication in the Presence of Noise, *Proc. IRE*, **37**:10 (1949).

2-8 Bracewell, R. N.: Two-Dimensional Aerial Smoothing in Radio Astronomy, *Australia J. Phys.*, **9**:297 (1956).

2-9 Peterson, D. P., and D. Middleton: Sampling and Reconstruction of Wave-Number-Limited Functions in N-Dimensional Euclidean Spaces, *Information and Control*, **5**:279 (1962).

2-10 Linden, D. A.: A Discussion of Sampling Theorems, *Proc. IRE*, **47**:1219 (1959).

APPENDIX

A. DIRAC DELTA FUNCTIONS

The one-dimensional Dirac delta function, widely used in electric circuit analysis, can be defined as the limit of a sequence of pulses of decreasing width, increasing height, and unit area. There are, of course, a multitude of different pulse shapes that can be used in the definition; three equally acceptable definitions are

$$\delta(t) = \lim_{N\to\infty} N \exp(-N^2\pi t^2) \tag{A-1a}$$

$$\delta(t) = \lim_{N\to\infty} N \operatorname{rect}(Nt) \tag{A-1b}$$

$$\delta(t) = \lim_{N\to\infty} N \operatorname{sinc}(Nt) \tag{A-1c}$$

While the δ function is used in circuit analysis to represent a sharp, intense pulse of current or voltage, the analogous concept in optics is a point source of light, or a *spatial* pulse of unit area. The definition of a δ function on a two-dimensional space is a simple extension of the one-dimensional case, although there is even greater latitude in the possible choice for the functional form of the pulses. Possible definitions of the spatial δ function include

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

$$\delta(x,y) = \lim_{N\to\infty} N^2 \operatorname{rect}(Nx)\operatorname{rect}(Ny) \tag{A-2b}$$

$$\delta(x,y) = \lim_{N\to\infty} N^2 \operatorname{sinc}(Nx)\operatorname{sinc}(Ny) \tag{A-2c}$$

$$\delta(x,y) = \lim_{N\to\infty} \frac{N^2}{\pi} \operatorname{circ}(N\sqrt{x^2+y^2}) \tag{A-2d}$$

$$\delta(x,y) = \lim_{N\to\infty} N \frac{J_1(2\pi N\sqrt{x^2+y^2})}{\sqrt{x^2+y^2}} \tag{A-2e}$$

Definitions (A-2a) to (A-2c) are separable in rectangular coordinates, while definitions (A-2d) and (A-2e) are circularly symmetric. In some applications one definition may be more convenient than others, and the definition best suited for the problem can be chosen.

Each of the above definitions of the spatial δ function has the following fundamental properties:

$$\delta(x,y) = \begin{cases} \infty & x = y = 0 \\ 0 & \text{otherwise} \end{cases} \tag{A-3}$$

$$\iint_{-\epsilon}^{\epsilon} \delta(x,y)\, dx\, dy = 1 \qquad \text{any } \epsilon > 0 \tag{A-4}$$

$$\iint_{-\infty}^{\infty} \mathbf{g}(\xi,\eta)\delta(x - \xi, y - \eta)\, d\xi\, d\eta = \mathbf{g}(x,y) \tag{A-5}$$

at each point of continuity of $\mathbf{g}$

Property (A-5) is often referred to as the *sifting* property of the δ function. An additional property of considerable importance can be proved from any of the definitions (cf. Prob. 2–1a), namely,

$$\delta(ax,by) = \frac{1}{|ab|}\,\delta(x,y) \tag{A-6}$$

There is, of course, no reason why the δ function cannot be defined on a space of higher dimensionality than two, but the properties of such functions are exactly analogous to their counterparts on spaces of lower dimensionality.

B. DERIVATION OF FOURIER TRANSFORM THEOREMS

In this section, brief proofs of basic Fourier transform theorems are presented. For more rigorous derivations, the reader should consult Ref. 2-1 or 2-2.

1. *Linearity theorem.* $\mathfrak{F}\{\alpha\mathbf{g} + \beta\mathbf{h}\} = \alpha\mathfrak{F}\{\mathbf{g}\} + \beta\mathfrak{F}\{\mathbf{h}\}$

Proof: This theorem follows directly from the linearity of the integrals that define the Fourier transform.

2. *Similarity theorem.* If $\mathfrak{F}\{\mathbf{g}(x,y)\} = \mathbf{G}(f_X,f_Y)$, then

$$\mathfrak{F}\{\mathbf{g}(ax,by)\} = \frac{1}{|ab|}\,\mathbf{G}\left(\frac{f_X}{a},\frac{f_Y}{b}\right)$$

Proof:

$$\mathfrak{F}\{\mathbf{g}(ax,by)\} = \iint_{-\infty}^{\infty} \mathbf{g}(ax,by) \exp\left[-j2\pi(f_X x + f_Y y)\right] dx\, dy$$

$$= \iint_{-\infty}^{\infty} \mathbf{g}(ax,by) \exp\left[-j2\pi\left(\frac{f_X}{a}\, ax + \frac{f_Y}{b}\, by\right)\right] \frac{dax}{|a|}\, \frac{dby}{|b|}$$

$$= \frac{1}{|ab|}\, \mathbf{G}\left(\frac{f_X}{a}, \frac{f_Y}{b}\right)$$

3. ***Shift theorem.*** If $\mathfrak{F}\{\mathbf{g}(x,y)\} = \mathbf{G}(f_X,f_Y)$, then

$$\mathfrak{F}\{\mathbf{g}(x - a, y - b)\} = \mathbf{G}(f_X,f_Y) \exp\left[-j2\pi(f_X a + f_Y b)\right]$$

Proof:

$$\mathfrak{F}\{\mathbf{g}(x - a, y - b)\}$$

$$= \iint_{-\infty}^{\infty} \mathbf{g}(x - a, y - b) \exp\left[-j2\pi(f_X x + f_Y y)\right] dx\, dy$$

$$= \iint_{-\infty}^{\infty} \mathbf{g}(x',y') \exp\left\{-j2\pi[f_X(x' + a) + f_Y(y' + b)]\right\} dx'\, dy'$$

$$= \mathbf{G}(f_X,f_Y) \exp\left[-j2\pi(f_X a + f_Y b)\right]$$

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

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

Proof:

$$\iint_{-\infty}^{\infty} |\mathbf{g}(x,y)|^2\, dx\, dy = \iint_{-\infty}^{\infty} \mathbf{g}(x,y)\, \mathbf{g}^*(x,y)\, dx\, dy$$

$$= \iint_{-\infty}^{\infty} dx\, dy \left[\iint_{-\infty}^{\infty} d\xi\, d\eta\, \mathbf{G}(\xi,\eta) \exp\left[j2\pi(x\xi + y\eta)\right]\right]$$

$$\left[\iint_{-\infty}^{\infty} d\alpha\, d\beta\, \mathbf{G}^*(\alpha,\beta) \exp\left[-j2\pi(x\alpha + y\beta)\right]\right]$$

$$= \iint_{-\infty}^{\infty} d\xi\, d\eta\, \mathbf{G}(\xi,\eta) \iint_{-\infty}^{\infty} d\alpha\, d\beta\, \mathbf{G}^*(\alpha,\beta)$$

$$\left[\iint_{-\infty}^{\infty} \exp\left\{j2\pi[x(\xi - \alpha) + y(\eta - \beta)]\right\} dx\, dy\right]$$

$$= \iint_{-\infty}^{\infty} d\xi\, d\eta\, \mathbf{G}(\xi,\eta) \iint_{-\infty}^{\infty} d\alpha\, d\beta\, \mathbf{G}^*(\alpha,\beta)\delta(\xi - \alpha, \eta - \beta)$$

$$= \iint_{-\infty}^{\infty} |\mathbf{G}(\xi,\eta)|^2\, d\xi\, d\eta$$

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

$$\mathfrak{F}\left\{\iint_{-\infty}^{\infty} \mathbf{g}(\xi,\eta)\mathbf{h}(x - \xi, y - \eta)\, d\xi\, d\eta\right\} = \mathbf{G}(f_X,f_Y)\mathbf{H}(f_X,f_Y)$$

Proof:

$$\mathfrak{F}\left\{\iint_{-\infty}^{\infty} \mathbf{g}(\xi,\eta)\mathbf{h}(x - \xi, y - \eta)\, d\xi\, d\eta\right\}$$

$$= \iint_{-\infty}^{\infty} \mathbf{g}(\xi,\eta)\mathfrak{F}\{\mathbf{h}(x - \xi, y - \eta)\}\, d\xi\, d\eta$$

$$= \iint_{-\infty}^{\infty} \mathbf{g}(\xi,\eta) \exp\,[-j2\pi(f_X\xi + f_Y\eta)]\, d\xi\, d\eta\, \mathbf{H}(f_X,f_Y)$$

$$= \mathbf{G}(f_X,f_Y)\mathbf{H}(f_X,f_Y)$$

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

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

Proof:

$$\mathfrak{F}\left\{\iint_{-\infty}^{\infty} \mathbf{g}(\xi,\eta)\mathbf{g}^*(\xi - x, \eta - y)\, d\xi\, d\eta\right\}$$

$$= \mathfrak{F}\left\{\iint_{-\infty}^{\infty} \mathbf{g}(\xi' + x, \eta' + y)\mathbf{g}^*(\xi',\eta')\, d\xi'\, d\eta'\right\}$$

$$= \iint_{-\infty}^{\infty} d\xi'\, d\eta'\, \mathbf{g}^*(\xi',\eta')\mathfrak{F}\{\mathbf{g}(\xi' + x, \eta' + y)\}$$

$$= \iint_{-\infty}^{\infty} d\xi'\, d\eta'\, \mathbf{g}^*(\xi',\eta') \exp\,[j2\pi(f_X\xi' + f_Y\eta')]\mathbf{G}(f_X,f_Y)$$

$$= \mathbf{G}^*(f_X,f_Y)\mathbf{G}(f_X,f_Y) = |\mathbf{G}(f_X,f_Y)|^2$$

7. Fourier integral theorem. At each point of continuity of $\mathbf{g}$,

$$\mathfrak{F}\mathfrak{F}^{-1}\{\mathbf{g}(x,y)\} = \mathfrak{F}^{-1}\mathfrak{F}\{\mathbf{g}(x,y)\} = \mathbf{g}(x,y)$$

At each point of discontinuity of $\mathbf{g}$, the two successive transformations yield the angular average of the value of $\mathbf{g}$ in a small neighborhood of that point.

Proof: Let the function $\mathbf{g}_R(x,y)$ be defined by

$$\mathbf{g}_R(x,y) = \iint\limits_{A_R} \mathbf{G}(f_X,f_Y) \exp\,[j2\pi(f_X x + f_Y y)]\, df_X\, df_Y$$

where A_R is a circle of radius R, centered at the origin of the $f_X f_Y$ plane. To prove the theorem, it suffices to show that at each point of continuity of $\mathbf{g}$,

$$\lim_{R\to\infty} \mathbf{g}_R(x,y) = \mathbf{g}(x,y)$$

and that at each point of discontinuity of $\mathbf{g}$,

$$\lim_{R\to\infty} \mathbf{g}_R(x,y) = \frac{1}{2\pi}\int_0^{2\pi} \mathbf{g}_o(\theta)\, d\theta$$

where $\mathbf{g}_o(\theta)$ is the angular dependence of $\mathbf{g}$ in a small neighborhood about the point in question.

Some initial straightforward manipulation can be performed as follows:

$$\mathbf{g}_R(x,y) = \iint\limits_{A_R} \left\{ \iint\limits_{-\infty}^{\infty} d\xi\, d\eta\, \mathbf{g}(\xi,\eta) \exp\,[-j2\pi(f_X\xi + f_Y\eta)]\right\} \exp\,[j2\pi(f_X x + f_Y y)]\, df_X\, df_Y$$

$$= \iint\limits_{-\infty}^{\infty} d\xi\, d\eta\, \mathbf{g}(\xi,\eta) \iint\limits_{A_R} df_X\, df_Y \exp\,\{j2\pi[f_X(x-\xi) + f_Y(y-\eta)]\}$$

Noting that

$$\iint\limits_{A_R} df_X\, df_Y \exp\,\{j2\pi[f_X(x-\xi) + f_Y(y-\eta)]\} = R\left[\frac{J_1(2\pi Rr)}{r}\right]$$

where $r = \sqrt{(x-\xi)^2 + (y-\eta)^2}$, we have

$$\mathbf{g}_R(x,y) = \iint\limits_{-\infty}^{\infty} d\xi\, d\eta\, \mathbf{g}(\xi,\eta) R\left[\frac{J_1(2\pi Rr)}{r}\right]$$

Suppose initially that (x,y) is a point of continuity of $\mathbf{g}$. Then

$$\lim_{R\to\infty} \mathbf{g}_R(x,y) = \iint_{-\infty}^{\infty} d\xi\, d\eta\, \mathbf{g}(\xi,\eta) \lim_{R\to\infty} R\left[\frac{J_1(2\pi Rr)}{r}\right]$$

$$= \iint_{-\infty}^{\infty} d\xi\, d\eta\, \mathbf{g}(\xi,\eta)\delta(x-\xi, y-\eta) = \mathbf{g}(x,y)$$

where Eq. (A-2*e*) has been used in the second step. Thus the first part of the theorem has been proved.

Consider next a point of discontinuity of $\mathbf{g}$. Without loss of generality that point can be taken to be the origin. Thus we write

$$\mathbf{g}_R(0,0) = \iint_{-\infty}^{\infty} d\xi\, d\eta\, \mathbf{g}(\xi,\eta) R\left[\frac{J_1(2\pi Rr)}{r}\right]$$

where $r = \sqrt{\xi^2 + \eta^2}$. But for sufficiently large R, the quantity in brackets has significant value only in a small neighborhood of the origin. In addition, in this small neighborhood the function $\mathbf{g}$ depends (approximately) only on the angle θ about that point, and therefore

$$\mathbf{g}_R(0,0) \cong \int_0^{2\pi} \mathbf{g}_o(\theta)\, d\theta \int_0^{\infty} rR\left[\frac{J_1(2\pi Rr)}{r}\right] dr$$

where $\mathbf{g}_o(\theta)$ represents the θ dependence of $\mathbf{g}$ about the origin. Finally, noting that

$$\int_0^{\infty} rR\left[\frac{J_1(2\pi Rr)}{r}\right] dr = \frac{1}{2\pi}$$

we conclude that

$$\lim_{R\to\infty} \mathbf{g}_R(0,0) = \frac{1}{2\pi}\int_0^{2\pi} \mathbf{g}_o(\theta)\, d\theta$$

and the proof is thus complete.