# Wigner distributions, linear canonical transforms, and phase-space optics.

Haldun M. Ozaktas
Bilkent University
Department of Electrical Engineering
TR-06533 Bilkent, Ankara, Turkey

M. Alper Kutay
Drexel University
Department of Electrical and Computer Engineering
Philadelphia, PA 19104, USA

Technical Report BU-CE-0006
Bilkent University
Department of Computer Engineering
TR-O6533 Bilkent, Ankara, Turkey

January 2000

# Contents

# Chapter 1

# Signals, Systems, and Transformations

## 1.1 Signals

### 1.1.1 Signals

Signals are information bearing entities which are usually represented by one or more functions of one or more independent variables. We will mostly deal with signals represented by a scalar function of one or two real variables. For instance, a voltage signal $\check{v}$ may be represented as a function of time $t$ by the function $\check{v}(t) = 2\cos(2\pi 10t)\,\mathrm{V}$, with $t$ being measured in seconds. It is common to think of and to refer to this function as the signal itself, as a consequence of the primacy we attach to time as an independent variable. However, the same information can be equally well represented by other functions of other variables. For instance, the same information can be represented as a function of temporal frequency $f$ (measured in Hertz) in the form $\check{V}(f) = [\delta(f-10)+\delta(f+10)]\,\mathrm{V}$, where $\check{V}(f)$ is the Fourier transform of $\check{v}(t)$. Thus, it will be more useful to use the term *signal* to refer to the signal $\check{v}$ as an information-bearing entity in the abstract, and to refer to the *function* $\check{v}(t)$ as the *time-domain representation of the signal.* When there is possibility of confusion, we will write $\check{v}_t(t)$ instead of $\check{v}(t)$, $\check{v}_f(f)$ instead of $\check{V}(f)$, and so on to identify the different functional representations of the signal $\check{v}$. This notion of a signal is similar to the notion of a vector in classical mechanics and the notion of a ket vector in quantum mechanics (Cohen-Tannoudji, Diu, and Laloë 1977). A vector $\mathbf{r}$ is a geometrical entity independent of any coordinate system. One way of representing it is with respect to a particular rectangular coordinate system in the form $\mathbf{r} = x\hat{\mathbf{u}}_x + y\hat{\mathbf{u}}_y + z\hat{\mathbf{u}}_z$ or $\mathbf{r} = (x, y, z)$, where $\hat{\mathbf{u}}_x, \hat{\mathbf{u}}_y, \hat{\mathbf{u}}_z$ are unit vectors along the coordinate axes. Many other representations with respect to many other coordinate systems are possible. When the distinction is not crucial, we will simply say "the signal $\check{f}(t)$" rather than "the time-domain representation $\check{f}(t)$ of the signal $\check{f}$."

### 1.1.2 Notation

Throughout this report, $u, v, w$ will be used as generic dimensionless coordinate variables, which may be referred to as time or space depending on the context. The associated frequency-domain variables will be denoted by $u_1, v_1, w_1$. The same variables will also be denoted by $\mu \equiv u_1, \nu \equiv v_1, \eta \equiv w_1$ in simpler contexts. Dimensionless coordinate and frequency vectors will be denoted by $\mathbf{q}$ and $\mathbf{q}_1$ respectively, with $\boldsymbol{\varsigma} \equiv \mathbf{q}_1$ in simpler contexts. Functions which take dimensionless arguments will simply be denoted by lowercase letters, such as $f(u)$, $g(u, v)$. Their Fourier transforms will be denoted by $F(\mu)$, $G(\mu, \nu)$.

The symbol $t$ will be used for the time coordinate, $f \equiv t_1$ for temporal frequency, $r_x \equiv x, r_y \equiv y, r_z \equiv z$ for the space coordinates, and $\sigma_x \equiv x_1, \sigma_y \equiv y_1, \sigma_z \equiv z_1$ for spatial frequencies. ($f$ will also commonly be used to represent a generic signal $f(\cdot)$ and the focal length of a lens, but this will cause no confusion.) Spatial coordinate and frequency vectors will be denoted by $\mathbf{r}$ and $\boldsymbol{\sigma}$ respectively. Functions whose arguments have the dimensions of time or distance will be denoted as $\check{f}(t)$, $\check{g}(x, y)$, $\check{F}(f)$, $\check{G}(\sigma_x, \sigma_y)$.

Integrals whose limits are not indicated will denote integrals from minus to plus infinity. Likewise, summations whose limits are not indicated will denote summations over the complete range of the indices, which is often either from minus to plus infinity or from zero to plus infinity.

When we use the square root function and unless we indicate otherwise, $\sqrt{z}$ will mean the square root of $z$ whose argument lies in the interval $(-\pi/2, \pi/2]$. We denote the imaginary unit by $i = \sqrt{-1}$.

We will use "$\equiv$" instead of simply "$=$" when it is important to emphasize that the left hand side is defined as the expression on the right hand side.

### 1.1.3 Some commonly used functions

The rectangle function $\mathrm{rect}(u)$ is defined to be equal to 1 in the interval $(-0.5, 0.5)$, 0.5 at $u = \pm 0.5$, and 0 elsewhere. The unit step function $\mathrm{step}(u)$ is defined to be equal to 1 when $u > 0$, 0.5 at $u = 0$, and 0 when $u < 0$ and the sign function $\mathrm{sgn}(u)$ is defined as $\mathrm{sgn}(u) = 2\mathrm{step}(u) - 1$. The sinc (or interpolation) function $\mathrm{sinc}(u)$ is defined as $\mathrm{sinc}(u) = \sin(\pi u)/(\pi u)$. The Gaussian function $\mathrm{gauss}(u)$ is given by $\mathrm{gauss}(u) = \exp(-\pi u^2)$, the harmonic function $\mathrm{har}(u)$ by $\mathrm{har}(u) = \exp(i2\pi u)$, and the chirp function $\mathrm{chirp}(u)$ by $\mathrm{chirp}(u) = e^{-i\pi/4} \exp(i\pi u^2)$.

The Dirac delta function $\delta(u)$ is a generalized function which is zero everywhere except at $u = 0$, such that its integral over any interval including $u = 0$ is equal to unity. It may be defined as the limit of parametric continuous functions:

$$\delta(u) = \lim_{c \to 0} c^{-1} \mathrm{rect}(u/c), \tag{1.1}$$

$$\delta(u) = \lim_{c \to 0} c^{-1} \mathrm{sinc}(u/c), \tag{1.2}$$

$$\delta(u) = \lim_{c \to 0} c^{-1} \mathrm{gauss}(u/c), \tag{1.3}$$

$$\delta(u) = \lim_{c\to 0} c^{-1}\mathrm{chirp}(u/c), \tag{1.4}$$

where $c > 0$. The last two equations can be rewritten for all real values of $c$ as

$$\delta(u) = \lim_{c\to 0} \frac{1}{\sqrt{|c|}} e^{-\pi u^2/|c|}, \tag{1.5}$$

$$\delta(u) = \lim_{c\to 0} e^{-i\pi \mathrm{sgn}(c)/4} \frac{1}{\sqrt{|c|}} e^{i\pi u^2/c} = \lim_{c\to 0} e^{-i\pi/4} \sqrt{\frac{1}{c}}\, e^{i\pi u^2/c}. \tag{1.6}$$

Alternatively, the delta function may be defined through its effect under the integral sign. For every continuous function $f(u)$,

$$f(u) = \int_{-\infty}^{\infty} \delta(u-u')f(u')\,du', \tag{1.7}$$

which is known as the *sifting property*. Table 1.1 is a list of some of the common properties of the delta function. Of particular importance is the following identity:

$$\delta(u) = \int_{-\infty}^{\infty} e^{\pm i2\pi u\mu}\,d\mu. \tag{1.8}$$

The first and higher-order derivatives of the delta function are denoted as $\delta'(u)$, $\delta''(u)$, and so on. The comb function is defined as $\mathrm{comb}(u) = \sum_{n=-\infty}^{\infty} \delta(u-n)$. The Kronecker

| | |
|---|---|
| 1. | $\delta(Mu) = \delta(u)/\lvert M\rvert$ |
| 2. | $f(u)\delta(u-\xi) = f(\xi)\delta(u-\xi)$ |
| 3. | $\int \delta(u-\xi)f(u)\,du = f(\xi)$ |
| 4. | $\int \delta(u-\xi)\delta(u-\xi')\,du = \delta(\xi-\xi')$ |
| 5. | $\int e^{\pm i2\pi(u-\xi)u'}\,du' = \delta(u-\xi)$ |
| 6. | $\int \delta'(u-u')f(u')\,du' = df(u)/du$ |
| 7. | $\int_{0^-}^{\infty} \delta(u-u')\,du' = \mathrm{step}(u)$ |
| 8. | $\delta(u) = d[\mathrm{step}(u)]/du$ |

Table 1.1: Properties of the Dirac delta function. $M, \xi, \xi'$ are real numbers.

delta $\delta_{ll'}$ is defined to be 0 when $l \neq l'$ and 1 when $l = l'$. A common identity valid for arbitrary real $\delta u$ is

$$\sum_{n=-\infty}^{\infty} \delta(u+n\delta u) = \frac{1}{\delta u} \sum_{n=-\infty}^{\infty} e^{i2\pi nu/\delta u}, \tag{1.9}$$

whose right hand side can be interpreted as the Fourier series of the comb function.

### 1.1.4 Analytic signals and the Hilbert transform

We will mostly deal with complex signals bearing the same information as the real physical signal. A real physical signal $f(u)$ and the associated complex signal $f_{\mathrm{as}}(u)$, also known

as the *analytic signal*, are related by

$$f_{\rm as}(u) = f(u) + if_{\rm H}(u), \tag{1.10}$$

$$f(u) = \Re[f_{\rm as}(u)], \tag{1.11}$$

where $\Re[\cdot]$ denotes the real part of a complex entity and

$$f_{\rm H}(u) = \int_{-\infty}^{\infty} \frac{1}{\pi(u-u')}\, f(u')\, du' \tag{1.12}$$

is the *Hilbert transform* of $f(u)$. The Fourier transform of $f_{\rm H}(u)$ is given by $-i{\rm sgn}(\mu)F(\mu)$ and $f(u)$ is orthogonal to its Hilbert transform: $\int f^*(u) f_{\rm H}(u)\, du = 0$. The Fourier transform $\mathcal{F}[f_{\rm as}(u)](\mu)$ of $f_{\rm as}(u)$ is obtained from $F(\mu)$ according to

$$\mathcal{F}[f_{\rm as}(u)](\mu) = 2{\rm step}(\mu)F(\mu). \tag{1.13}$$

In the event that $f_{\rm as}(u)$ is a narrowband signal whose spectrum is centered around some center frequency $\mu_0$, it is convenient to express it in the form

$$f_{\rm as}(u) = A_c(u)e^{i2\pi\mu_0 u}, \tag{1.14}$$

where $A_c(u)$ is known as the *complex envelope*, and is a lowpass (slowly varying) function. In general, the analytic signal of $A(u)\cos[2\pi\mu_0 u + \phi(u)]$ is not $A(u)\exp[i2\pi\mu_0 u + i\phi(u)]$, where $A(u)$ and $\phi(u)$ are real functions. However, this is approximately true when $A(u)$ and $\phi(u)$ are slowly varying functions (Cohen 1989).

A *monochromatic* signal is one which consists of only a single frequency: $f(u) = A_0\cos(2\pi\mu_0 u + \phi_0)$ for some particular $\mu_0$, $A_0$, and $\phi_0$. The associated complex representation (analytic signal) is given by $A_0\exp(i2\pi\mu_0 u + i\phi_0)$ and the complex envelope is simply $A_c(u) = A_0\exp(i\phi_0)$. In this case the complex envelope is also known as the *phasor* of $f(u)$. The real signal is recovered by multiplying the phasor by $\exp(i2\pi\mu_0 u)$ and taking the real part. Since we will mostly deal with complex representations, we will often omit the subscript "as" and simply write $f(u)$ instead of $f_{\rm as}(u)$.

### 1.1.5 Signal spaces

A vector space is a set of entities for which addition and scalar multiplication have been defined such that certain axioms are satisfied (see the appendix to this chapter). A set of signals which constitute a vector space is referred to as a *signal space*. Although we will not rigorously specify what this means mathematically, we will restrict ourselves to the space of signals whose members are "physically realizable," which in particular implies that they have finite energy, and that their representations are smooth, and negligible outside some finite interval (Cohen-Tannoudji, Diu, and Laloë 1977, page 94). We will also use certain physically unrealizable signals which have infinite energy, but which nevertheless serve as useful intermediaries (such as the delta and harmonic functions), and occasionally deal

with discontinuous functions. The fact that these signals are not "physically realizable" will not overly concern us in this report.

For concreteness, we will concentrate on the space of functions consisting of the representations of the members of a signal space in discrete and continuous domains (such as the time domain). We will mostly use $l, m, \ldots$ to denote the independent variable(s) in a discrete domain. Likewise, we will mostly use $u, v, \ldots$ to denote the independent variable(s) in a continuous domain. The addition of two functions is defined as ordinary arithmetic addition and scalar multiplication is defined as ordinary arithmetic multiplication with a complex number. The *inner product* $\langle f, g \rangle$ of two signals $f$ and $g$ may be defined in terms of their discrete or continuous functional representations in the $l$ or $u$ domain as

$$\langle f, g \rangle = \sum_l f^*(l) g(l), \tag{1.15}$$

$$\langle f, g \rangle = \int f^*(u) g(u)\, du, \tag{1.16}$$

respectively. The *energy* $\|f\|^2$ and *norm* $\|f\|$ of a signal $f$ are defined by $\|f\|^2 = \langle f, f \rangle$. Two signals whose inner product is zero are called *orthogonal* to each other. The distance between two signals $f$ and $g$ is defined to be the norm of their difference: $\|f - g\|$. We will later show that the definitions of the inner product, norm, and energy of a signal are independent of the particular functional representation or domain in which we calculate it (by using equation 1.15 or 1.16).

## 1.2 Systems

### 1.2.1 Systems

A *system* is a process, event, mechanism or the like that maps a given signal into another signal. Mathematically, a system is a rule for assigning to any element $f$ of some set of signals, an element $g$ of (another or the same) set of signals. The signal $f$ is referred to as the input, and the signal $g$ is referred to as the output of the system. In other words, a system is a mapping from the input set of signals to the output set of signals (Papoulis 1977).

The rule relating the output signal $g$ to the input signal $f$ is denoted as

$$g = \mathcal{S}[f]. \tag{1.17}$$

In this notation, $f$ becomes the argument of the (possibly many-to-one) relation $\mathcal{S}[\cdot]$ which characterizes the system. Alternatively, by interpreting $\mathcal{S}$ as an operator that operates on objects to its right, we may write the above in the form

$$g = \mathcal{S}f. \tag{1.18}$$

The same relationships can be written more explicitly in terms of the time-domain (or space-domain) representations of the signals in the form

$$g(u) = \{\mathcal{S}[f(u)]\}(u) \tag{1.19}$$

or more simply as $g(u) = \{\mathcal{S}[f]\}(u)$, or even $g(u) = \mathcal{S}[f(u)]$ when there is no room for confusion. In operator notation one may write

$$g(u) = \{\mathcal{S}f\}(u), \tag{1.20}$$

or more simply $g(u) = \mathcal{S}f(u)$.

### 1.2.2 Linearity and superposition integrals

For a linear system $\mathcal{L}$, the output corresponding to a linear superposition of a sequence of inputs $f_j$, is the same linear superposition of the corresponding sequence of outputs $g_j$:

$$g_j = \mathcal{L}[f_j] \quad \text{for all } j \qquad \Rightarrow \qquad \sum_j \alpha_j g_j = \mathcal{L}\left[\sum_j \alpha_j f_j\right], \tag{1.21}$$

where $\alpha_j$ are arbitrary complex coefficients. If we have continuously many inputs $f_\upsilon$, the summations above should be replaced with integrals over $\upsilon$.

$$g_\upsilon = \mathcal{L}[f_\upsilon] \quad \text{for all } \upsilon \qquad \Rightarrow \qquad \int \alpha_\upsilon g_\upsilon \, d\upsilon = \mathcal{L}\left[\int \alpha_\upsilon f_\upsilon \, d\upsilon\right], \tag{1.22}$$

where $\alpha_\upsilon$ are arbitrary complex coefficients.

Let us express the input function $f(u)$ as a linear superposition of shifted delta functions as

$$f(u) = \int_{-\infty}^{\infty} \delta(u - u') f(u') \, du', \tag{1.23}$$

and let $h(u, u')$ denote the output of a linear system when the input is $\delta(u - u')$:

$$h(u, u') = \mathcal{L}[\delta(u - u')]. \tag{1.24}$$

It follows that the output $g(u)$ is related to the input $f(u)$ by the relation

$$g(u) = \int_{-\infty}^{\infty} h(u, u') f(u') \, du'. \tag{1.25}$$

Now, let the signal $f$ be input to the linear system $\mathcal{L}_1$ and the output be input to a second linear system $\mathcal{L}_2$:

$$\mathcal{L}_2\left[\mathcal{L}_1[f]\right], \tag{1.26}$$

or simply $\mathcal{L}_2\mathcal{L}_1[f]$ or $\mathcal{L}_2\mathcal{L}_1 f$. The kernel $h(u, u')$ corresponding to the composite system $\mathcal{L} = \mathcal{L}_2\mathcal{L}_1$ can be shown to be given by

$$h(u, u') = \int h_2(u, u'') h_1(u'', u') \, du''. \tag{1.27}$$

### 1.2.3 Some special linear systems

A number of systems that will be of special interest are tabulated in table 1.2 together with their linear transform kernels $h(u,u')$ and inverses. The inverse $\mathcal{L}^{-1}$ of a system $\mathcal{L}$, if it exists, satisfies $\mathcal{L}\mathcal{L}^{-1} = \mathcal{L}^{-1}\mathcal{L} = \mathcal{I}$. The kernel of a system $h(u,u')$ and the kernel of its inverse $h^{-1}(u,u')$ satisfy the relation

$$\int h(u,u'')h^{-1}(u'',u')\,du'' = \delta(u-u'). \tag{1.28}$$

| Symbol | Kernel | Inverse Kernel |
|---|---|---|
| $\mathcal{I}$ | $\delta(u-u')$ | $\delta(u-u')$ |
| $\mathcal{P}$ | $\delta(u+u')$ | $\delta(u+u')$ |
| $\mathcal{M}_M$ | $\sqrt{\lvert M\rvert}\,\delta(u-Mu')$ | $(1/\sqrt{\lvert M\rvert}\,)\delta(u-u'/M)$ |
| $\mathcal{SH}_\xi$ | $\delta(u-u'+\xi)$ | $\delta(u-u'-\xi)$ |
| $\mathcal{PH}_\xi$ | $\exp(i2\pi\xi u)\delta(u-u')$ | $\exp(-i2\pi\xi u)\delta(u-u')$ |
| $\Lambda_h$ | $h(u)\delta(u-u')$ | $[1/h(u)]\delta(u-u')$ |
| $\mathcal{Q}_q$ | $\exp(-i\pi qu^2)\delta(u-u')$ | $\exp(i\pi qu^2)\delta(u-u')$ |
| $\Lambda_H$ | $h(u-u')$ | $h^{-1}(u-u')$ |
| $\mathcal{R}_r$ | $e^{-i\pi/4}\sqrt{1/r}\,\exp[i\pi(u-u')^2/r]$ | $e^{i\pi/4}(1/\sqrt{r}\,)\exp[-i\pi(u-u')^2/r]$ |
| $\mathcal{U}$ | $u\delta(u-u')$ | $u^{-1}\delta(u-u')$ |
| $\mathcal{D}$ | $(i2\pi)^{-1}\delta'(u-u')$ | $(i2\pi)\mathrm{step}(u-u')$ |
| $\mathcal{F}$ | $\exp(-i2\pi uu')$ | $\exp(i2\pi uu')$ |

Table 1.2: Special linear systems and their kernels. $\mathcal{I}$: Identity, $\mathcal{P}$: Parity, $\mathcal{M}_M$: Scaling, $\mathcal{SH}_\xi$: Shift or Translation, $\mathcal{PH}_\xi$: Phase shift, $\Lambda_h$: Multiplicative filter, $\mathcal{Q}_q$: Chirp multiplication, $\Lambda_H$: Convolutive filter, $\mathcal{R}_r$: Chirp convolution, $\mathcal{U}$: Coordinate multiplication, $\mathcal{D}$: Differentiation, $\mathcal{F}$: Fourier transform. $M$, $\xi$, $q$, $r$ are real parameters and $\delta'(u-u') = d[\delta(u-u')]/du$. $h^{-1}(u)$ is related to $h(u)$ through $\int h(u-u')h^{-1}(u')\,du' = \delta(u)$.

Among the systems listed in the table we comment only on chirp convolution, also known as the Fresnel transform or the Fresnel integral:

$$g(u) = \sqrt{\frac{1}{r}}\,\mathrm{chirp}(u/\sqrt{r}\,) * f(u) = e^{-i\pi/4}\sqrt{\frac{1}{r}}\int_{-\infty}^{\infty} f(u')e^{i\pi(u-u')^2/r}\,du'. \tag{1.29}$$

The Fresnel transform satisfies many properties (Gori 1994) of which we will need to know

$$h^{-1}(u,u';r) = h(u,u';-r) = h^*(u,u',r), \tag{1.30}$$

$$\mathcal{R}_{r_1}\mathcal{R}_{r_2} = \mathcal{R}_{r_1+r_2}, \tag{1.31}$$

$$\lim_{r\to 0}\mathcal{R}_r = \mathcal{I}. \tag{1.32}$$

Here $h(u, u'; r)$ explicitly shows the dependence of the Fresnel transform kernel on the parameter $r$. We also note that as $r \to 0$, the transform approaches the identity transform characterized by the kernel $\delta(u - u')$.

Another important class of systems, which includes most of the above as special cases, is the class of linear canonical transforms, which we will discuss at length in chapter 2.

### 1.2.4 Shift-invariance and convolution

Let the output of a system corresponding to the input $\delta(u)$ be denoted by $h(u)$. This system is called shift-invariant (or time-invariant or space-invariant) if the output $h(u, u')$ corresponding to the input $\delta(u - u')$ is equal to $h(u - u')$ for all $u'$. In this case $h(u)$ is called the impulse response and the relation between the output and the input becomes

$$g(u) = f(u) * h(u) \equiv \int_{-\infty}^{\infty} h(u - u') f(u')\, du'. \tag{1.33}$$

We say that $g(u)$ is the *convolution* of the two functions $f(u)$ and $h(u)$. The *correlation* of $f(u)$ and $h(u)$ is denoted by $R_{fh}(u)$ or $f(u) \star h(u)$ and is defined as

$$\begin{aligned} R_{fh}(u) \equiv f(u) \star h(u) \equiv f(u) * h^*(-u) &= \int_{-\infty}^{\infty} f(u + u') h^*(u')\, du' \\ &= \int_{-\infty}^{\infty} f(u' + u/2) h^*(u' - u/2)\, du'. \end{aligned} \tag{1.34}$$

Some properties of the convolution and correlation operations are summarized in table 1.3. We might also recall that the Fourier transform of $f(u) * h(u)$ is $F(\mu)H(\mu)$ and the Fourier transform of $f(u) \star h(u)$ is $F(\mu)H^*(\mu)$.

| | |
|---|---|
| 1. | $f(u) * h(u) = h(u) * f(u)$ |
| 2. | $f(-u) * h(-u) = g(-u)$ |
| 3. | $f(u) * [h_1(u) * h_2(u)] = [f(u) * h_1(u)] * h_2(u)$ |
| 4. | $f(u) * [h_1(u) + h_2(u)] = f(u) * h_1(u) + f(u) * h_2(u)$ |
| 5. | $f(u - \xi) * h(u) = g(u - \xi)$ |
| 6. | $R_{fh}(u) = f(u) \star h(u) = h^*(-u) \star f^*(-u) = R_{h^* f^*}(-u)$ |
| 7. | $f(-u) \star h(-u) = R_{fh}(-u)$ |
| 8. | $f(u) * [h_1(u) \star h_2(u)] = [f * h_1(u)] \star h_2(u)$ |
| 9. | $f(u) \star [h_1(u) + h_2(u)] = f(u) \star h_1(u) + f(u) \star h_2(u)$ |
| 10. | $f(u - \xi) \star h(u) = R_{fh}(u - \xi)$ |
| 11. | $R_{ff}(u) = R^*_{ff}(-u)$ |
| 12. | $\max[R_{ff}] = R_{ff}(0) = \int \lvert f(u) \rvert^2\, du$ |

Table 1.3: Properties of the convolution and correlation operations. $g(u) = f(u) * h(u)$, $R_{fh} = f(u) \star h(u)$, and $\xi$ is real.

Let us now consider the eigenvalue equation for a linear shift-invariant system with impulse response $h(u)$:

$$\{\mathcal{L}[f(u)]\}(u) = \lambda f(u), \tag{1.35}$$

which we may simply write as $\mathcal{L}f(u) = h(u)*f(u) = \lambda f(u)$ in operator notation. Rewriting the right hand side of equation 1.33 as $\int f(u-u')h(u')\,du'$, it is easy to show that $f(u) = \exp(i2\pi\mu u)$ is a solution for all real $\mu$ with eigenvalue $\lambda_\mu$ given by

$$\lambda_\mu = \int_{-\infty}^{\infty} h(u)e^{-i2\pi\mu u}\,du. \tag{1.36}$$

Interpreted as a function of $\mu$, we see that $\lambda_\mu$ is nothing but $H(\mu)$, the Fourier transform of $h(u)$. Thus we see that harmonic functions are eigenfunctions of linear shift-invariant systems, with the eigenvalues being given by the Fourier transform of the impulse response.

## 1.3 Representations and transformations

### 1.3.1 Systems versus transformations

Signals can be represented in many different ways which are distinct in appearance but nevertheless contain the same information. A common example is given by the time- and frequency-domain representations of a signal. Another common example is given by the two functional forms of an image with respect to two coordinate systems which are rotated with respect to each other. In both cases the two representations both contain the same information and either representation can be obtained from the other. The act of obtaining one representation from the other is called a *transformation.*

Different representations of a signal correspond to different coordinate systems or basis sets. Once an appropriate basis set is chosen, the signals may be expressed as a linear superposition of the elements of the basis set. In other words, the signal may be expanded in terms of the elements of the basis set. The coefficients appearing in this superposition or expansion, which uniquely specify the signal, constitute the representation of the signal with respect to this basis set.

It is important to distinguish clearly between systems and transformations, although mathematically they can take similar forms. A system is a rule that maps an input signal into an output signal. A system is usually a mathematical abstraction of a physical system which alters a physical input in a certain way to produce a physical output. For instance, a live television broadcasting system tries to reproduce the event in front of the camera as faithfully as possible on the retinas of human observers watching their televisions at home. A careful study of its physical components will enable characterization of this system and how it departs from this ideal. Other systems will intentionally alter the input, such as a pattern recognition system whose inputs are images and outputs are labels of recognized images. A system can alter the information content of the input signal in producing the output signal, so that systems need not always be invertible. A system, much like a signal,

is an abstract entity whose existence is independent of which coordinate system or basis set we choose to work with. The output may be represented in either the same representation as the input, or a different one, without affecting the nature of the system.

A transform(ation), on the other hand, is merely a change of the coordinate system or basis set used, with which we move from one representation of a signal to another. The signal is not altered, but expressed in another form bearing the same information. As such, transformations are usually invertible. Despite this clear distinction between systems and transformations, they are often mathematically expressed in the same way, and both are often represented by abstract operators (which we denote by calligraphic letters). As an example, consider an image signal $f$. Let this image be input to a system $\mathcal{L}$ which rotates the input image by $\pi/4$ in the clockwise direction to obtain the output image. Notice that the definition of the system is not tied to any particular coordinate system or representation. To relate the output image to the input image mathematically, we may choose a particular rectangular coordinate system in which the image $f$ is represented by the function $f(u, v)$. Then, the output image will be represented by the function $g(u, v)$ which is related to the input image by

$$g(u, v) = f\left(\cos(\pi/4)u - \sin(\pi/4)v, \sin(\pi/4)u + \cos(\pi/4)v\right). \tag{1.37}$$

Now, let us set the system $\mathcal{L}$ aside and consider a transformation $\mathcal{T}$ which rotates the coordinate axes by $\pi/4$ in the counterclockwise direction. The new coordinate axes $u', v'$ are related to $u, v$ as follows

$$\begin{aligned} u' &= +\cos(\pi/4)u + \sin(\pi/4)v, \\ v' &= -\sin(\pi/4)u + \cos(\pi/4)v, \end{aligned} \tag{1.38}$$

and the representation of the signal with respect to the new coordinate axes, which we denote by $f'(u', v')$, is given by

$$f'(u', v') = f\left(\cos(\pi/4)u' - \sin(\pi/4)v', \sin(\pi/4)u' + \cos(\pi/4)v'\right), \tag{1.39}$$

which we see is identical in form to the output $g(u, v)$ of system $\mathcal{L}$.

One can always define a system based on a transformation (but not necessarily the other way around). For instance, consider the Fourier transformation which relates the frequency-domain representation $F(\mu)$ of a signal to its space-domain representation $f(u)$ as follows:

$$F(\mu) = \int f(u) e^{-i2\pi\mu u}\, du. \tag{1.40}$$

Let us rewrite the same with a change in dummy variables as

$$g(u) = \int f(u') e^{-i2\pi u u'}\, du'. \tag{1.41}$$

This latter equation can be interpreted as the rule relating the output of a system $g$ to its input $f$, expressed in a particular coordinate system. Thus a transformation is employed

as the rule that relates the output of the system to the input of the system in a particular representation.

In other words, although we more often think of the Fourier transform as a transformation, it can also be interpreted as a system. This is particularly useful in physical contexts. For instance, the simple "$2f$" setup used to implement Fourier transforms optically (chapter 3) is a physical system which alters an input distribution of light in a particular way to produce an output distribution of light. This physical *system* can be characterized by (a scaled version of) equation 1.41. However, the purpose for which this system is most often employed is to compute the Fourier *transformation* of the input and present it as the output; that is, to obtain the frequency-domain representation of a signal from its space-domain representation.

In physics, the distinction between systems and transformations is often referred to as the distinction between *active* and *passive* transformations (Wolf 1979). Active transformations are produced by operators which bodily move the vectors or signals, and correspond to what we have called "systems." Passive transformations arise from a change in the basis used for the description of the space, and correspond to what we have simply called "transformations."

### 1.3.2 Basis sets and representations

A set of discretely (countably) many signals, denoted by $\{\psi_l\}$, is said to be *orthonormal* if all of its members have unit norm and are orthogonal to each other:

$$\langle \psi_l, \psi_{l'} \rangle = \delta_{ll'}. \tag{1.42}$$

A set of continuously (uncountably) many signals, denoted by $\{\Psi_\upsilon\}$, is likewise orthonormal if

$$\langle \Psi_\upsilon, \Psi_{\upsilon'} \rangle = \delta(\upsilon - \upsilon'). \tag{1.43}$$

Using the inner product definition given in equation 1.16, these conditions may be written in the time or space domain as

$$\int \psi_l^*(u)\psi_{l'}(u)\,du = \delta_{ll'}, \tag{1.44}$$

$$\int \Psi_\upsilon^*(u)\Psi_{\upsilon'}(u)\,du = \delta(\upsilon - \upsilon'). \tag{1.45}$$

The set $\{\psi_l\}$ (or $\{\Psi_\upsilon\}$) is said to constitute a *basis* for a signal space if every member of the signal space can be expanded in one and only one way in terms of the elements of this set. In the discrete case, this expansion may be expressed as

$$f = \sum_l f_\psi(l)\,\psi_l, \tag{1.46}$$

where $f_\psi(l)$ are the expansion coefficients of the signal $f$ with respect to the basis $\{\psi_l\}$. These expansion coefficients, interpreted as a function of the discrete variable $l$, constitute

the representation of the signal $f$ in the basis $\{\psi_l\}$. Alternatively, we may say that they represent the signal $f$ in this basis, or with respect to this basis. Likewise, in the continuous case,

$$f = \int f_\Psi(\upsilon)\, \Psi_\upsilon \, d\upsilon, \tag{1.47}$$

where $f_\Psi(\upsilon)$ are the expansion coefficients of the signal $f$ with respect to the basis $\{\Psi_\upsilon\}$. These coefficients constitute the representation of signal $f$ in the basis $\{\Psi_\upsilon\}$. The above equations which are written in terms of abstract signals may be specialized to a particular domain, such as the time domain:

$$f(u) = \sum_l f_\psi(l)\psi_l(u), \tag{1.48}$$

$$f(u) = \int f_\Psi(\upsilon)\Psi_\upsilon(u)\, d\upsilon. \tag{1.49}$$

If every function $f(u)$ belonging to the function space of interest can be expanded in terms of the elements of the set of functions $\{\psi_l(u)\}$ this set of functions is said to constitute a *complete* set of functions. Alternatively, it is sometimes said that the set of functions *spans* the space of interest. A linearly independent set of functions which spans the space of interest, such that the expansion not only exists but is unique, constitutes a basis for that space. Similar statements can be made for the set of functions $\{\Psi_\upsilon(u)\}$. The act of going from the $f(u)$ representation to the $f_\psi(l)$ or $f_\Psi(\upsilon)$ representation is referred to as a *transformation.*

In order to obtain the expansion coefficients for an orthonormal basis set, we take the inner product of both sides of equations 1.46 or 1.47 with a particular member of the basis set. For the discrete and continuous cases respectively,

$$\langle \psi_{l'}, f\rangle = \sum_l f_\psi(l)\langle \psi_{l'}, \psi_l\rangle = \sum_l f_\psi(l)\delta_{l'l} = f_\psi(l'), \tag{1.50}$$

$$\langle \Psi_{\upsilon'}, f\rangle = \int f_\Psi(\upsilon)\langle \Psi_{\upsilon'}, \Psi_\upsilon\rangle\, d\upsilon = \int f_\Psi(\upsilon)\delta(\upsilon' - \upsilon)\, d\upsilon = f_\Psi(\upsilon'), \tag{1.51}$$

so that

$$f_\psi(l) = \langle \psi_l, f\rangle = \int \psi_l^*(u) f(u)\, du, \tag{1.52}$$

$$f_\Psi(\upsilon) = \langle \Psi_\upsilon, f\rangle = \int \Psi_\upsilon^*(u) f(u)\, du, \tag{1.53}$$

where the rightmost forms are expressed in the time domain using the inner product definition given in equation 1.16. The orthonormality conditions given by equations 1.42 and 1.43 have been used in deriving these results.

Now, let us substitute equations 1.52 and 1.53 in equations 1.48 and 1.49 respectively and use the orthonormality relations to obtain

$$\sum_l \psi_l^*(u')\psi_l(u) = \delta(u - u'), \tag{1.54}$$

$$\int \Psi_\upsilon^*(u')\Psi_\upsilon(u)\, d\upsilon = \delta(u - u'), \tag{1.55}$$

for all $u, u'$. Treating $u$ and $u'$ as parameters, the summation and integral can be interpreted as inner products so that we can formally write the above equations in the form of inner products:

$$\langle \psi_l(u'), \psi_l(u) \rangle = \delta(u - u'), \tag{1.56}$$

$$\langle \Psi_\upsilon(u'), \Psi_\upsilon(u) \rangle = \delta(u - u'), \tag{1.57}$$

for all $u, u'$. These relations are known as *closure* or *completeness* conditions. It is worth comparing and contrasting the closure conditions with the orthonormality conditions given by equations 1.42 and 1.43. These conditions are indeed a statement of completeness of the set of functions and their constituting an orthonormal basis set. (The *braket notation* employed in quantum mechanics provides a very elegant means of expressing such relations, see Cohen-Tannoudji, Diu, and Laloë 1977.)

Let us summarize by repeating the following two key relations for discrete bases:

$$f_\psi(l) = \int \psi_l^*(u) f(u)\, du, \tag{1.58}$$

$$f(u) = \sum_l f_\psi(l) \psi_l(u). \tag{1.59}$$

The first of these equations gives the coefficient $f_\psi(l)$ appearing in the expansion of $f(u)$ given in the second equation. If this second equation is a given; that is, if we know that $f(u)$ *can* be expanded in terms of the orthonormal set $\{\psi_l(u)\}$, then the first equation for $f_\psi(l)$ is derived simply by orthonormality. The second equation expresses the somewhat more subtle fact that the projections $f_\psi(l)\psi_l(u)$ for all $l$ add up to $f(u)$ itself. This is what we mean when we say that the set $\{\psi_l(u)\}$ is a basis. (For example, in $\mathbf{R}^3$, any two of the common unit vectors $\hat{\mathbf{u}}_x$, $\hat{\mathbf{u}}_y$, $\hat{\mathbf{u}}_z$ do not constitute a basis but all three of them do.)

Let us now assume that we are given the representations $f_\psi(l)$ and $g_\psi(l)$ of two signals $f$ and $g$ in the basis $\{\psi_l\}$, or the representations $f_\Psi(\upsilon)$ and $g_\Psi(\upsilon)$ in the basis $\{\Psi_\upsilon\}$, and that we wish to calculate the inner product $\langle f, g \rangle$ directly in terms of these representations. By substituting equations 1.46 and 1.47 in equations 1.15 and 1.16 and using the orthonormality conditions we can easily show that

$$\langle f, g \rangle = \sum_l f_\psi^*(l) g_\psi(l), \tag{1.60}$$

$$\langle f, g \rangle = \int f_\Psi^*(\upsilon) g_\Psi(\upsilon)\, d\upsilon, \tag{1.61}$$

which we see are identical in form to equations 1.15 and 1.16. Thus, the expression for the inner product and hence the norm $\|f\| = \sqrt{\langle f, f \rangle}$ is independent of the particular basis set in which we represent the signal. In other words, no matter which representation of the signal we use in equations 1.15 and 1.16, we will always obtain the same result. This justifies our previous assertion that inner products and norms are properties of the signals in the abstract, and not tied to any particular representation. When it comes to actually

calculating them, we can calculate inner products and norms in any representation we find convenient.

We now turn our attention to the representation of systems with respect to particular bases. Let a linear system $\mathcal{L}$ mapping an input signal $f$ to an output $g = \mathcal{L}f$ be defined with respect to the basis set $\{\psi_l\}$ as

$$g_\psi(l) = \sum_{l'} L_\psi(l, l') f_\psi(l'), \tag{1.62}$$

where $L_\psi(l, l')$ is the representation of this linear system with respect to the basis set $\{\psi_l\}$. This equation is the most general linear relation between the representation of $g$ and the representation of $f$. To see how $L_\psi(l, l')$ can be expressed in terms of the members of the basis set $\{\psi_l\}$, let us start from the system equation in abstract form $g = \mathcal{L}f$ and substitute the expansions of $g$ and $f$ to obtain

$$\sum_l g_\psi(l)\, \psi_l = \mathcal{L}\left[\sum_{l'} f_\psi(l')\, \psi_{l'}\right] = \sum_{l'} f_\psi(l')\, \mathcal{L}\psi_{l'}. \tag{1.63}$$

Now, taking the inner product of both sides from the left with $\psi_l$ we obtain

$$g_\psi(l) = \sum_{l'} \langle \psi_l, \mathcal{L}\psi_{l'} \rangle f_\psi(l'), \tag{1.64}$$

from which we recognize

$$L_\psi(l, l') = \langle \psi_l, \mathcal{L}\psi_{l'} \rangle. \tag{1.65}$$

This expression shows how the representation of a system with respect to a particular basis is related to the abstract system operator $\mathcal{L}$ and the members of the basis set.

The trace of a system is a representation-invariant quantity defined by

$$\mathrm{Tr}[\mathcal{L}] = \sum_l L_\psi(l, l), \tag{1.66}$$

$$\mathrm{Tr}[\mathcal{L}] = \int L_\Psi(u, u)\, du, \tag{1.67}$$

for discrete and continuous bases respectively. It is easy to show that the trace is the same no matter which representation it is calculated in. That is, $\sum_l L_\psi(l, l) = \sum_l L_\phi(l, l)$ for any two discrete basis sets $\{\psi_l\}$ and $\{\phi_l\}$. It is also know that the trace is equal to the summation of the eigenvalues. Corresponding results hold for continuous bases.

We finally note that if we have a set of signals spanning a certain space, it is possible to obtain an orthonormal basis set by using a process known as *Gram-Schmidt orthogonalization* (Naylor and Sell 1982).

### 1.3.3 Impulse and harmonic bases

We will now illustrate some of the above concepts with two familiar examples. First, we consider the set of signals $\Psi_v = \delta_v$ which are defined through their representations in the

time domain as $\delta_\upsilon(u) = \delta(u - \upsilon)$. This set of signals constitutes an orthonormal basis set as they obviously satisfy the orthonormality and closure conditions:

$$\langle \delta_\upsilon, \delta_{\upsilon'} \rangle = \int \delta(u - \upsilon)\delta(u - \upsilon')\, du = \delta(\upsilon - \upsilon'), \tag{1.68}$$

$$\langle \delta_\upsilon(u), \delta_\upsilon(u') \rangle = \int \delta(u - \upsilon)\delta(u' - \upsilon)\, d\upsilon = \delta(u - u'). \tag{1.69}$$

This basis set will be referred to as the impulse basis. The expansion of a signal $f$ in this basis takes the form

$$f = \int f_\delta(\upsilon)\, \delta_\upsilon\, d\upsilon, \tag{1.70}$$

$$f_\delta(\upsilon) = \langle \delta_\upsilon, f \rangle. \tag{1.71}$$

Expressed in the time domain, these expression take the form

$$f(u) = \int f_\delta(\upsilon)\delta(u - \upsilon)\, d\upsilon, \tag{1.72}$$

$$f_\delta(\upsilon) = \int \delta(u - \upsilon) f(u)\, du, \tag{1.73}$$

from which we see that $f_\delta(\upsilon) = f(\upsilon)$. The expansion coefficients of $f$ corresponding to the impulse basis set is simply $f(\upsilon)$, the representation of the signal in the time domain. Alternatively, we may say that what we conventionally call the time-domain representation of the signal $f$ and denote by $f(u)$, is nothing but the representation of the signal in the impulse basis. Equivalently, the impulse basis $\{\delta_\upsilon\}$ is the basis set associated with what is conventionally called the time domain. Thus the time-domain representation is no different from any other representation in terms of the status accorded to it in our framework. It does not have a special place and is on an equal footing with other representations.

As a second example, we consider the set of signals $\Psi_\upsilon = \mathrm{har}_\upsilon$ which are defined to correspond in the time domain to the set of eigenfunctions $\mathrm{har}_\upsilon(u) = e^{i2\pi\upsilon u}$ of linear shift-invariant systems. This set of signals constitutes an orthonormal basis set as they satisfy the orthonormality and closure conditions:

$$\langle \mathrm{har}_\upsilon, \mathrm{har}_{\upsilon'} \rangle = \int e^{-i2\pi\upsilon u} e^{i2\pi\upsilon' u}\, du = \delta(\upsilon - \upsilon'), \tag{1.74}$$

$$\langle \mathrm{har}_\upsilon(u), \mathrm{har}_\upsilon(u') \rangle = \int e^{-i2\pi\upsilon u} e^{i2\pi\upsilon u'}\, d\upsilon = \delta(u - u'). \tag{1.75}$$

These equations are simply two different instances of equation 1.8. This basis set will be referred to as the harmonic basis. The expansion of a signal $f$ in this basis takes the form

$$f = \int f_{\mathrm{har}}(\upsilon)\, \mathrm{har}_\upsilon\, d\upsilon, \tag{1.76}$$

$$f_{\mathrm{har}}(\upsilon) = \langle \mathrm{har}_\upsilon, f \rangle. \tag{1.77}$$

Expressed in the time domain, these expression take the form

$$f(u) = \int f_{\text{har}}(\upsilon) e^{i2\pi\upsilon u}\, d\upsilon, \tag{1.78}$$

$$f_{\text{har}}(\upsilon) = \int e^{-i2\pi\upsilon u} f(u)\, du, \tag{1.79}$$

from which we see that $f_{\text{har}}(\upsilon)$ is equal to $F(\upsilon)$, the Fourier transform of $f(u)$. The expansion coefficients of $f$ corresponding to the harmonic basis set is simply $F(\upsilon)$, the representation of the signal in the frequency domain. Alternatively, we may say that what we conventionally call the frequency-domain representation of the signal $f$ and denote by $F(\mu)$, is nothing but the representation of the signal in the harmonic basis. Equivalently, the harmonic basis $\{\text{har}_\upsilon\}$ is the basis set associated with what is conventionally called the frequency domain.

Notice that members of both the impulse set and the harmonic set have infinite norms and energies; they are not square integrable. They are not physically realizable signals, but are mathematical idealizations which are found to be quite indispensable as intermediaries. We cannot physically realize impulse or harmonics functions, but we can expand physically realizable functions in terms of them. Both of them are examples of continuous bases; an example of a discrete basis set will be given in section 1.5.2.

### 1.3.4 Transformations between representations

We have seen that the coefficients appearing in the expansion of a signal in terms of an orthonormal basis set constitute the representation of the signal with respect to that basis set. We will now examine more closely the relations between different representations and transformations between them.

Let us assume that a new orthonormal basis $\{\phi_l\}$ is defined in terms of the orthonormal basis $\{\psi_l\}$ through the relation

$$\psi_l = \mathcal{T}\phi_l \qquad \text{for all } l, \tag{1.80}$$

where $\mathcal{T}$ is a linear system. To ensure that the set $\{\phi_l\}$ as defined is indeed an orthonormal basis, $\mathcal{T}$ must satisfy certain properties. A linear system $\mathcal{T}$ which maps any orthonormal basis into another orthonormal basis, is called a *unitary* system. (Conversely, a unitary system will always map an orthonormal basis into another orthonormal basis.) A unitary system always has an inverse $\mathcal{T}^{-1}$ so that we can write

$$\phi_l = \mathcal{T}^{-1}\psi_l \qquad \text{for all } l. \tag{1.81}$$

In the continuous case, a new orthonormal basis $\{\Phi_\upsilon\}$ may be defined in terms of the orthonormal basis $\{\Psi_\upsilon\}$ through

$$\Psi_\upsilon = \mathcal{T}\Phi_\upsilon \qquad \text{for all } \upsilon. \tag{1.82}$$

Let us now consider two discrete orthonormal basis sets $\{\psi_l\}$ and $\{\phi_l\}$ and consider the expansion of $f$ in terms of both of these basis sets:

$$f = \sum_{l'} f_\psi(l')\psi_{l'}, \tag{1.83}$$

$$f = \sum_{l} f_\phi(l)\phi_l. \tag{1.84}$$

We wish to find the relation between $f_\psi(l)$ and $f_\phi(l)$. One way is to expand each member of one of the sets in terms of members of the other:

$$\psi_{l'} = \sum_{l} \langle \phi_l, \psi_{l'} \rangle \phi_l, \tag{1.85}$$

and substitute this in equation 1.83 to recognize

$$f_\phi(l) = \sum_{l'} \langle \phi_l, \psi_{l'} \rangle f_\psi(l') \tag{1.86}$$

from equation 1.84. This equation allowing us to compute $f_\phi(l)$ in terms of $f_\psi(l)$ is an explicit expression of the transformation from the $\{\psi_l\}$ basis to the $\{\phi_l\}$ basis. The inner products $\langle \phi_l, \psi_{l'} \rangle = \langle \psi_{l'}, \phi_l \rangle^*$ constitute a two-dimensional array of coefficients, which we will define as the *transformation coefficients* $T(l, l')$ of the transformation from the $\{\psi_l\}$ representation to the $\{\phi_l\}$ representation: $T(l, l') \equiv \langle \phi_l, \psi_{l'} \rangle$. Using equation 1.81, these inner products can also be written as $\langle \mathcal{T}^{-1}\psi_l, \psi_{l'} \rangle = \langle \phi_l, \mathcal{T}\phi_{l'} \rangle$. Equation 1.86 can now be rewritten as

$$f_\phi(l) = \sum_{l'} T(l, l') f_\psi(l'). \tag{1.87}$$

The coefficients $T^{-1}(l, l')$ of the inverse transformation (equation 1.81) from the $\{\phi_l\}$ representation to the $\{\psi_l\}$ representation are likewise given by $T^{-1}(l, l') \equiv \langle \psi_l, \phi_{l'} \rangle = \langle \phi_{l'}, \psi_l \rangle^*$ from which we conclude that

$$T^{-1}(l, l') = T^*(l', l). \tag{1.88}$$

To obtain the array of transformation coefficients for the inverse transformation, we simply take the conjugate transpose of the array of coefficients for the forward transformation. Equation 1.88 also implies

$$\sum_{l'} T^*(l, l') T(j, l') = \delta_{lj}, \tag{1.89}$$

$$\sum_{l} T^*(l, l') T(l, j') = \delta_{l'j'}, \tag{1.90}$$

from which we see that the rows and columns of the array of coefficients are orthogonal to each other.

We have showed that if $\mathcal{T}$ is unitary; that is, if $\mathcal{T}$ maps any orthonormal basis set into another orthonormal basis set, it satisfies equation 1.88. Conversely, it is also possible to

show that an array of coefficients satisfying this equation does indeed map any orthonormal basis into another orthonormal basis. (This amounts to showing that if the basis $\{\psi_l\}$ satisfies orthonormality and closure relations, then so does the basis $\{\phi_l\}$, a task which we leave to the reader.) For this reason, equation 1.88 is often taken to be the defining property of a unitary system. Likewise, the transformation expressed by equation 1.87, with $T(l,l')$ satisfying equation 1.88, is called a unitary transformation.

For continuous orthonormal basis sets we can analogously write

$$f_\Phi(u) = \int \langle \Phi_u, \Psi_{u'} \rangle f_\Psi(u')\, du'. \tag{1.91}$$

Defining the transformation coefficients $T(u,u') \equiv \langle \Phi_u, \Psi_{u'} \rangle = \langle \Psi_{u'}, \Phi_u \rangle^*$, we can write

$$f_\Phi(u) = \int T(u,u') f_\psi(u')\, du', \tag{1.92}$$

and so on. We note that again $T^{-1}(u,u') = T^*(u',u)$. Transformations between a discrete set to a continuous set and the other way around are similarly handled.

A transformation is linear if the relation between the two representations is linear, as it is in equations 1.92 and 1.87. This implies that if $f_{j_\phi}(l) = \sum_{l'} T(l,l') f_{j_\psi}(l')$ for some sequence of signals $f_j$, then

$$\sum_{l'} T(l,l') \left[ \sum_j \alpha_j f_{j_\psi}(l') \right] = \sum_j \alpha_j f_{j_\phi}(l), \tag{1.93}$$

where $\alpha_j$ are arbitrary complex coefficients.

We now turn our attention to the transformation of the representations of systems, rather than signals, from one basis set to another. Let the output $g$ of a linear system $\mathcal{L}$ be related to the input $f$ through the relation $g = \mathcal{L}f$. This can be expressed as

$$g_\psi(l) = \sum_{l'} L_\psi(l,l') f_\psi(l'), \tag{1.94}$$

$$g_\phi(l) = \sum_{l'} L_\phi(l,l') f_\phi(l'), \tag{1.95}$$

in the $\{\psi_l\}$ and $\{\phi_l\}$ representations respectively. Our aim is to find the relation between $L_\phi(l,l')$ and $L_\psi(l,l')$. We can write two instances of equation 1.87 as

$$f_\psi(l) = \sum_{l'} T^{-1}(l,l') f_\phi(l'), \tag{1.96}$$

$$g_\phi(l) = \sum_{l'} T(l,l') g_\psi(l'), \tag{1.97}$$

and use these in equation 1.94 to obtain

$$g_\phi(l) = \sum_{l'} \sum_{l''} \sum_{l'''} T(l,l') L_\psi(l',l'') T^{-1}(l'',l''') f_\phi(l'''), \tag{1.98}$$

from which we recognize the desired result as

$$L_\phi(l,l') = \sum_{l''}\sum_{l'''} T(l,l'')L_\psi(l'',l''')T^{-1}(l''',l'), \tag{1.99}$$

or since $T^{-1}(l,l') = T^*(l',l)$

$$L_\phi(l,l') = \sum_{l''}\sum_{l'''} T(l,l'')L_\psi(l'',l''')T^*(l',l'''). \tag{1.100}$$

For the continuous case we can likewise derive

$$L_\Phi(u,u') = \iint T(u,u'')L_\Psi(u'',u''')T^*(u',u''')\,du''\,du'''. \tag{1.101}$$

An alternative derivation of the above result, which we leave to the reader, takes as a starting point the closure relation and equation 1.7.

As a simple example, consider the kernel $h(u,u') = h(u-u')$ in the time domain ($g(u) = \int h(u,u')f(u')\,du'$), which becomes the kernel $H(\mu,\mu') = H(\mu)\delta(\mu-\mu')$ in the frequency domain ($G(\mu) = \int H(\mu,\mu')F(\mu')\,d\mu'$), where $F(\mu)$, $G(\mu)$, $H(\mu)$ are the Fourier transforms of $f(u)$, $g(u)$, $h(u)$. The reader may illustrate the above general results for this special case where the Fourier transform plays the role of the unitary transformation.

## 1.4 Operators

### 1.4.1 Operators

Operators are mathematical objects that can be used to denote either systems or transformations. They are denoted by calligraphic letters such as $\mathcal{S}$ or $\mathcal{T}$. For instance, the clockwise rotation system or the counterclockwise coordinate transformation discussed on page 10 may both be denoted by the calligraphic symbol $\mathcal{ROT}_{\pi/4}$. In the case of systems, they denote a system in the abstract, without reference to any particular representation or basis set. In the case of transformations, they denote the underlying system through which the new basis set is related to the old (equation 1.80).

A linear operator is an operator denoting a linear system or transformation. In the case of systems, an operator relates an output signal $g$ to an input signal $f$ in the form $g = \mathcal{L}f$. If we choose to represent signals with respect to a particular continuous basis set $\{\Psi_u\}$ in the form $f(u)$, $g(u)$, and so forth, we may write this relation more explicitly in one of several forms

$$g(u) = \mathcal{L}f(u) \equiv (\mathcal{L}f)(u) \equiv \mathcal{L}[f](u) \equiv \{\mathcal{L}[f]\}(u). \tag{1.102}$$

We are writing simply $f(u)$, $g(u)$ instead of $f_\Psi(u)$, $g_\Psi(u)$ since only one representation is involved in this context. Also note that $u$ is a dummy variable; we could have written $v$ or some other symbol instead. The rightmost form in the above equation explicitly denotes the $\Psi$-representation of the abstract signal $g = \mathcal{L}[f]$. The form preceding it is essentially

the same but the brackets have been omitted. In the two forms preceding these, we have employed the convention that $\mathcal{L}f$ stands for $\mathcal{L}[f]$. The form $\mathcal{L}f(u)$ is interpreted as the $\Psi$-representation of the abstract signal $\mathcal{L}f$. This form also allows another interpretation. Let us define $\mathcal{L}$ such that it acts on the function $f(u)$, rather than the abstract signal $f$, to result in another function $g(u) = \mathcal{L}[f(u)]$ in the obvious manner:

$$\mathcal{L}[f(u)] \equiv (\mathcal{L}f)(u), \tag{1.103}$$

Thus the action of a system, on a function which is the representation of a signal with respect to some basis, will be defined to be, the representation (with respect to the same basis) of the output of that system when the input is the underlying signal $f$. In other words, since a signal is fully characterized by any of its representations, we may write one of these representations in place of the signal and agree that this means that the result is also expressed in the same representation. Sometimes it will be useful to denote the output explicitly:

$$g(u) = \mathcal{L}f(u) \equiv \mathcal{L}[f(u)] \equiv \mathcal{L}[f(u)](u) \equiv \{\mathcal{L}[f(u)]\}(u). \tag{1.104}$$

The form $\mathcal{L}f(u)$ is ambiguous in that it allows $\mathcal{L}$ to be interpreted both as an operator that acts on abstract signals and as an operator that acts on functions. Since both interpretations are consistent and useful, this expression will be used to denote the $\Psi$-representation of $g = \mathcal{L}f$, or equivalently, $\mathcal{L}[f(u)]$.

Likewise, an expression such as $\langle f(u), g(u)\rangle$ will simply be interpreted as the inner product of $f$ and $g$ evaluated in the $\Psi$-representation. (Of course, the inner product is always the same no matter what representation it is evaluated in.)

The representation of the system $\mathcal{L}$ with respect to this basis set will be denoted by $L(u, u')$ and the output $g(u)$ will be related to the input $f(u)$ as in equation 1.25 or the continuous version of equation 1.62:

$$g(u) = \int L(u, u')f(u')\,du'. \tag{1.105}$$

When there is possibility of confusion, we will employ explicit labels such as $f_\Psi(u)$, $L_\Psi(u, u')$, and so on, as introduced earlier.

The interpretation of $\mathcal{L}$ as a mapping from functions to functions (representations to representations) is convenient also in the case of transformations. Let us repeat the above forms for the transformation $\mathcal{T}$:

$$f_\Phi(v) = \mathcal{T}f_\Psi(u) \equiv \mathcal{T}[f_\Psi(u)] \equiv \mathcal{T}[f_\Psi(u)](v) \equiv \{\mathcal{T}[f_\Psi(u)]\}(v). \tag{1.106}$$

We have used distinct variables $u$ and $v$ for the two representations since this reminds us that the functions inhabit distinct spaces, but this is not of any deeper significance since both $u$ and $v$ are dummy variables. For a transformation between a basis set $\{\Psi_u\}$ to another basis set $\{\Phi_v\}$, the representation of the transformation $T(v, u)$ with respect

to these basis sets will be denoted by $T(v,u)$ and $f_\Phi(v)$ will be related to $f_\Psi(u)$ as in equation 1.92:

$$f_\Phi(v) = \mathcal{T} f_\Psi(u) = \int T(v,u) f_\Psi(u)\, du. \tag{1.107}$$

When there is possibility of confusion, we will employ explicit labels such as $T_{\Psi\to\Phi}(v,u)$, and so on. Analogous expressions may be written for discrete representations.

To summarize, we have now defined the effect of operators (whether they represent systems or transformations) on signals and on functions in a consistent manner. The expression $\mathcal{L}f(u)$ can be interpreted as the $\Psi$-representation of the abstract signal $\mathcal{L}f$, or the action of the operator $\mathcal{L}$ on the function $f(u)$. The expression $\mathcal{T} f_\Psi(u)$ denotes the $\Phi$-representation of $f$.

The formal similarity between systems and transformations is sometimes useful, although their physical interpretations are distinct. It is sometimes useful to think of transformations as if they were systems in mathematical manipulations. $f_\Psi(u)$ is interpreted as the input, and $f_\Phi(v)$ as the output. For instance, the system which rotates the input by $\pi/4$ in the clockwise direction is associated with the transformation which corresponds to rotation of the rectangular coordinate system by $\pi/4$ in the counterclockwise direction. The operator notation embodies the formal similarity between systems and transformations and allows them to be treated in a unified manner, so that in the course of symbolic manipulations we do not need to distinguish between systems and transformations.

The *Hermitian conjugate* (or *Hermitian transpose* or *conjugate transpose* or *adjoint*) $\mathcal{L}^{\mathrm{H}}$ of a linear operator $\mathcal{L}$ with representation $L(u,u')$ (or $L(l,l')$) is defined as the operator whose representation is $L^{\mathrm{H}}(u,u') = L^*(u',u)$ (or $L^{\mathrm{H}}(l,l') = L^*(l',l)$). Hermitian conjugation satisfies the following properties:

$$(\mathcal{L}^{\mathrm{H}})^{\mathrm{H}} = \mathcal{L}, \tag{1.108}$$

$$(\mathcal{L}_1\mathcal{L}_2\cdots\mathcal{L}_n)^{\mathrm{H}} = \mathcal{L}_n^{\mathrm{H}}\cdots\mathcal{L}_2^{\mathrm{H}}\mathcal{L}_1^{\mathrm{H}}, \tag{1.109}$$

$$\langle f, \mathcal{L}g\rangle = \langle \mathcal{L}^{\mathrm{H}} f, g\rangle, \tag{1.110}$$

$$\langle \mathcal{L}f, g\rangle = \langle f, \mathcal{L}^{\mathrm{H}} g\rangle, \tag{1.111}$$

where $f$ and $g$ are any two signals. It can be shown, for instance by using equations 1.100 or 1.101, that this definition is independent of the representation in which the conjugate transpose is taken. In fact, it is also possible to take the representation-independent equation 1.110 or equation 1.111 as the definition of Hermitian conjugation, which can be readily shown to be equivalent to the definition we have given above. (Choosing a particular basis $\{\psi_l\}$ and using equation 1.65, $L^{\mathrm{H}}(l,l') = \langle \psi_l, \mathcal{L}^{\mathrm{H}}\psi_{l'}\rangle = \langle \mathcal{L}\psi_l, \psi_{l'}\rangle = \langle \psi_{l'}, \mathcal{L}\psi_l\rangle^* = L^*(l',l)$, and similar for the continuous case.)

An operator $\mathcal{H}$ is called *Hermitian* if it is equal to its Hermitian conjugate: $\mathcal{H}^{\mathrm{H}} = \mathcal{H}$. Thus the representation of such an operator satisfies the relation $H(u,u') = H^*(u',u)$. For an operator denoting a system, being Hermitian is a property of the system in the abstract, and is not tied to a particular representation $H(u,u')$. It is easy to see from

equation 1.110 or equation 1.111 that Hermitian operators satisfy $\langle f, \mathcal{H}g\rangle = \langle \mathcal{H}f, g\rangle$. In fact, this equality can be taken as a representation-independent definition of Hermitian operators. ($H(l,l') \equiv \langle \psi_l, \mathcal{H}\psi_{l'}\rangle = \langle \mathcal{H}\psi_l, \psi_{l'}\rangle = \langle \psi_{l'}, \mathcal{H}\psi_l\rangle^* = H^*(l',l)$, and similar for the continuous case.) If two operators are Hermitian so is their sum and difference. The quantity $\langle f, \mathcal{H}f\rangle$ for arbitrary $f$ is always real as can be seen easily by writing $\langle \mathcal{H}f, f\rangle^* = \langle f, \mathcal{H}f\rangle = \langle \mathcal{H}f, f\rangle$.

An operator $\mathcal{T}$ is called *unitary* if its inverse equals its Hermitian conjugate: $\mathcal{T}^{\mathrm{H}} = \mathcal{T}^{-1}$, or equivalently $\mathcal{T}\mathcal{T}^H = \mathcal{T}^H\mathcal{T} = \mathcal{I}$ where $\mathcal{I}$ is the identity operator. Thus the representation of such an operator satisfies the relation $T^*(u,u') = T^{-1}(u',u)$. Operators denoting linear transformations (in the sense of expressing a signal with respect to a new orthonormal basis set) are always unitary, as we have seen in association with equations 1.87 and 1.92. If two operators are unitary so is their product. It is easy to verify that if $\mathcal{H}$ is Hermitian and $\mathcal{T}$ is unitary, then $\mathcal{T}^{-1}\mathcal{H}\mathcal{T}$ is also Hermitian, $\langle \mathcal{T}f, g\rangle = \langle f, \mathcal{T}^{-1}g\rangle$, $\langle \mathcal{T}f, \mathcal{T}g\rangle = \langle f, \mathcal{T}^{-1}\mathcal{T}g\rangle = \langle f, g\rangle$, and $\|\mathcal{T}f\|^2 = \|f\|^2$, for arbitrary $f$ and $g$. The latter properties mean that inner products and norms are conserved when the signals in question are acted upon by unitary operators. This property is what underlies their being interpretable as transformations from one basis to another, as we have already seen. A kernel $T(l,l')$ whose columns (or rows) constitute an orthonormal set is unitary since it can be directly shown that $\mathcal{T}^{\mathrm{H}}\mathcal{T} = \mathcal{I} = \mathcal{T}\mathcal{T}^H = \mathcal{I}$. Conversely, the columns (or rows) of a unitary kernel constitute an orthonormal set.

Particularly important properties of Hermitian and unitary operators (whether they denote systems or transformations) are those regarding their eigenfunctions, which will be discussed separately further below.

Another correspondence between systems and transformations is that between equation 1.65, which we repeat for convenience:

$$L_\psi(l,l') = \langle \psi_l, \mathcal{L}\psi_{l'}\rangle, \tag{1.112}$$

and the corresponding

$$T_{\psi\to\phi}(l,l') \equiv T(l,l') = \langle \phi_l, \mathcal{T}\phi_{l'}\rangle = \langle \psi_l, \mathcal{T}\psi_{l'}\rangle, \tag{1.113}$$

$$T_{\phi\to\psi}(l,l') \equiv T^{-1}(l,l') = \langle \psi_l, \mathcal{T}^{-1}\psi_{l'}\rangle = \langle \phi_l, \mathcal{T}^{-1}\phi_{l'}\rangle, \tag{1.114}$$

which can be derived from the definition of the transformation matrix $T(l,l') = \langle \phi_l, \psi_{l'}\rangle$ and $\psi_l = \mathcal{T}\phi_l$. Let us now write $f = \sum_l f_\phi(l)\,\phi_l$ and apply the unitary operator $\mathcal{T}$ on both sides to obtain

$$\mathcal{T}f = \sum_l f_\phi(l)\,\psi_l, \tag{1.115}$$

implying $\langle \psi_l, \mathcal{T}f\rangle = f_\phi(l)$ or

$$[\mathcal{T}f]_\psi(l) = f_\phi(l). \tag{1.116}$$

How is this last equation to be interpreted? It says that the $\psi$-representation of the signal $\mathcal{T}f$ is functionally identical in appearance to the $\phi$-representation of the signal $f$. Looking

at the same equation from the other way around, the $\phi$-representation of the signal $f$ may be found by finding the $\psi$-representation of the signal $\mathcal{T}f$. The operator $\mathcal{T}$ interpreted as a system, is related to the operator $\mathcal{T}$ interpreted as a transformation, in the same way that the rotational system and rotational transformation discussed on page 10 are related. $\mathcal{T}f$ is that signal whose $\psi$-representation looks exactly like the $\phi$-representation of the signal $f$. Thus, if we wish to find the $\phi$-representation of the signal $f$, we might obtain the $\psi$ representation of the signal $\mathcal{T}f$ instead.

To show the utility of this formalism, let us rederive two previous results. We have already shown that the inner products and norms of signals are independent of which representation they are calculated in. This is particularly easy to see by using the properties of unitary operators. Let $\mathcal{T}$ denote the unitary transformation between two representations of the signals $f$ and $g$ as $f_\Phi(u) = \mathcal{T}f_\Psi(u)$ and $g_\Phi(u) = \mathcal{T}g_\Psi(u)$. Then

$$\langle f_\Psi(u), g_\Psi(u)\rangle = \langle \mathcal{T}^{-1}f_\Phi(u), \mathcal{T}^{-1}g_\Phi(u)\rangle = \langle f_\Phi(u), \mathcal{T}\mathcal{T}^{-1}g_\Phi(u)\rangle = \langle f_\Phi(u), g_\Phi(u)\rangle, \tag{1.117}$$

proving the desired result. Now, let us consider the transformation of the representation of a linear system from one basis to another, which is also particularly transparent in operator notation. With $f_\Phi(u) = \mathcal{T}f_\Psi(u)$, $g_\Phi(u) = \mathcal{T}g_\Psi(u)$, and $g_\Psi(u) = \int L_\Psi(u,u')f_\Psi(u')\,du'$ we obtain

$$L_\Phi(u,u') = \mathcal{T}L_\Psi(u,u')\mathcal{T}^{-1}, \tag{1.118}$$

whose explicit form was derived earlier (equation 1.101).

### 1.4.2 Eigenvalue equations

Let us consider the eigenvalue equation for the linear operator $\mathcal{L}$:

$$\mathcal{L}f = \lambda f. \tag{1.119}$$

$f$ is called the eigenvector or eigensignal, and $\lambda$ the eigenvalue. The representation of an eigensignal with respect to a particular basis is referred to as an eigenfunction. To solve this abstract equation, we must first write it in a particular representation. For instance, in the discrete $\psi$-representation we have

$$\mathcal{L}f_\psi(l) = \lambda f_\psi(l), \tag{1.120}$$

or more explicitly

$$\sum_{l'} L_\psi(l,l')f_\psi(l') = \lambda f_\psi(l), \tag{1.121}$$

and in the continuous $\Psi$-representation we have

$$\mathcal{L}f_\Psi(u) = \lambda f_\Psi(u), \tag{1.122}$$

or more explicitly

$$\int L_\Psi(u,u')f_\Psi(u')\,du' = \lambda f_\Psi(u). \tag{1.123}$$

Let us consider the discrete case and let $\mathcal{T}$ represent the (unitary) transformation from the $\psi$-representation to the $\phi$-representation so that $f_\phi(u) = \mathcal{T} f_\psi(u)$ and $L_\phi(u,u') = \mathcal{T} L_\psi(u,u')\mathcal{T}^{-1}$. With these equation 1.121 becomes

$$\sum_{l'} \mathcal{T}^{-1} L_\phi(l,l')\,\mathcal{T}\mathcal{T}^{-1} f_\phi(l) = \lambda \mathcal{T}^{-1} f_\phi(l), \tag{1.124}$$

$$\sum_{l'} \mathcal{T}^{-1} L_\phi(l,l') f_\phi(l') = \lambda \mathcal{T}^{-1} f_\phi(l), \tag{1.125}$$

$$\sum_{l'} L_\phi(l,l') f_\phi(l') = \lambda f_\phi(l), \tag{1.126}$$

which is of the same form as equation 1.121. We have simply rewritten the eigenvalue equation in another representation. Clearly, if we have a solution $f_\psi(l)$ of equation 1.121 with eigenvalue $\lambda$, then $f_\phi(l)$ will be a solution of equation 1.126 *with the same eigenvalue.* Thus the eigenvalues and eigensignals of a system are properties of the system in the abstract, and are not tied to the particular representation in which we solve the eigenvalue equation.

From now on we restrict our attention to Hermitian or unitary operators. The eigenvalues of Hermitian operators are always real and the eigenvalues of unitary operators are always of unit magnitude, as can be easily verified. In general there will be several values of $\lambda$ for which a solution to the eigenvalue equation can be found. For such operators, the eigensignals corresponding to distinct eigenvalues are always orthogonal to each other, but this cannot be said for two eigensignals which share the same eigenvalue. However, it is possible to show that within the subspace spanned by all eigensignals which share the same $m$-degenerate eigenvalue, it is always possible to find $m$ linearly independent eigensignals. These $m$ linearly independent eigensignals can be orthogonalized among themselves so that for such operators, it is always possible to find an orthogonal set of eigensignals. When speaking of the eigensignals of Hermitian or unitary operators, we will always assume that the eigensignals have been chosen so that they constitute an orthonormal set. In general, it may not always be the case that this orthonormal set constitutes a basis for the space of signals we are interested in (Cohen-Tannoudji, Diu, and Laloë 1977, page 137). We will however assume that this is the case for the operators we are dealing with.

### 1.4.3 Diagonalization and spectral expansion

We will now assume that $\{\psi_l\}$ is an orthonormal set of eigensignals of the operator $\mathcal{L}$, constituting a basis for the signal space we are interested in. Eigensignal decompositions are often a convenient way of finding the output of a linear system in response to an arbitrary input. If we expand the input signal in terms of the eigensignals of the system in the form

$$f = \sum_l f_\psi(l)\,\psi_l, \tag{1.127}$$

we can easily obtain the output $g$ corresponding to this input by applying the linear system operator $\mathcal{L}$ to both sides of the above equation to obtain

$$g = \mathcal{L}f = \sum_l f_\psi(l)\, \mathcal{L}\psi_l, \tag{1.128}$$

and since $\psi_l$ is an eigensignal of $\mathcal{L}$ with eigenvalue $\lambda_l$, we have

$$g = \mathcal{L}f = \sum_l f_\psi(l)\, \lambda_l \psi_l, \tag{1.129}$$

which we can compare with $g = \sum_l g_\psi(l)\, \psi_l$ to recognize

$$g_\psi(l) = f_\psi(l)\, \lambda_l. \tag{1.130}$$

We see that the output $g$ of the system is simply a signal whose expansion coefficient is $f_\psi(l)\, \lambda_l$. The effect of a linear system on an input signal turns out to be particularly simple if we know the representation of the signal in the eigensignal basis. Let us also find the representation of $\mathcal{L}$ with respect to the eigensignal basis. We earlier showed that $L_\psi(l, l') = \langle \psi_l, \mathcal{L}\psi_{l'} \rangle$. But since $\psi_{l'}$ is an eigensignal of $\mathcal{L}$ we have

$$L_\psi(l, l') = \lambda_l \delta_{ll'}. \tag{1.131}$$

Likewise, for a continuous set of eigensignals,

$$L_\Psi(u, u') = \lambda_u \delta(u - u'). \tag{1.132}$$

We see that the representation of $\mathcal{L}$ is diagonal in the eigensignal basis. The representation of $\mathcal{L}$ in another basis will not be diagonal; the act of transforming to the eigensignal basis is thus referred to as *diagonalization* and takes the form given in equation 1.100. The transformation kernel $T_{\phi\to\psi}(l, l')$ from an arbitrary basis $\{\phi_l\}$ to the eigensignal basis $\{\psi_l\}$ is given by $\langle \psi_l, \phi_{l'} \rangle = \langle \psi_l, \mathcal{T}_{\phi\to\psi}\psi_{l'} \rangle$. Interpreted as a function of $l$ with $l'$ a parameter, we recognize this kernel as the representation of $\phi_{l'}$ in the $\{\psi_l\}$ basis. Interpreted as a transformation matrix (with infinite dimensions), we see that $T_{\phi\to\psi}(l, l')$ consists of the orthonormal eigensignals of $\mathcal{L}$ as its columns. The orthonormality of the $\psi_l$ is consistent with the unitarity of $T_{\phi\to\psi}(l, l')$. If $\mathcal{L}$ is Hermitian, a set of orthonormal eigensignals always exists, so that $\mathcal{L}$ can always be diagonalized by a unitary transformation whose columns consist of the orthonormal set of eigensignals of the Hermitian matrix.

If the eigensignals of $\mathcal{L}$ constitute an orthonormal basis for the signal space of interest, then knowing the eigensignals and eigenvalues of $\mathcal{L}$ is sufficient to completely characterize the system. Let us start by expanding the input $f$ in terms of the eigensignal basis $\{\psi_l\}$ in the form $f = \sum_l f_\psi(l)\psi_l$ where $f_\psi(l) = \langle \psi_l, f \rangle$ and write

$$g = \mathcal{L}f = \sum_l f_\psi(l)\mathcal{L}\psi_l = \sum_l \langle \psi_l, f \rangle \mathcal{L}\psi_l = \sum_l \langle \psi_l, f \rangle \lambda_l \psi_l. \tag{1.133}$$

Now, let us represent this abstract equation in any representation we find convenient to work with. For instance, in the time domain,

$$g(u) = \sum_l \langle \psi_l, f \rangle \lambda_l \psi_l(u) = \int \left[ \sum_l \lambda_l \psi_l(u) \psi_l^*(u') \right] f(u')\, du', \tag{1.134}$$

from which we can recognize the time-domain kernel $L(u, u')$ as

$$L(u, u') = \sum_l \lambda_l \psi_l(u) \psi_l^*(u'). \tag{1.135}$$

Likewise, with respect to a continuous set of eigensignals we can show that

$$L(u, u') = \int \lambda_\upsilon \psi_\upsilon(u) \psi_\upsilon^*(u')\, d\upsilon. \tag{1.136}$$

Such expansions of a kernel are known as *spectral expansions* (or *spectral decompositions* or *singular value decompositions*).

Another way of interpreting the spectral expansion is as follows. Note that the effect of $\psi_\upsilon(u)\psi_\upsilon^*(u')$ under the integral $\int du'$ is precisely to find the projection of $f(u)$ along $\psi_\upsilon(u)$. Letting $\mathcal{PR}_\upsilon$ denote the projection operators whose kernels are $\mathcal{PR}_\upsilon = \psi_\upsilon(u)\psi_\upsilon^*(u')$, we can write the spectral expansion as

$$\mathcal{L} = \int \lambda_\upsilon \mathcal{PR}_\upsilon\, d\upsilon, \tag{1.137}$$

where we have assumed that all eigenvalues are distinct. (If there are multiple eigenvalues, then we employ projection operators onto the distinct eigenspaces.) Thus each term in equation 1.134 is the projection of the signal onto one of the eigensignals, multiplied with the corresponding eigenvalue. These are added back together to obtain $g(u)$. (Also note that the projection operators satisfy $\mathcal{PR}_\upsilon \mathcal{PR}_{\upsilon'} = \mathcal{PR}_\upsilon$ if $\upsilon = \upsilon'$ and $\mathcal{PR}_\upsilon \mathcal{PR}_{\upsilon'} = 0$ if $\upsilon \neq \upsilon'$.)

As an example, let us apply the above procedure to a linear shift-invariant system $\mathcal{L}$ whose eigenfunctions we saw were $\exp(i2\pi\mu u)$. An arbitrary input $f(u)$ may be expanded in terms of these eigenfunctions as

$$f(u) = \int F(\mu) e^{i2\pi\mu u}\, d\mu, \tag{1.138}$$

$$F(\mu) = \int e^{-i2\pi\mu u} f(u)\, du. \tag{1.139}$$

The eigenfunction representation of $\mathcal{L}$ is diagonal; its kernel has the particularly simple form $L_{\text{har}}(\mu, \mu') = H(\mu)\delta(\mu - \mu')$ where $H(\mu)$ is the eigenvalue given by equation 1.36. The effect of $\mathcal{L}$ in this representation is simply expressed as $G(\mu) = \int L_{\text{har}}(\mu, \mu') F(\mu')\, d\mu' = H(\mu)F(\mu)$. This final relation is nothing but the so-called convolution property of the Fourier transform. An interesting exercise is to start from equation 1.136 and specialize it

to obtain the spectral expansion of a linear shift-invariant system with impulse response $h(u)$. With $\psi_{\text{har}}(u) = \exp(i2\pi\mu u)$ and $\lambda_{\text{har}} = H(\mu)$ we obtain

$$h(u) = \int H(\mu) e^{i2\pi\mu u} d\,\mu, \tag{1.140}$$

which is nothing but the inverse of equation 1.36. Note that the expansion coefficient $H(\mu)$, which is nothing but the frequency-domain representation of $h(u)$, is also equal to the eigenvalue associated with the eigenfunction $\exp(i2\pi\mu u)$ (page 9). If we expand the impulse response of a particular system in terms of the eigenfunctions of that system, the expansion coefficients will correspond to the eigenvalues.

As another example, let us consider a system with time-domain representation $L(u, u')$ and whose eigenfunctions constitute a discrete set $\{\psi_l\}$. (We will later see systems whose eigensignals constitute such a discrete set, most notably a system defined by the fractional Fourier transform.) The eigenfunction representation of the system will be of the form $L_\psi(l, l') = \lambda_l \delta_{ll'}$, as we now show. The unitary transformation from the time-domain representation to the eigenfunction representation is given by $T(l, u) = \langle \psi_l(u'), \delta(u'-u) \rangle = \int \psi_l^*(u')\delta(u' - u)\, du' = \psi_l^*(u)$ and exhibits one continuous and one discrete variable. The eigenfunction representation of the system is found as

$$L_\psi(l, l') = \int\int \psi_l^*(u) L(u, u') \psi_{l'}(u')\, du\, du', \tag{1.141}$$

which can be shown without difficulty to be simply equal to

$$L_\psi(l, l') = \lambda_l \delta_{ll'}. \tag{1.142}$$

### 1.4.4 Functions of operators

Integer powers of operators are simply defined as their repeated application. Thus $\mathcal{L}^2 = \mathcal{L}\mathcal{L}$, $\mathcal{L}^3 = \mathcal{L}\mathcal{L}^2$ and so on. $\mathcal{L}^{-1}$ is defined as the inverse of $\mathcal{L}$, so that this definition can be easily generalized to negative integers. If $\mathcal{L}$ is Hermitian or unitary, $\mathcal{L}^n$ is also so.

Now, Let $\Upsilon(z)$ denote a function of a complex variable whose polynomial series is defined everywhere:

$$\Upsilon(z) = \sum_{n=0}^{\infty} \Upsilon_n z^n. \tag{1.143}$$

Then, we can take

$$\Upsilon(\mathcal{L}) \equiv \sum_{n=0}^{\infty} \Upsilon_n \mathcal{L}^n. \tag{1.144}$$

as the definition of $\Upsilon(\mathcal{L})$. It can be shown that if $\lambda_l$ is an eigenvalue of $\mathcal{L}$ and $\psi_l$ the corresponding eigensignal, then $\mathcal{L}^n$ has the same eigensignal with the eigenvalue $\lambda_l^n$. This property can also be made the basis of an alternative definition. We have earlier seen that, provided they constitute a basis set for the set of signals under consideration, specifying

the eigenvalues and eigensignals of an operator are sufficient to fully characterize it. Thus we may define $\Upsilon(\mathcal{L})$ by specifying its eigensignals and eigenvalues:

$$\Upsilon(\mathcal{L})\,\psi_l = \Upsilon(\lambda_l)\,\psi_l \qquad \text{for all } l. \tag{1.145}$$

In order to calculate $\Upsilon(\mathcal{L})\,f$, we first decompose $f$ in terms of the eigensignals of $\mathcal{L}$ and use the preceding equation:

$$f = \sum_l f_\psi(l)\,\psi_l, \tag{1.146}$$

$$\Upsilon(\mathcal{L})\,f = \sum_l f_\psi(l)\Upsilon(\lambda_l)\,\psi_l. \tag{1.147}$$

Equation 1.52 allows us to write this in the time domain as

$$\Upsilon(\mathcal{L})\,f(u) = \int \Upsilon(u,u')f(u')\,du', \tag{1.148}$$

$$\text{with} \qquad \Upsilon(u,u') = \sum_l \Upsilon(\lambda_l)\psi_l(u)\psi_l^*(u'). \tag{1.149}$$

Alternatively, we may start from the diagonalized $L_\psi(l,l')$ given in equation 1.142, replace the eigenvalues $\lambda_l$ with $\Upsilon(\lambda_l)$, and then transform back to the time domain:

$$\Upsilon(u,u') = \sum_l \sum_{l'} \psi_l(u)\Upsilon(l,l')\psi_{l'}^*(u'), \tag{1.150}$$

$$\text{with} \qquad \Upsilon(l,l') = \Upsilon(\lambda_l)\delta_{ll'}, \tag{1.151}$$

which can be shown to yield equation 1.149.

If the operator $\mathcal{H}$ is Hermitian, the operator $\exp(ia\mathcal{H})$ with real $a$, is unitary. (In different applications $a$ may represent time, the axis of propagation, or the order of a parametric transform such as the fractional Fourier transform).

The derivative of an operator $\mathcal{L}$ is the operator whose representation is the common derivative of the representation of $\mathcal{L}$. Thus if the time-domain representation of $\mathcal{L}$ is $L(u,u')$, the time-domain representation of $d\mathcal{L}/da$, where $a$ is some real parameter implicit in $\mathcal{L}$, is $dL(u,u')/da$. Manipulations involving derivatives of operators and functions of operators are easily carried out by considering series expansion of the functions. One can show, for instance, that some common rules of differentiation apply to operators and their functions:

$$\frac{d}{da}e^{\mathcal{A}a} = \mathcal{A}e^{\mathcal{A}a}, \tag{1.152}$$

where $\mathcal{A}$ is assumed not to depend on $a$.

The *commutator* of two operators $\mathcal{A}$ and $\mathcal{B}$ is another operator denoted by $[\mathcal{A},\mathcal{B}]$ and defined as

$$[\mathcal{A},\mathcal{B}] \equiv \mathcal{A}\mathcal{B} - \mathcal{B}\mathcal{A}. \tag{1.153}$$

Two operators whose commutator is the zero operator are said to commute: $\mathcal{AB} = \mathcal{BA}$. It is always the case that $[\mathcal{A}, \Upsilon(\mathcal{A})] = 0$, operators commute with functions of themselves. If two operators $\mathcal{A}$ and $\mathcal{B}$ commute, then $[\mathcal{B}, \Upsilon(\mathcal{A})] = 0$ and

$$e^{\mathcal{A}} e^{\mathcal{B}} = e^{\mathcal{A}+\mathcal{B}} = e^{\mathcal{B}+\mathcal{A}} = e^{\mathcal{B}} e^{\mathcal{A}}. \tag{1.154}$$

If $[\mathcal{A}, \mathcal{B}] = \pm i\mathcal{I}$, then we have

$$[\mathcal{A}, \mathcal{B}^n] = \pm in\mathcal{B}^{n-1}, \qquad [\mathcal{A}, \Upsilon(\mathcal{B})] = \pm i\Upsilon'(\mathcal{B}), \tag{1.155}$$

where $\Upsilon'(\cdot)$ is the derivative of the function $\Upsilon(\cdot)$. The latter equality can be shown by expanding $\Upsilon(\cdot)$ into a power series. An example of two operators satisfying equation 1.155 is the coordinate multiplication and differentiation operators $\mathcal{U}$ and $\mathcal{D}$.

We conclude with some additional results applying to two operators $\mathcal{A}$ and $\mathcal{B}$ which commute ($[\mathcal{A}, \mathcal{B}] = 0$). It is possible to show that if $f$ is an eigensignal of $\mathcal{A}$ with eigenvalue $\lambda$, then $\mathcal{B}f$ is also an eigensignal of $\mathcal{A}$ with the same eigenvalue, since $\mathcal{A}(\mathcal{B}f) = \lambda(\mathcal{B}f)$. Furthermore, if $\lambda$ is a non-degenerate eigenvalue, then $\mathcal{B}f \propto f$ so that $f$ is also an eigensignal of $\mathcal{B}$. For any two commuting operators which are Hermitian or unitary, it is always possible to find a common set of orthonormal eigensignals. Thus in this case, one can find a representation in which both of these operators are diagonal. (Cohen-Tannoudji, Diu, and Laloë 1977)

## 1.5 The Fourier transform

### 1.5.1 Definition and properties

The Fourier transform(ation) $F(\mu)$ of the function $f(u)$ is defined as (Bracewell 1986)

$$F(\mu) = \int_{-\infty}^{\infty} f(u) e^{-i2\pi\mu u}\, du. \tag{1.156}$$

The function $f(u)$ can be recovered from its Fourier transform by

$$f(u) = \int_{-\infty}^{\infty} F(\mu) e^{i2\pi\mu u}\, d\mu. \tag{1.157}$$

To see this, substitute either of these equations in the other and use equation 1.8. These relations are valid for the set of finite-energy functions mentioned on page 4, as well as a more general set of functions which include pure harmonic functions, delta functions, chirp functions, and so on, whose exact nature we will not precisely define (see Dym and McKean 1972).

The Fourier transform can be interpreted either as a system or a transformation. As a system, the Fourier transform of a signal $f$ will be denoted by $F \equiv \mathcal{F}f \equiv \mathcal{F}[f]$. This relation may be expressed in the time domain as

$$F(u) = \mathcal{F}f(u) \equiv (\mathcal{F}f)(u) \equiv \mathcal{F}[f](u) \equiv \{\mathcal{F}[f]\}(u). \tag{1.158}$$

As a transformation from the time domain to the frequency domain we write

$$F(\mu) = \mathcal{F}f(u) \equiv \mathcal{F}[f(u)] \equiv \mathcal{F}[f(u)](\mu) \equiv \{\mathcal{F}[f(u)]\}(\mu) \equiv \{\mathcal{F}[f]\}(\mu) \equiv \{\mathcal{F}f\}(\mu), \tag{1.159}$$

where $\mathcal{F}[f] \equiv \mathcal{F}f$ simply stand for $F$.

The inverse Fourier transform operator will be denoted by $\mathcal{F}^{-1}$ and satisfies $\mathcal{F}^{-1}\mathcal{F} = \mathcal{F}\mathcal{F}^{-1} = \mathcal{I}$. Integer powers of the Fourier transform are defined through repeated application: $\mathcal{F}^0 = \mathcal{I}$ and $\mathcal{F}^j = \mathcal{F}\mathcal{F}^{j-1}$ for integer $j$. Properties of the Fourier transform are summarized in table 1.4 and 1.5.

| | $f(u)$ | $F(\mu)$ |
|---|---|---|
| 1. | $f(-u)$ | $F(-\mu)$ |
| 2. | $\lvert M\rvert^{-1}f(u/M)$ | $F(M\mu)$ |
| 3. | $f(u-\xi)$ | $\exp(-i2\pi\mu\xi)F(\mu)$ |
| 4. | $\exp(i2\pi\xi u)f(u)$ | $F(\mu-\xi)$ |
| 5. | $u^n f(u)$ | $(-i2\pi)^{-n}d^nF(\mu)/d\mu^n$ |
| 6. | $(i2\pi)^{-n}d^nf(u)/du^n$ | $\mu^n F(\mu)$ |
| 7. | $f^*(u)$ | $F^*(-\mu)$ |
| 8. | $f^*(-u)$ | $F^*(\mu)$ |
| 9. | $[f(u)+f(-u)]/2$ | $[F(\mu)+F(-\mu)]/2$ |
| 10. | $[f(u)-f(-u)]/2$ | $[F(\mu)-F(-\mu)]/2$ |
| 11. | $f(u)*h(u)$ | $F(\mu)H(\mu)$ |
| 12. | $f(u)h(u)$ | $F(\mu)*H(\mu)$ |
| 13. | $f(u)\star h(u)$ | $F(\mu)H^*(\mu)$ |
| 14. | $R_{ff}(u) = f(u)\star f(u)$ | $\lvert F(\mu)\rvert^2$ |

Table 1.4: Properties of the Fourier transform I. The expressions on the right are Fourier transforms of the expressions on the left. $M, \xi$ are real but $M \neq 0, \pm\infty$, and $n$ is a positive integer.

Properties 1–6 in table 1.4 can be expressed in operator notation as well (see table 1.2):

$$\mathcal{F}\mathcal{P} = \mathcal{P}\mathcal{F}, \tag{1.160}$$

$$\mathcal{F}\mathcal{M}_M = \mathcal{M}_{1/M}\mathcal{F}, \tag{1.161}$$

$$\mathcal{F}\,\mathcal{S}\mathcal{H}_\xi = \mathcal{P}\mathcal{H}_\xi\mathcal{F}, \tag{1.162}$$

$$\mathcal{F}\,\mathcal{P}\mathcal{H}_\xi = \mathcal{S}\mathcal{H}_{-\xi}\mathcal{F}, \tag{1.163}$$

$$\mathcal{F}\mathcal{U}^n = (-\mathcal{D})^n\mathcal{F}, \tag{1.164}$$

$$\mathcal{F}\mathcal{D}^n = \mathcal{U}^n\mathcal{F}. \tag{1.165}$$

Examination of the kernel $\exp(-i2\pi\mu u)$ of the Fourier transform and the kernel $\exp(i2\pi\mu u)$ of the inverse Fourier transform reveal that the Fourier transform operator is unitary. Parseval's relation (table 1.5.5) is a direct consequence of this fact.

| | | |
|---|---|---|
| 1. | $\mathcal{F}[\sum_j \alpha_j f_j] = \sum_j \alpha_j \mathcal{F} f_j$ | $\mathcal{F}[\sum_j \alpha_j f_j(u)] = \sum_j \alpha_j F_j(\mu)$ |
| 2. | $\mathcal{F}^2 = \mathcal{P}$ | $\mathcal{F}^2 f(u) = \mathcal{F} F(u) = f(-u)$ |
| 3. | $\mathcal{F}^3 = \mathcal{P}\mathcal{F} = \mathcal{F}\mathcal{P}$ | $\mathcal{F}^3 f(u) = \mathcal{F} f(-u) = \mathcal{F}^2 F(u) = F(-u)$ |
| 4. | $\mathcal{F}^4 = \mathcal{F}^0 = \mathcal{I}$ | $\mathcal{F}^4 f(u) = f(u)$ |
| 5. | $\langle f, g\rangle = \langle F, G\rangle$ | $\int f^*(u) g(u)\, du = \int F^*(\mu) G(\mu)\, d\mu$ |
| 6. | $\mathrm{En}[f] = \mathrm{En}[F]$ | $\int \lvert f(u)\rvert^2\, du = \int \lvert F(u)\rvert^2\, d\mu$ |

Table 1.5: Properties of the Fourier transform II. The same properties are expressed in abstract signal and operator form on the left and explicitly on the right, where the Fourier transform has been interpreted as a system. $\alpha_j$ are arbitrary complex constants.

Further properties may be derived from those given. For instance, if $f(u)$ is an analytic signal expressed in the form $A(u)\exp[i\phi(u)]$ where $A(u)$ and $\phi(u)$ are real functions, then (Cohen 1989)

$$\int \mu |F(\mu)|^2\, d\mu = \frac{1}{2\pi} \int \frac{d\phi(u)}{du} |f(u)|^2\, du, \tag{1.166}$$

whose proof we leave to the reader. $d\phi(u)/du$ can be interpreted as the instantaneous frequency of $f(u)$, if $f(u)$ is a narrowband signal. Two other properties concern the magnitudes of the function and its derivatives (Bracewell 1986):

$$|f(u)| \leq \int |F(\mu)|\, d\mu, \tag{1.167}$$

$$\frac{df(u)}{du} \leq 2\pi \int |\mu F(\mu)|\, d\mu. \tag{1.168}$$

The derivatives of a function provide information on its rate of change. So does the frequency spectrum. For instance, if the Fourier transform of $f(u)$ is zero for $|\mu| \geq \mu_{\max}$, and the function is bounded by $f_{\max}$ such that $|f(u)| \leq f_{\max}$ for all $u$, then $|df(u)/du| \leq \mu_{\max} f_{\max}$ for all $u$ (Papoulis 1968, page 131). This result shows the relationship between the frequency spectrum and derivatives of the function.

Common Fourier transform pairs are given in table 1.6. Pair number 6 in table 1.6 is intimately related to equation 1.6, as can be seen by writing it as

$$\mathcal{F}\left[e^{-i\pi/4} \frac{1}{|s|} e^{i\pi(u/s)^2}\right] = e^{-i\pi(s\mu)^2}, \tag{1.169}$$

where $s$ is real and considering $s \to 0$. The finite delta train appearing in pair 10 is equal to $\mathrm{rect}[u/(2N+1)\delta u]\, \delta u\, \mathrm{comb}(u/\delta u)$, whose Fourier transform is given by $(2N+1)\delta u\, \mathrm{sinc}[(2N+1)\delta u\, \mu] * \mathrm{comb}(\delta u\, \mu)$. Thus the right hand side of pair 10 can also be written as

$$\frac{\sin[\pi(2N+1)\delta u\, \mu]}{\sin[\pi \delta u\, \mu]} = \sum_{n=-\infty}^{\infty} (2N+1)\delta u\, \mathrm{sinc}\left[(2N+1)\delta u(\mu - n/\delta u)\right], \tag{1.170}$$

| | $f(u)$ | $F(\mu)$ |
|---|---|---|
| 1. | $\delta(u-\xi)$ | $\mathrm{har}(-\xi\mu) = \exp(-i2\pi\xi\mu)$ |
| 2. | $\mathrm{har}(\xi u) = \exp(i2\pi\xi u)$ | $\delta(\mu-\xi)$ |
| 3. | $\mathrm{rect}(u)$ | $\mathrm{sinc}(\mu)$ |
| 4. | $\mathrm{sinc}(u)$ | $\mathrm{rect}(\mu)$ |
| 5. | $\mathrm{gauss}(u) = \exp(-\pi u^2)$ | $\mathrm{gauss}(\mu) = \exp(-\pi\mu^2)$ |
| 6. | $e^{\mp i\pi/4}\exp(\pm i\pi u^2)$ | $\exp(\mp i\pi\mu^2)$ |
| 7. | $\exp[i\pi(\chi u^2 + 2\xi u)]$ | $(e^{i\pi/4}/\sqrt{\chi}\,)\exp[-i\pi(u-\xi)^2/\chi]$ |
| 8. | $\mathrm{comb}(u) = \sum_{n=-\infty}^{\infty}\delta(u-n)$ | $\mathrm{comb}(\mu) = \sum_{n=-\infty}^{\infty}\delta(\mu-n)$ |
| 9. | $i/\pi u$ | $\mathrm{sgn}(\mu)$ |
| 10. | $\sum_{n=-N}^{N}\delta(u-n\delta u)$ | $\frac{\sin[\pi(2N+1)\delta u\,\mu]}{\sin[\pi\delta u\,\mu]}$ |

Table 1.6: Some common Fourier pairs. $\xi, \chi, \delta u$ are real, $N$ is a positive integer.

showing that it consists of periodically replicated sinc functions. This function is illustrated in figure 1.1.

We might also mention that the Fourier transform of $P_n(u)\exp(-\pi u^2)$ where $P_n(u)$ is a polynomial of degree $n$ is always of the same form $R_n(\mu)\exp(-\pi\mu^2)$ where $R_n(\mu)$ is another polynomial of $n$th degree. To prove this, one first writes the polynomial as a series and then takes the Fourier transform term by term noting that the Fourier transform of $u^n f(u)$ corresponds to the $n$th derivative of $F(\mu)$. Of course, the $n$th order derivative of $\exp(-\pi u^2)$ is simply a polynomial of that order times $\exp(-\pi u^2)$, which upon collecting terms results in the form $R_n(u)\exp(-\pi u^2)$. The Hermite polynomials to be introduced below are special in that they reproduce themselves: $R_n(u) \propto P_n(u)$.

Equation 1.157 is a linear superposition of functions of the form $\exp(i2\pi\mu u)$. Thus the Fourier transform $F(\mu)$ is essentially the expansion coefficient when we expand the function $f(u)$ in terms of these functions. Equation 1.156 which shows us how to calculate the expansion coefficients is essentially an inner product between the function $f(u)$ and the basis functions $\exp(i2\pi\mu u)$. Thus equations 1.156 and 1.157 are a special case of equations 1.49 and 1.53.

This expansion is of special interest when we are dealing with linear shift-invariant systems. Since complex exponential functions are the eigenfunctions of such systems, application of such a system with impulse response $h(u)$ on both sides of equation 1.157 gives

$$\mathcal{L}f(u) = \int H(\mu)F(\mu)e^{i2\pi\mu u}\,d\mu. \tag{1.171}$$

The right hand side is simply the function whose Fourier transform is $H(\mu)F(\mu)$, which we know is $f(u) * h(u)$. This is nothing but a derivation of the convolution property given in table 1.4.11. The process of calculating the convolution $f(u) * h(u)$ by multiplying its Fourier transform with a Fourier domain filter function $H(\mu)$ is referred to as *multiplicative*

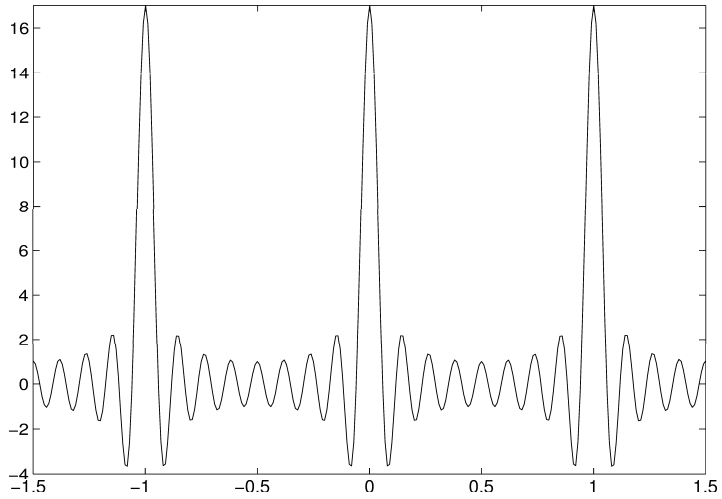


Figure 1.1: The function given in equation 1.170 with $N = 8$ and $\delta u = 1$.

*filtering in the Fourier domain.*

An interesting consequence of equation 1.9 is

$$\sum_{n=-\infty}^{\infty} f(u + n\delta u) = \frac{1}{\delta u} \sum_{n=-\infty}^{\infty} F(n/\delta u) e^{i2\pi nu/\delta u}. \tag{1.172}$$

This equation, as well as equation 1.9, are referred to as Poisson's sum(mation) formulas.

The discrete Fourier transform (DFT) is a mapping from $\mathbf{R}^N$ to $\mathbf{R}^N$. If we let $f(l)$ denote the $l$th component of a vector in $\mathbf{R}^N$, then its discrete Fourier transform is defined by

$$F(j) \equiv \frac{1}{\sqrt{N}} \sum_{l=0}^{N-1} W(j,l) f(l), \qquad j = 0, 1, \ldots, N-1, \tag{1.173}$$
$$W(j,l) \equiv W^{jl}, \qquad W \equiv e^{-i2\pi/N},$$

where $F(j)$ denotes the $j$th component of the discrete Fourier transform of $f(l)$. The DFT can be computed in $O(N \log N)$ time on a serial computer, and $O(\log N)$ time on a parallel computer by using the fast Fourier transform (FFT) algorithm (Bracewell 1986, Iizuka 1987, Oppenheim and Shafer 1989). We will discuss the relationship between the continuous and discrete Fourier transforms in chapter 2.

### 1.5.2 Eigenfunctions of the Fourier transform

In this section we will examine the eigenfunctions of the Fourier transform, interpreted either as a system or a transformation. That is, we are looking for solutions of equation 1.119 for the Fourier transform operator. The solutions are denoted by $\psi_n(u)$ for

$n = 0, 1, 2, \ldots$ and are known as the Hermite-Gaussian functions:

$$\mathcal{F}\psi_n(u) = e^{-in\pi/2}\psi_n(u). \tag{1.174}$$

These functions may be written explicitly as

$$\psi_n(u) = A_n H_n(\sqrt{2\pi}\,u)\, e^{-\pi u^2}, \qquad A_n = \frac{2^{1/4}}{\sqrt{2^n n!}}, \tag{1.175}$$

where $H_n(u)$ denotes the $n$th order Hermite polynomial. $H_n(u)$ is an $n$th degree polynomial and is an odd function when $n$ is odd and is an even function when $n$ is even. $H_n(u)$ has $n$ real zeroes between which are interposed the $n-1$ zeroes of $H_{n-1}(u)$. Properties of the Hermite polynomials are given in table 1.7. Using property 8, we can see that the Hermite-Gaussian functions $\psi_n(u)$ constitute an orthonormal set. An excellent introduction to Hermite polynomials and Hermite-Gaussian functions may be found in Cohen-Tannoudji, Diu, and Laloë 1977 and Wolf 1979, the former being of a more elementary nature.

| | |
|---|---|
| 1. | $H_0(u) = 1,\ H_1(u) = 2u,\ H_2(u) = 4u^2 - 2,\ H_3(u) = 8u^3 - 12u, \ldots$ |
| 2. | $H_n(u) = (-1)^n e^{u^2} \frac{d^n}{du^n} e^{-u^2}$ |
| 3. | $e^{-\xi^2+2\xi u} = \sum_{n=0}^{\infty} \frac{\xi^n}{n!} H_n(u) \qquad H_n(u) = \left[\frac{\partial^n}{\partial \xi^n} e^{-\xi^2+2\xi u}\right]_{\xi=0}$ |
| 4. | $\frac{dH_n(u)}{du} = 2nH_{n-1}(u)$ |
| 5. | $H_{n+1}(u) = 2uH_n(u) - 2nH_{n-1}(u)$ |
| 6. | $H_n(u) = (2u - \frac{d}{du})H_{n-1}(u)$ |
| 7. | $(\frac{d^2}{du^2} - 2u\frac{d}{du} + 2n)H_n(u) = 0$ |
| 8. | $\int e^{-u^2} H_n(u) H_{n'}(u)\, du = 2^n n! \sqrt{\pi}\, \delta_{nn'}$ |
| 9. | $\int e^{-(u-u')^2} H_n(u')\, du' = 2^n \pi^{1/2} u^n$ |
| 10. | $\lim_{n\to\infty} \frac{(-1)^n\sqrt{n}}{4^n n!} H_{2n}(u/2\sqrt{n}\,) = \frac{1}{\sqrt{\pi}} \cos(u)$ |
| 11. | $\lim_{n\to\infty} \frac{(-1)^n}{4^n n!} H_{2n+1}(u/2\sqrt{n}\,) = \frac{2}{\sqrt{\pi}} \sin(u)$ |
| 12. | $\sum_{j=0}^{\infty} \frac{1}{2^{n/2}} \binom{n}{j} H_j(\sqrt{2}\,u) H_{n-j}(\sqrt{2}\,u') = H_n(u+u')$ |
| 13. | $\sum_{j=0}^{n} \frac{H_j(u)H_j(u')}{2^j j!} = \frac{H_{n+1}(u)H_n(u') - H_n(u)H_{n+1}(u')}{2^{n+1} n!(u-u')}$ |
| 14. | $\pi^{-1/2} e^{-(u^2+u'^2)/2} \sum_{n=0}^{\infty} \frac{1}{2^n n!} H_n(u) H_n(u') = \delta(u-u')$ |
| 15. | $\sum_{n=0}^{\infty} \frac{e^{in\alpha}}{2^n n!} H_n(u) H_n(u') = (1 - e^{2i\alpha})^{-1/2} \exp\left[\frac{2uu' e^{i\alpha} - e^{2i\alpha}(u^2+u'^2)}{1-e^{2i\alpha}}\right]$ |
| 16. | $\sum_{n=0}^{\infty} \frac{z^n}{2^n n!} H_n(u) H_n(u') = (1-z^2)^{-1/2} \exp\left[u'^2 - \frac{(u'-zu)^2}{1-z^2}\right]$ |

Table 1.7: Properties of Hermite polynomials. $\xi, \alpha$ are real, $z$ is complex.

The Hermite-Gaussian functions are often recognized as the solutions of the following differential equation

$$\frac{d^2 f(u)}{du^2} + 4\pi^2 \left(\frac{2n+1}{2\pi} - u^2\right) f(u) = 0, \tag{1.176}$$

associated with the quantum-mechanical harmonic oscillator or propagation in quadratic graded-index media. For the moment, we treat $(2n+1)/2\pi$ as a single constant. By taking the Fourier transform of this equation, and using elementary identities regarding the transforms of $d^2f(u)/du^2$ and $(-i2\pi u)^2 f(u)$, we can show

$$\frac{d^2F(\mu)}{d\mu^2} + 4\pi^2\left(\frac{2n+1}{2\pi} - \mu^2\right)F(\mu) = 0. \tag{1.177}$$

Since this equation is identical to the previous one in form, it is easy to accept the well known fact that solutions of this equation, the Hermite-Gaussian functions, are eigenfunctions of the Fourier transform operation (Wiener 1933, Dym and McKean 1972). It is also possible to show directly that the Hermite-Gaussian functions are indeed eigenfunctions of the Fourier transform by directly substituting them in the definition of the Fourier transform.

It is also not difficult to see that the Hermite-Gaussian functions $\psi_n(u) \propto H_n(\sqrt{2\pi}\,u)\exp(-\pi u^2)$ are indeed solutions of the above equation. When we substitute them, we obtain

$$\frac{d^2H_n(u)}{du^2} + (-2u)\frac{dH_n(u)}{du} + (2n)H_n(u) = 0, \tag{1.178}$$

which is nothing but item 7 in table 1.7. (This equation is sometimes taken as the defining equation of the Hermite polynomials.)

The Hermite-Gaussian functions constitute an orthonormal basis for the set of finite-energy functions:

$$\langle\psi_n, \psi_{n'}\rangle = \int \psi_n(u)\psi_{n'}(u)\,du = \delta_{nn'}, \tag{1.179}$$

$$\sum_{n=0}^{\infty} \psi_n(u)\psi_n(u') = \delta(u-u'). \tag{1.180}$$

Thus any finite-energy signal can be expanded in the form

$$f = \sum_{n=0}^{\infty} f_\psi(n)\,\psi_n, \tag{1.181}$$

$$f_\psi(n) = \langle\psi_n, f\rangle = \int \psi_n(u)f(u)\,du. \tag{1.182}$$

The transformation from $f(u)$ to $f_\psi(n)$ is unitary. In the $\psi$-representation, the Fourier transform has a particularly simple form. The representation of the Fourier transform of $f_\psi(n)$ is simply $\exp(-in\pi/2)f_\psi(n)$; that is, the kernel for Fourier transformation is diagonal in this representation.

The coefficients $f_\psi(n)$ constitute the representation of the signal $f$ in the Hermite-Gaussian basis set. The act of obtaining these coefficients from the time-domain representation of the signal $f(u)$ is an example of a transformation from a continuous to a discrete representation. Clearly, as with any orthonormal basis expansion,

$$\sum_{n=0}^{\infty} |f_\psi(n)|^2 = \int |f(u)|^2\,du. \tag{1.183}$$

Other properties of the Hermite-Gaussian functions are given in table 1.8.

| | |
|---|---|
| 1. | $\psi_n(u) = \frac{(-1)^n A_n}{(\sqrt{2\pi})^n} e^{\pi u^2} \frac{d^n}{du^n} e^{-2\pi u^2}$ |
| 2. | $\psi_n(u) = A_n \left[ \frac{\partial^n}{\partial \xi^n} e^{-\xi^2} e^{u(2\xi - \pi u)} \right]_{\xi=0}$ |
| 3. | $(\sqrt{2\pi}\, u + \frac{1}{\sqrt{2\pi}} \frac{d}{du}) \psi_n(u) = 2n \frac{A_n}{A_{n-1}} \psi_{n-1}(u)$ |
| 4. | $\psi_n(u) = 2\sqrt{2\pi}\, u \frac{A_n}{A_{n-1}} \psi_{n-1}(u) - 2(n-1) \frac{A_n}{A_{n-2}} \psi_{n-2}(u)$ |
| 5. | $\psi_n(u) = \frac{A_n}{A_{n-1}} (\sqrt{2\pi}\, u - \frac{1}{\sqrt{2\pi}} \frac{d}{du}) \psi_{n-1}(u)$ |
| 6. | $[\frac{d^2}{du^2} + 4\pi^2 (\frac{2n+1}{2\pi} - u^2)] \psi_n(u) = 0$ |
| 7. | $\sum_{n=0}^{\infty} \psi_n(u) \psi_n(u') = \delta(u - u')$ |
| 8. | $\sum_{n=0}^{\infty} e^{-in\pi/2} \psi_n(u) \psi_n(u') = e^{-i2\pi uu'}$ |
| 9. | $\sum_{n=0}^{\infty} e^{-in\alpha} \psi_n(u) \psi_n(u') = \sqrt{1 - i \cot \alpha}\, e^{i\pi(\cot \alpha\, u^2 - 2 \csc \alpha\, uu' + \cot \alpha\, u'^2)}$ |

Table 1.8: Properties of Hermite-Gaussian functions. $\xi, \alpha$ are real.

Another interesting property is that "most" of the energy of the $n$th order Hermite-Gaussian function is concentrated between the bounds $-\sqrt{(n+1/2)/\pi}$ and $\sqrt{(n+1/2)/\pi}$. It is within these bounds that the function shows oscillatory behavior. Outside of these bounds, the exponential factor dominates and its value decays quickly. (See Ozaktas and Mendlovic 1993b for a physical justification.) The first 4 Hermite-Gaussian functions are shown in figure 1.2.

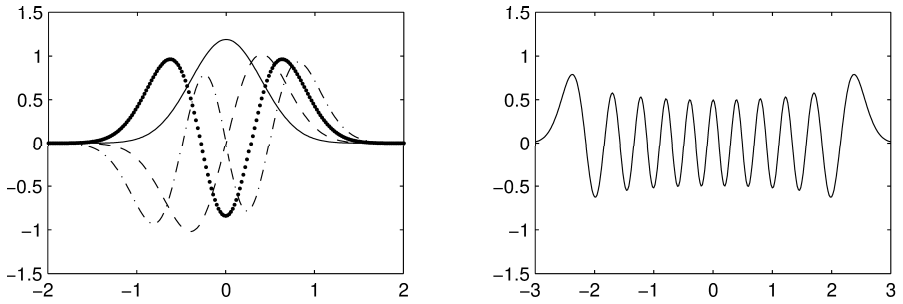


Figure 1.2: Hermite-Gaussian functions of order (left panel) $n = 0$ (solid), $n = 1$ (dashed), $n = 2$ (dotted), $n = 3$ (dot-dashed), and (right panel) $n = 20$.

As a final comment on the Hermite-Gaussian functions, we note that if one starts with the set of functions $u^n \exp(-\pi u^2)$, for $n = 0, 1, 2, \ldots$ and uses the Gram-Schmidt orthonormalization process to construct an orthonormal set of functions $Q_n(u)$, what one obtains is precisely the Hermite-Gaussian functions; that is $Q_n(u) = \psi_n(u)$ (Wiener 1933, page 54).

The Fourier transform of any function can be obtained by first expanding it in terms of the Hermite-Gaussian functions. Upon application of the eigenvalue equation, and

substituting for $f_\psi(n)$ we can show that

$$e^{-i2\pi\mu u} = \sum_{n=0}^{\infty} e^{-in\pi/2}\psi_n(\mu)\psi_n(u). \tag{1.184}$$

This is called the spectral expansion of the kernel of the Fourier transformation.

The eigenfunctions of the Fourier transform have received special interest under the name "self-Fourier functions" or "self-reciprocal Fourier functions" (Caola 1991; Cincotti, Gori, and Santarsiero 1992; Lohmann and Mendlovic 1992a, b, 1994a; Lakhtakia 1993; Lipson 1993; Coffey 1994; Choudhury, Puntambekar, and Chakraborty 1995).

## 1.6 Some important operators

### 1.6.1 Coordinate multiplication and differentiation operators

We will now investigate the operators $\mathcal{U}$ and $\mathcal{D}$ which may be defined by specifying their effect in the time domain (Cohen-Tannoudji, Diu, and Laloë 1977, page 149):

$$g(u) = (\mathcal{U}f)(u) = uf(u), \tag{1.185}$$

$$g(u) = (\mathcal{D}f)(u) = (i2\pi)^{-1}\frac{df(u)}{du}, \tag{1.186}$$

corresponding to the kernels

$$h_{\mathcal{U}}(u,u') = u\delta(u-u'), \tag{1.187}$$

$$h_{\mathcal{D}}(u,u') = (i2\pi)^{-1}\delta'(u-u'). \tag{1.188}$$

These operators are duals in the sense that their effects in the frequency domain are

$$G(\mu) = (\mathcal{U}f)(\mu) = (-i2\pi)^{-1}\frac{dF(\mu)}{d\mu}, \tag{1.189}$$

$$G(\mu) = (\mathcal{D}f)(\mu) = \mu F(\mu), \tag{1.190}$$

corresponding to the kernels

$$H_{\mathcal{U}}(\mu,\mu') = (-i2\pi)^{-1}\delta'(\mu-\mu'), \tag{1.191}$$

$$H_{\mathcal{D}}(\mu,\mu') = \mu\delta(\mu-\mu'). \tag{1.192}$$

If we had defined a frequency multiplication operator in the frequency domain, this would be identical to the derivative operator defined in the time domain. Likewise if we had defined a derivative operator in the frequency domain, this would be identical to the coordinate multiplication operator defined in the time domain.

The kernels in the frequency domain are related to the kernels in the time domain through

$$H(\mu,\mu') = \iint e^{-i2\pi\mu u}h(u,u')e^{i2\pi\mu' u'}\,du\,du', \tag{1.193}$$

which is a special case of equation 1.101.

The impulse signals $\delta_\upsilon$ defined earlier are the eigensignals of $\mathcal{U}$. This is most easily seen in the time domain:

$$(\mathcal{U}\delta_\upsilon)(u) = u\delta(u-\upsilon) = \lambda\delta(u-\upsilon). \tag{1.194}$$

with the eigenvalue $\lambda = \upsilon$. The set of eigensignals of $\mathcal{U}$, namely the impulse set, was earlier shown to constitute an orthonormal basis. It is also instructive to write the eigensignal equation in the frequency domain:

$$(\mathcal{U}\delta_\upsilon)(\mu) = \frac{1}{-i2\pi}\frac{d}{d\mu}\, e^{-i2\pi\upsilon\mu} = \lambda e^{-i2\pi\upsilon\mu}. \tag{1.195}$$

with $\lambda = \upsilon$.

Likewise, the harmonic signals $\mathrm{har}_\upsilon$ are the eigensignals of $\mathcal{D}$. This is most easily seen in the frequency domain:

$$(\mathcal{D}\mathrm{har}_\upsilon)(\mu) = \mu\delta(\mu-\upsilon) = \lambda\delta(\mu-\upsilon). \tag{1.196}$$

with the eigenvalue $\lambda = \upsilon$. The set of eigensignals of $\mathcal{D}$, namely the harmonic set, was earlier shown to constitute an orthonormal basis. In the time domain:

$$(\mathcal{D}\mathrm{har}_\upsilon)(u) = \frac{1}{i2\pi}\frac{d}{du}\, e^{i2\pi\upsilon u} = \lambda e^{i2\pi\upsilon u}. \tag{1.197}$$

with $\lambda = \upsilon$.

Since we know that the time-domain representation and the frequency-domain representation are associated with the impulse set and the harmonic set respectively, we are not surprised by the fact that the $\mathcal{U}$ and $\mathcal{D}$ operators have such simple expressions in these domains.

It is easy to show that the $\mathcal{U}$ and $\mathcal{D}$ operators are Hermitian by examining their kernels. Their commutator is given by

$$[\mathcal{U},\mathcal{D}] = \frac{i}{2\pi}\mathcal{I}. \tag{1.198}$$

Some properties of Hermitian operators satisfying this relation can be found in Cohen-Tannoudji, Diu, and Laloë 1977 (also see equation 1.155 and the following paragraphs).

We will now consider functions of $\mathcal{U}$ and $\mathcal{D}$. The effect of a function $\Upsilon(\mathcal{U})$ on a signal $f$ can be found easily by considering the power series of $\Upsilon(\cdot)$. Thus

$$g = \Upsilon(\mathcal{U})\, f = \sum_{n=0}^{\infty} \Upsilon_n \mathcal{U}^n f. \tag{1.199}$$

For instance, in the time domain,

$$g(u) = \left[\sum_{n=0}^{\infty} \Upsilon_n u^n\right] f(u) = \Upsilon(u) f(u), \tag{1.200}$$

where we have used $(\mathcal{U}^n f)(u) = u^n f(u)$, which can be easily derived by considering repeated application of equation 1.185. The kernel corresponding to $\Upsilon(\mathcal{U})$ in the time domain is likewise easily shown to be given by

$$h_{\Upsilon(\mathcal{U})}(u, u') = \Upsilon(u)\delta(u - u'). \tag{1.201}$$

Let us also consider the effect of $\Upsilon(\mathcal{U})$ in the frequency domain.

$$G(\mu) = \left[\sum_{n=0}^{\infty} \Upsilon_n (-i2\pi)^{-n} \frac{d^n}{d\mu^n}\right] F(\mu) = \Upsilon\left[(-i2\pi)^{-1}\frac{d}{d\mu}\right] F(\mu). \tag{1.202}$$

The corresponding kernel is difficult to derive from this result, but can be obtained from a common property of the Fourier transform. Since we know that $G(\mu) = (\mathcal{F}\Upsilon)(\mu) * F(\mu)$, we have

$$H_{\Upsilon(\mathcal{U})}(\mu, \mu') = (\mathcal{F}\Upsilon)(\mu - \mu'). \tag{1.203}$$

The effect of $\Upsilon(\mathcal{D})$ can be found similarly. In the frequency domain,

$$G(\mu) = \left[\sum_{n=0}^{\infty} \Upsilon_n \mu^n\right] F(\mu) = \Upsilon(\mu)F(\mu), \tag{1.204}$$

with the corresponding kernel

$$H_{\Upsilon(\mathcal{D})}(\mu, \mu') = \Upsilon(\mu)\delta(\mu - \mu'). \tag{1.205}$$

In the time domain

$$g(u) = \left[\sum_{n=0}^{\infty} \Upsilon_n (i2\pi)^{-n} \frac{d^n}{du^n}\right] f(u) = \Upsilon\left[(i2\pi)^{-1}\frac{d}{du}\right] f(u), \tag{1.206}$$

where we have used $(\mathcal{D}^n f)(u) = (i2\pi)^{-n} d^n f(u)/du^n$, which can be easily derived by considering repeated application of equation 1.186. The corresponding kernel is difficult to derive from this result, but can be obtained from a common property of the Fourier transform. Since we know that $g(u) = (\mathcal{F}^{-1}\Upsilon)(u) * f(u)$, we have

$$h_{\Upsilon(\mathcal{D})}(u, u') = (\mathcal{F}^{-1}\Upsilon)(u - u'). \tag{1.207}$$

It is not difficult to show that $\Upsilon(\mathcal{U})$ and $\Upsilon(\mathcal{D})$ have the same eigensignals as $\mathcal{U}$ and $\mathcal{D}$ respectively:

$$\Upsilon(\mathcal{U})\delta_\upsilon = \Upsilon(\upsilon)\,\delta_\upsilon, \tag{1.208}$$

$$\Upsilon(\mathcal{D})\mathrm{har}_\upsilon = \Upsilon(\upsilon)\,\mathrm{har}_\upsilon. \tag{1.209}$$

The various kernels are summarized in table 1.9.

In this context it is also worth discussing the so-called moment theorem, which is closely related to the above considerations (Papoulis 1977). It can be stated in various

| | $\mathcal{U}$ | $h(\mathcal{U})$ | $\mathcal{D}$ | $H(\mathcal{D})$ |
|---|---|---|---|---|
| $u$ domain | $u\delta(u-u')$ | $h(u)\delta(u-u')$ | $(i2\pi)^{-1}\delta'(u-u')$ | $h(u-u')$ |
| $\mu$ domain | $(-i2\pi)^{-1}\delta'(\mu-\mu')$ | $H(\mu-\mu')$ | $\mu\delta(\mu-\mu')$ | $H(\mu)\delta(\mu-\mu')$ |

Table 1.9: Kernels associated with the $\mathcal{U}$ and $\mathcal{D}$ operators and their functions.

forms and derived directly from the properties of the Fourier transform without resorting to operator concepts. Consider the expansion of $F(\mu)$:

$$F(\mu) = \sum_{n=0}^{\infty} \frac{\mu^n}{n!} \left. \frac{d^n F(\mu)}{d\mu^n} \right|_{\mu=0}. \tag{1.210}$$

Using

$$m_f^n \equiv \int u^n f(u)\, du = \frac{1}{(-i2\pi)^n} \left. \frac{d^n F(\mu)}{d\mu^n} \right|_{\mu=0}, \tag{1.211}$$

where $m_f^n$ denotes the $n$th moment of $f(u)$, we obtain the moment theorem:

$$F(\mu) = \sum_{n=0}^{\infty} \frac{(-i2\pi\mu)^n}{n!} m_f^n. \tag{1.212}$$

This result can also be derived by starting from

$$F(\mu) = \int f(u) e^{-i2\pi\mu u}\, du, \tag{1.213}$$

and replacing $\exp(-i2\pi\mu u)$ with its series expansion $\sum_{n=0}^{\infty}(-i2\pi\mu u)^n/n!$. The dual of the moment theorem is

$$f(u) = \sum_{n=0}^{\infty} \frac{(i2\pi u)^n}{n!} m_F^n, \tag{1.214}$$

where $m_F^n$ denotes the $n$th moment of $F(\mu)$.

As an additional exercise, consider the shift-invariant system characterized in the time and frequency domains respectively as

$$g(u) = h(u) * f(u) = \int h(u-u') f(u')\, du', \tag{1.215}$$

$$G(\mu) = H(\mu) F(\mu). \tag{1.216}$$

Let us consider the series expansion of $H(\mu)$

$$H(\mu) = \sum_{n=0}^{\infty} m_h^n \frac{(-i2\pi\mu)^n}{n!} \tag{1.217}$$

so that

$$G(\mu) = \sum_{n=0}^{\infty} m_h^n \frac{(-i2\pi\mu)^n}{n!} F(\mu), \tag{1.218}$$

$$g(u) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} m_h^n \frac{d^n f(u)}{du^n}, \tag{1.219}$$

an expression which, though not obvious from its appearance, is equivalent to the convolution $g(u) = h(u) * f(u)$.

An alternative approach will be instructive. In table 1.9, the kernel associated with the operator $H(\mathcal{D})$ is denoted by $h(u-u')$ in the $u$ domain. But we can also expand the series of $H(\mathcal{D})$ to obtain

$$\left(H(\mathcal{D})\right)(u,u') = \sum_{n=0}^{\infty} K_n (i2\pi)^{-n} \delta^{(n)}(u-u') \tag{1.220}$$

where $K_n$ is the series coefficient given by

$$K_n = \frac{1}{n!} \left.\frac{d^n H(\mu)}{d\mu^n}\right|_{\mu=0} = \frac{1}{n!}(-i2\pi)^n m_h^n, \tag{1.221}$$

so that

$$\left(H(\mathcal{D})\right)(u,u') = h(u-u') = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} m_h^n \delta^{(n)}(u-u'), \tag{1.222}$$

$$h(u) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} m_h^n \delta^{(n)}(u), \tag{1.223}$$

the latter which, when convolved with $f(u)$, results in equation 1.219.

### 1.6.2 Phase shift, translation, chirp multiplication and convolution operators

We now consider the phase shift operator $\mathcal{PH}_\xi = \exp(i2\pi\xi\mathcal{U})$ and translation operator $\mathcal{SH}_\xi = \exp(i2\pi\xi\mathcal{D})$ (Cohen-Tannoudji, Diu, and Laloë 1977, page 187). Their effect in the time domain can be shown to be

$$(\mathcal{PH}_\xi f)(u) = e^{i2\pi\xi u} f(u), \tag{1.224}$$

$$(\mathcal{SH}_\xi f)(u) = \exp\left[i2\pi\xi(i2\pi)^{-1} d/du\right] f(u) = f(u+\xi). \tag{1.225}$$

The last equality can be shown by expanding $\exp(\xi d/du)$ in a series, applying it to $f(u)$, and comparing it to the series of $f(u+\xi)$ with respect to the variable $\xi$. Since this is important, it is worth going through carefully:

$$e^{\xi d/du} f(u) = \sum_{n=0}^{\infty} \frac{\xi^n}{n!} \frac{d^n f(u)}{du^n} = f(u+\xi), \tag{1.226}$$

where the last equality follows from the fact that the summation is simply the series expansion of $f(u+\xi)$ around $u$. Another method is to consider $\exp(\xi d/du) f(u)$ as a function $k(u,\xi)$ of two variables. Now, differentiate $k(u,\xi)$ with respect to $\xi$. Since $\partial \exp(\xi d/du)/\partial\xi = (d/du)\exp(\xi d/du)$, we have

$$\frac{\partial k(u,\xi)}{\partial \xi} = \frac{\partial k(u,\xi)}{\partial u}. \tag{1.227}$$

Noting that $k(u,0) = f(u)$, we find the solution of this equation as $k(u,\xi) = f(u+\xi)$, which is again the desired result. However, perhaps the most direct way of seeing this result is to note that the effect of the $\exp(i2\pi\xi\mathcal{D})$ operator in the frequency domain is to multiply with $\exp(i2\pi\xi\mu)$, which we know corresponds to a shift by $\xi$ in the time domain: $f(u+\xi)$.

The associated kernels are

$$h_{\mathcal{PH}_\xi}(u,u') = e^{i2\pi\xi u}\delta(u-u'), \tag{1.228}$$

$$h_{\mathcal{SH}_\xi}(u,u') = \delta(u+\xi-u'). \tag{1.229}$$

That these two operators are also duals is easily seen by examining their effect in the frequency domain:

$$(\mathcal{PH}_\xi F)(\mu) = \exp\left[i2\pi\xi(-i2\pi)^{-1}d/d\mu\right]\, F(\mu) = F(\mu-\xi), \tag{1.230}$$

$$(\mathcal{SH}_\xi F)(\mu) = e^{i2\pi\xi\mu}F(\mu), \tag{1.231}$$

with associated kernels

$$H_{\mathcal{PH}_\xi}(\mu,\mu') = \delta(\mu-\xi-\mu'), \tag{1.232}$$

$$H_{\mathcal{SH}_\xi}(\mu,\mu') = e^{i2\pi\xi\mu}\delta(\mu-\mu'). \tag{1.233}$$

The kernels in the time and frequency domains are related by equation 1.193. Since $\mathcal{U}$ and $\mathcal{D}$ are Hermitian, $\mathcal{PH}_\xi$ and $\mathcal{SH}_\xi$ are unitary. The commutator of $\mathcal{PH}_\xi$ and $\mathcal{SH}_{\xi'}$ is

$$[\mathcal{PH}_\xi, \mathcal{SH}_{\xi'}] = [1 - e^{i2\pi\xi\xi'}]\mathcal{PH}_\xi\mathcal{SH}_{\xi'}. \tag{1.234}$$

Obviously these operations do not commute.

Now we consider the chirp multiplication operator $\mathcal{Q}_q = \exp(-i\pi q\mathcal{U}^2)$ and the chirp convolution (Fresnel) operator $\mathcal{R}_r = \exp(-i\pi r\mathcal{D}^2)$. Their effect in the time-domain is

$$(\mathcal{Q}_q f)(u) = e^{-i\pi q u^2} f(u), \tag{1.235}$$

$$(\mathcal{R}_r f)(u) = \exp\left[-i\pi r[(i2\pi)^{-2}d^2/du^2]\right]\, f(u) = e^{-i\pi/4}\sqrt{1/r}\, e^{i\pi u^2/r} * f(u). \tag{1.236}$$

The last equality is a special case of equation 1.207. The associated kernels are

$$h_{\mathcal{Q}_q}(u,u') = e^{-i\pi q u^2}\delta(u-u'), \tag{1.237}$$

$$h_{\mathcal{R}_r}(u,u') = e^{-i\pi/4}\sqrt{1/r}\, e^{i\pi(u-u')^2/r}. \tag{1.238}$$

That these two operators are also duals is easily seen by examining their effect in the frequency domain:

$$(\mathcal{R}_r F)(\mu) = e^{-i\pi r\mu^2} F(\mu), \tag{1.239}$$

$$(\mathcal{Q}_q F)(\mu) = \exp\left[-i\pi q[(-i2\pi)^{-2}d^2/d\mu^2]\right]\, F(\mu) = e^{-i\pi/4}\sqrt{1/q}\, e^{i\pi\mu^2/q} * F(\mu), \tag{1.240}$$

with associated kernels

$$H_{\mathcal{R}_r}(\mu,\mu') = e^{-i\pi r\mu^2}\delta(\mu-\mu'), \tag{1.241}$$

$$H_{\mathcal{Q}_q}(\mu,\mu') = e^{-i\pi/4}\sqrt{1/q}\, e^{i\pi(\mu-\mu')^2/q}. \tag{1.242}$$

The kernels in the time and frequency domains are once again related by equation 1.193. Since $\mathcal{U}$ and $\mathcal{D}$ are Hermitian, $\mathcal{R}_r$ and $\mathcal{Q}_q$ are unitary. The time-domain kernel of their commutator is given by

$$h_{[\mathcal{R}_r,\mathcal{Q}_q]}(u,u') = e^{-i\pi/4}\sqrt{\frac{1}{r}}\, e^{i\pi(u-u')^2/r}\left(e^{-i\pi qu'^2} - e^{-i\pi qu^2}\right). \tag{1.243}$$

It is important to note that since these two operators do not commute, an expression such as

$$e^{i\pi(r\mathcal{D}^2+q\mathcal{U}^2)} = e^{i\pi(q\mathcal{U}^2+r\mathcal{D}^2)} \tag{1.244}$$

cannot be written as $\exp(i\pi r\mathcal{D}^2)\exp(i\pi q\mathcal{U}^2)$ or $\exp(i\pi q\mathcal{U}^2)\exp(i\pi r\mathcal{D}^2)$. However, several relations can be used to manipulate such expressions, of which one important example is

$$e^{-i\theta\pi(\mathcal{U}^2+\mathcal{D}^2)} = e^{-i\pi\tan\theta\,\mathcal{U}^2}e^{-i\pi\ln(\cos\theta)\,(\mathcal{U}\mathcal{D}+\mathcal{D}\mathcal{U})}e^{-i\pi\tan\theta\,\mathcal{D}^2}. \tag{1.245}$$

Such relationships are in general known as Baker-Campbell-Hausdorff formulas (Wilcox 1967, Gilmore 1974, Wolf 1979, Stoler 1981). Similar considerations apply to and similar relations exist for the phase shift and translation operators discussed previously. In general, such a formula exists to correspond with each one of the decompositions we will come across in section 2.4.4, from which they can be derived. Another celebrated example is Glauber's formula (Cohen-Tannoudji, Diu, and Laloë 1977, page 174). If $\mathcal{A}$ and $\mathcal{B}$ both commute with their commutator ($[\mathcal{A},[\mathcal{A},\mathcal{B}]] = 0$ and $[\mathcal{B},[\mathcal{A},\mathcal{B}]] = 0$), then $[\mathcal{A},\Upsilon(\mathcal{B})] = [\mathcal{A},\mathcal{B}]\Upsilon'(\mathcal{B})$ and

$$e^{\mathcal{A}}e^{\mathcal{B}} = e^{\mathcal{A}+\mathcal{B}}e^{[\mathcal{A},\mathcal{B}]/2}. \tag{1.246}$$

At this point the reader may wish to look back at tables 1.2 and 1.4. Most properties of the Fourier transform are examples of finding the effect of a linear system on the Fourier transform. On the left hand column of table 1.4, there is a linear alteration of the function $f(u)$. The right hand column shows how the same alteration looks like in the Fourier domain. It is in particular interesting to examine the effects of the six operations of coordinate multiplication, differentiation, phase shift, translation, chirp multiplication, and chirp convolution. The latter four of these operators are sometimes referred to as *hyperdifferential operators*. A useful source is Wolf 1979. For a self-consistent operational calculus based on such operators, see Yosida 1984. Such operators have been made the basis of a study of optical systems in a series of papers by Nazarathy and Shamir (1980, 1982a, b).

We end by mentioning the eigenvalues and eigenfunctions of these operators. The eigenfunctions of the $\mathcal{Q}_q$ operator in the space domain are of course $\delta(u-v)$ (as for all

functions of the $\mathcal{U}$ operator) with eigenvalue $\exp(-i\pi q v^2)$. Likewise, the eigenfunctions of the $\mathcal{R}_r$ operator in the frequency domain are $\delta(\mu - v)$ (as for all functions of the $\mathcal{D}$ operator) with eigenvalue $\exp(-i\pi r v^2)$. Notice that these latter eigenfunctions are harmonic functions in the space domain.

### 1.6.3 Annihilation and creation operators

Properties 3 and 5 of table 1.8 can be rewritten as

$$\mathcal{A}\psi_n(u) = \sqrt{n}\,\psi_{n-1}(u), \tag{1.247}$$

$$\mathcal{A}^{\mathrm{H}}\psi_{n-1}(u) = \sqrt{n}\,\psi_n(u), \tag{1.248}$$

$$\text{or} \quad \mathcal{A}^{\mathrm{H}}\psi_n(u) = \sqrt{n+1}\,\psi_{n+1}(u),$$

where

$$\mathcal{A} \equiv \sqrt{2\pi}\,\frac{\mathcal{U} + i\mathcal{D}}{\sqrt{2}}, \tag{1.249}$$

$$\mathcal{A}^{\mathrm{H}} \equiv \sqrt{2\pi}\,\frac{\mathcal{U} - i\mathcal{D}}{\sqrt{2}}. \tag{1.250}$$

These operators are respectively referred to as the *annihilation* and *creation* operators, because of their effect on $\psi_n(u)$. Their commutator can be easily shown to be $[\mathcal{A}, \mathcal{A}^{\mathrm{H}}] = \mathcal{I}$ by using $[\mathcal{U}, \mathcal{D}] = (i/2\pi)\mathcal{I}$.

Now, let us form the products

$$\mathcal{A}^{\mathrm{H}}\mathcal{A} = \pi(\mathcal{U}^2 + \mathcal{D}^2) - \frac{1}{2}, \tag{1.251}$$

$$\mathcal{A}\mathcal{A}^{\mathrm{H}} = \pi(\mathcal{U}^2 + \mathcal{D}^2) + \frac{1}{2}, \tag{1.252}$$

by using the definitions given in equations 1.249 and 1.250. Alternately, using equations 1.247 and 1.248, we can obtain

$$(\mathcal{A}^{\mathrm{H}}\mathcal{A})\psi_n(u) = n\psi_n(u), \tag{1.253}$$

$$(\mathcal{A}\mathcal{A}^{\mathrm{H}})\psi_n(u) = (n+1)\psi_n(u), \tag{1.254}$$

from which we see that $\psi_n(u)$ is an eigenfunction of $(\mathcal{A}^{\mathrm{H}}\mathcal{A})$ with eigenvalue $n$. Comparing the first of these with the first of the preceding pair of equations, we obtain

$$\left[\pi(\mathcal{U}^2 + \mathcal{D}^2) - \frac{1}{2}\right]\psi_n(u) = n\psi_n(u), \tag{1.255}$$

$$\left[\pi(\mathcal{U}^2 + \mathcal{D}^2)\right]\psi_n(u) = (n + 1/2)\psi_n(u), \tag{1.256}$$

from which we note that $\psi_n(u)$ is also an eigenfunction of $\pi(\mathcal{U}^2 + \mathcal{D}^2)$ with eigenvalue $(n + 1/2)$. The reader may also wish to note that the final eigenvalue equation is the same as equation 1.176.

It is also possible to show $[(\mathcal{A}^{\mathrm{H}}\mathcal{A}), \mathcal{A}] = -\mathcal{A}$ and $[(\mathcal{A}^{\mathrm{H}}\mathcal{A}), \mathcal{A}^{\mathrm{H}}] = \mathcal{A}^{\mathrm{H}}$. Using $(\mathcal{A}^{\mathrm{H}}\mathcal{A})\psi_n(u) = n\psi_n(u)$, these commutation relations can be shown to imply

$$(\mathcal{A}^{\mathrm{H}}\mathcal{A})\mathcal{A}\psi_n(u) = (n-1)\mathcal{A}\psi_n(u), \tag{1.257}$$

$$(\mathcal{A}^{\mathrm{H}}\mathcal{A})\mathcal{A}^{\mathrm{H}}\psi_n(u) = (n+1)\mathcal{A}^{\mathrm{H}}\psi_n(u), \tag{1.258}$$

which state that $\mathcal{A}\psi_n(u)$ and $\mathcal{A}^{\mathrm{H}}\psi_n(u)$ are also eigenfunctions of $(\mathcal{A}^{\mathrm{H}}\mathcal{A})$ with the eigenvalues $(n-1)$ and $(n+1)$ respectively (consistent with equations 1.247 and 1.248). Finally, we note that use of equation 1.248 in the form $\psi_n(u) = n^{-1/2}\mathcal{A}^{\mathrm{H}}\psi_{n-1}(u)$ leads to the expression $\psi_n(u) = (n!)^{-1/2}(\mathcal{A}^{\mathrm{H}})^n\psi_0(u)$ which allows the calculation of Hermite-Gaussian functions of arbitrary order.

These operators are commonly used in quantum mechanics texts to solve the equation of the quantum-mechanical harmonic oscillator (Cohen-Tannoudji, Diu, and Laloë 1977).

## 1.7 Uncertainty relations

In the context of deterministic signals, uncertainty relations are bounds on the concentration or spread of the energy of a signal in two domains, commonly the time and frequency domains. In this section we will present a rather general result and then discuss some of its special cases. First, however, we introduce some definitions which may already be familiar to readers who have studied quantum mechanics (Cohen-Tannoudji, Diu, and Laloë 1977). Means $\eta_{\mathcal{A}}$, mean squares $m^2_{\mathcal{A}}$, and squared standard deviations (variances) $\sigma^2_{\mathcal{A}}$ of an operator $\mathcal{A}$ are defined as the weighted averages of $\mathcal{A}$, $\mathcal{A}^2$ and $(\mathcal{A}-\eta_{\mathcal{A}})^2$, normalized by the energy $\mathrm{En}[f] = \|f\|^2 = \int |f(u)|^2\,du$ of the signal $f$ under consideration:

$$\eta_{\mathcal{A}} \equiv \left[\int f^*(u)\mathcal{A}f(u)\,du\right] / \|f\|^2, \tag{1.259}$$

$$m^2_{\mathcal{A}} \equiv \left[\int f^*(u)\mathcal{A}^2 f(u)\,du\right] / \|f\|^2, \tag{1.260}$$

$$\sigma^2_{\mathcal{A}} \equiv \left[\int f^*(u)(\mathcal{A}-\eta_{\mathcal{A}})^2 f(u)\,du\right] / \|f\|^2 = m^2_{\mathcal{A}} - \eta^2_{\mathcal{A}}. \tag{1.261}$$

For example, if $\mathcal{A} = \mathcal{U}$, the coordinate multiplication operator, $\eta_{\mathcal{U}}$ is simply $\int u|f(u)|^2\,du/\|f\|^2$, the center of gravity of $|f(u)|^2$, and $\sigma_{\mathcal{U}}$ is a measure of the spread of $|f(u)|^2$. Likewise, if $\mathcal{A} = \mathcal{D}$, the derivative operator, Parseval's relation allows us to write $\int f^*(u)[\mathcal{D}f(u)]\,du = \int F^*(\mu)[\mu F(\mu)]\,d\mu$, so that $\eta_{\mathcal{D}}$ will give us the center of gravity of $|F(\mu)|^2$ and similarly $\sigma_{\mathcal{D}}$ is a measure of the spread of $|F(\mu)|^2$.

Now, let us denote the commutator of two operators $\mathcal{A}$ and $\mathcal{B}$ as $[\mathcal{A}, \mathcal{B}] = i\mathcal{C}$. Then, the standard deviations $\sigma_{\mathcal{A}}$ and $\sigma_{\mathcal{B}}$ of $\mathcal{A}$ and $\mathcal{B}$ satisfy

$$\sigma^2_{\mathcal{A}}\sigma^2_{\mathcal{B}} \geq \frac{|\eta_{\mathcal{C}}|^2}{4}, \tag{1.262}$$

where $\eta_{\mathcal{C}}$ is the mean of $\mathcal{C}$ (Dym and McKean 1972, page 119). As an example we again consider the coordinate multiplication and differentiation operators $\mathcal{U}$ and $\mathcal{D}$ whose commutator was given before as $[\mathcal{U},\mathcal{D}] = (i/2\pi)\mathcal{I}$, so that $\mathcal{C} = (1/2\pi)\mathcal{I}$. These lead to an uncertainty relation of the form $\sigma_{\mathcal{U}}\sigma_{\mathcal{D}} \geq 1/4\pi$ or more explicitly

$$\sigma_{\mathcal{U}}\sigma_{\mathcal{D}} = \frac{\left[\int (u-\eta_{\mathcal{U}})^2 |f(u)|^2\,du\right]^{1/2}\left[\int (\mu-\eta_{\mathcal{D}})^2 |F(\mu)|^2\,d\mu\right]^{1/2}}{\left[\int |f(u)|^2\,du\right]^{1/2}\left[\int |F(\mu)|^2\,d\mu\right]^{1/2}} \geq \frac{1}{4\pi}, \tag{1.263}$$

which means that the spread of $|f(u)|^2$ and the spread of $|F(\mu)|^2$ cannot simultaneously be very small.

If we define a measure of spread $\Delta u$ and $\Delta\mu$ as $\sqrt{4\pi}$ times the standard deviations of $|f(u)|^2$ and $|F(\mu)|^2$ respectively, then the uncertainty relation expressed in terms of these measures of spread takes the form

$$\Delta u\,\Delta\mu \geq 1. \tag{1.264}$$

The absolute square of the unit-energy Gaussian function $2^{1/4}\Delta u^{-0.5}\exp(-\pi u^2/\Delta u^2)$, has standard deviation $\Delta u/\sqrt{4\pi}$. In the case of Gaussian functions, the above inequality is satisfied with equality: $\Delta u\,\Delta\mu = 1$. (At $u = \Delta u/2$, the Gaussian function drops to 0.46 of its value at $u = 0$ and 92.5% of the energy of the Gaussian is contained in the interval $[-\Delta u/2, \Delta u/2]$.)

We can further manipulate the numerator of the central term in equation 1.263 by first letting $f_1(u) \equiv f(u+\eta_{\mathcal{U}})$, and noting that $F_1(\mu)$ has the same magnitude as $F(\mu)$:

$$\left[\int u^2 |f_1(u)|^2\,du\right]^{1/2}\left[\int (\mu-\eta_{\mathcal{D}})^2 |F_1(\mu)|^2\,d\mu\right]^{1/2}. \tag{1.265}$$

Now, with $F_2(\mu) \equiv F_1(\mu+\eta_{\mathcal{D}})$, the above becomes

$$\left[\int u^2 |f_2(u)|^2\,du\right]^{1/2}\left[\int \mu^2 |F_2(\mu)|^2\,d\mu\right]^{1/2}, \tag{1.266}$$

which, by eliminating the subscripts, leads to

$$\left[\int u^2 |f(u)|^2\,du\right]\left[\int \mu^2 |F(\mu)|^2\,d\mu\right] \geq (16\pi^2)^{-1}\|f\|^4, \tag{1.267}$$

with equality for $f(u)$ a constant multiple of $\exp(-\chi u^2)$ where $\chi > 0$ (Dym and McKean 1972, pages 117–118). This is an alternate form of the uncertainty relation given in equation 1.263.

We will now take a slightly different approach to arrive at yet a third form. Let us first note that

$$m_{\mathcal{D}}^2\|f\|^2 = \int f^*(u)\mathcal{D}^2 f(u)\,du = \langle f, \mathcal{D}^2 f\rangle = \langle \mathcal{D}^{\mathrm{H}} f, \mathcal{D} f\rangle. \tag{1.268}$$

In the time domain we know that $\mathcal{D}f(u) = (i2\pi)^{-1}df(u)/du$ and $\mathcal{D}^{\mathrm{H}}f(u) = (-i2\pi)^{-1}df(u)/du$, so that

$$m_{\mathcal{D}}^2\|f\|^2 = \int \mathcal{D}^{\mathrm{H}} f^*(u)\mathcal{D}f(u)\,du = \int |(i2\pi)^{-1}df(u)/du|^2\,du. \tag{1.269}$$

Likewise, it is not difficult to show that $m_{\mathcal{U}}^2\|f\|^2 = \int |uf(u)|^2\,du$. Noting that $|\eta_{\mathcal{I}}|^2/4 = 1/4$, we can write

$$(m_{\mathcal{U}}^2 - \eta_{\mathcal{U}}^2)(m_{\mathcal{D}}^2 - \eta_{\mathcal{D}}^2) \geq \left(\frac{1}{2\pi}\right)^2 \frac{1}{4}. \tag{1.270}$$

Now, noting that $m_{\mathcal{U}}^2 - \eta_{\mathcal{U}}^2 \leq m_{\mathcal{U}}^2$ and $m_{\mathcal{D}}^2 - \eta_{\mathcal{D}}^2 \leq m_{\mathcal{D}}^2$, and using the above we ultimately obtain

$$\|uf(u)\|\,\|df(u)/du\| \geq \|f(u)\|^2/2, \tag{1.271}$$

again with equality when $f(u)$ is proportional to $\exp(-\chi u^2)$ (Born and Wolf 1980, page 773).

Intuitively, the uncertainty relation states that both a function and its Fourier transform cannot be simultaneously concentrated. Generally speaking, the narrower one is, the broader must be the other. A function $f(u)$ of approximate duration $\Delta u$ will necessarily exhibit non-negligible frequency components around $1/\Delta u$ so that its Fourier transform will exhibit a spread of $\Delta\mu \simeq 1/\Delta u$, implying $\Delta u\,\Delta\mu \geq 1$.

There are other kinds of uncertainty relations which embody the same basic concepts. For instance, assume $\|f\| = 1$ and define

$$\alpha^2 = \int_{-\Delta u/2}^{\Delta u/2} |f(u)|^2\,du, \tag{1.272}$$

$$\beta^2 = \int_{-\Delta\mu/2}^{\Delta\mu/2} |F(\mu)|^2\,d\mu. \tag{1.273}$$

Let us fix $\Delta u, \Delta\mu > 0$. Now, $\alpha = \beta = 1$ is clearly not possible. But how close to unity can they simultaneously be? Such bounds on $\alpha$ and $\beta$ for given $\Delta u$ and $\Delta\mu$ are discussed by Slepian and Pollak (1961, 1962), Landau and Pollak (1961), Slepian (1964, 1978), and more briefly by Dym and McKean (1972, page 122).

A simpler, but related result states that a function and its Fourier transform cannot be both compact (unless they are identically zero). A compact function is one which is zero outside a finite interval around the origin. Such a function can be expressed as itself multiplied by a sufficiently wide rectangle function. Upon Fourier transforming the product, we see that the Fourier transform of the function is equal to itself convolved with a sinc function, which cannot be compact.

An instructive treatment of uncertainty relations may be found in Vakman 1968.

## 1.8 Random processes

### 1.8.1 Fundamental definitions

A random process can be considered as a parametric random variable (Papoulis 1991). That is, if $f(u)$ is a random process, then $f(u_0)$ for a particular value of $u_0$, is a random variable with a probability density function $P_{f(u_0)}[f(u_0)]$. Thus, the probability that $f(u_0)$ will lie in $[f(u_0), f(u_0)+\Delta f(u_0)]$ is given by $P_{f(u_0)}[f(u_0)]\Delta f(u_0)$ and $\int P_{f(u_0)}[f(u_0)]\,df(u_0) = 1$. (In this section $u_0, u_1, u_2$ will denote particular instances of $u$.)

The mean of a random process $f(u)$ is defined through the ensemble average

$$\eta_f(u) \equiv \langle f(u) \rangle \equiv \int f(u)\, P_{f(u)}[f(u)]\, df(u), \tag{1.274}$$

where $u$ is interpreted as a parameter. The autocorrelation is defined as

$$R_{ff}(u_1, u_2) \equiv \langle f(u_1) f^*(u_2) \rangle, \tag{1.275}$$

where the calculation of the expectation value will this time involve the joint probability density of $f(u_1)$ and $f(u_2)$. Since the instantaneous power of a deterministic signal is given by $|f(u)|^2$, we interpret $R_{ff}(u,u) = \langle |f(u)|^2 \rangle \equiv m_f^2(u)$ as the expected power of $f(u)$. The autocorrelation is a non-negative definite function. It is Hermitian: $R_{ff}(u_1, u_2) = R_{ff}^*(u_2, u_1)$. It also satisfies $R_{ff}(u_1, u_2) \leq R_{ff}(u_1, u_1)$. The cross correlation of $f$ and $h$ is defined as

$$R_{fh}(u_1, u_2) = \langle f(u_1) h^*(u_2) \rangle. \tag{1.276}$$

$f(u)$ and $h(u)$ are orthogonal if $R_{fh}(u_1, u_2) = 0$ for all $u_1, u_2$.

A process $f(u)$ is said to be *wide-sense stationary* if and only if (i) $m_f(u)$ is finite, (ii) $\eta_f(u)$ is a constant independent of $u$, (iii) the autocorrelation is a function of $u_1 - u_2$ only and not a function of $u_1$ and $u_2$ separately: $R_{ff}(u_1, u_2) = R_{ff}(u_1 - u_2)$. It follows from the properties of the autocorrelation that $R_{ff}(u) = R_{ff}^*(-u)$, $R_{ff}(u) \leq R_{ff}(0)$, and $R_{ff}(0) = m_f^2(0)$.

The time-averaged mean and autocorrelation of a wide-sense stationary process are denoted by $\overline{f(u)}$ and $\overline{f(u + \upsilon/2) f^*(u - \upsilon/2)}$ and defined as

$$\overline{f(u)} = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} f(u)\, du, \tag{1.277}$$

$$\overline{f(u + \upsilon/2) f^*(u - \upsilon/2)} = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} f(u + \upsilon/2) f^*(u - \upsilon/2)\, du. \tag{1.278}$$

The wide-sense stationary process $f(u)$ will be called ergodic in the mean if $\langle f(u) \rangle = \overline{f(u)}$ and equals a constant. Likewise, the wide-sense stationary process $f(u)$ will be called ergodic in the autocorrelation if $\langle f(u + \upsilon/2) f^*(u - \upsilon/2) \rangle = \overline{f(u + \upsilon/2) f^*(u - \upsilon/2)}$ and is a function of $\upsilon$ only. Thus, ergodicity is a concept involving interchangeability of time and ensemble averages. The present discussion of ergodicity is not rigorous or completely accurate; the reader should consult standard texts on random processes for further details.

### 1.8.2 Power spectral density

Let $f(u)$ be a wide-sense stationary random process with mean $\eta_f$ and autocorrelation $R_{ff}(u)$. The power spectral density $S_{ff}(\mu)$ is defined as the Fourier transform of the autocorrelation:

$$S_{ff}(\mu) \equiv \int R_{ff}(u) e^{-i2\pi\mu u}\, du, \tag{1.279}$$

$$R_{ff}(u) = \int S_{ff}(\mu) e^{i2\pi\mu u}\, d\mu. \tag{1.280}$$

It can be shown from the properties of the autocorrelation that $S_{ff}(\mu)$ is real, always $\geq 0$, and that

$$\langle |f(u)|^2 \rangle = R_{ff}(0) = \int S_{ff}(\mu)\, d\mu, \tag{1.281}$$

which suggests that $S_{ff}(u)$ deserves its name in that it can indeed be interpreted as the spectral density of power. Further confirmation to this end comes from the relation between $S_{ff}(\mu)$ and the expectation value of $|F(\mu)|^2$, discussed in many standard texts.

### 1.8.3 Linear systems with random inputs

Let the random process $f(u)$ be input to the linear system $\mathcal{L}$ whose output $g(u)$ is given by

$$g(u) = \int h(u, u') f(u')\, du'. \tag{1.282}$$

It is possible to show that the mean of the output $\eta_g$ is given by $\eta_g = \eta_f \int h(u, u')\, du'$. As for the autocorrelation of the output process $R_{gg}(u_1, u_2)$, it is given by

$$R_{gg}(u_1, u_2) = \iint R_{ff}(u_1', u_2') h(u_1, u_1') h^*(u_2, u_2')\, du_1'\, du_2'. \tag{1.283}$$

In the event that the processes are wide-sense stationary, then this simplifies to

$$R_{gg}(v) = R_{ff}(v) * h(v) * h^*(-v), \tag{1.284}$$

which implies the following relation between input and output power spectral densities:

$$S_{gg}(\mu) = |H(\mu)|^2 S_{ff}(\mu). \tag{1.285}$$

## 1.9 Generalization to two dimensions

Most of the definitions and results given in this and other chapters can be generalized easily to two and higher dimensions in a trivial manner. Expressions involving abstract signals are of course not affected in any way. An expression of the form $g = \mathcal{L}f$ does not imply any dimensionality. As for expressions involving functional representations of signals, it is often possible to write two-dimensional versions of them by simply replacing the variables $u$ and $\mu$ with the pairs of variables $u, v$ and $\mu, \nu$, and by replacing integrals and summations with double integrals and summations. For instance, equations 1.15 and 1.16 become

$$\langle f, g \rangle = \sum_l \sum_m f^*(l, m) g(l, m), \tag{1.286}$$

$$\langle f, g \rangle = \iint f^*(u, v) g(u, v)\, du\, dv. \tag{1.287}$$

The output $g(u, v)$ of a linear system is related to its input by a relation of the form

$$g(u, v) = \iint h(u, v; u', v') f(u', v')\, du'\, dv'. \tag{1.288}$$

Most of the commonly used functions introduced at the beginning of this chapter are generalized such that they are separable. For instance, $\text{rect}(u,v) \equiv \text{rect}(u)\text{rect}(v)$ and so on for $\text{sinc}(u,v)$, $\text{gauss}(u,v)$, $\text{har}(u,v)$, $\text{chirp}(u,v)$, and $\delta(u,v) \equiv \delta(u)\delta(v)$. However, some new functions which are not separable can also be defined. Introducing polar coordinates $q^2 = u^2 + v^2$ and $\tan\phi = u/v$, it is possible to define the function $\text{rect}(q)$ which is unity inside the circle of radius $1/2$ centered at the origin and zero outside. It is also convenient to define the jinc function as $\text{jinc}(q) = J_1(\pi q)/2q$, where $J_1(\cdot)$ is the first order Bessel function of the first kind. Furthermore, non-separable versions of the chirp and Gaussian functions also exist which have terms in the exponent not only in $u^2$ and $v^2$ but also in $uv$.

Likewise, some of the systems defined in table 1.2 may be generalized in a separable manner. For instance, the kernel of $\mathcal{I}$ and $\mathcal{P}$ become $\delta(u \pm u', v \pm v') \equiv \delta(u \pm u')\delta(v \pm v')$. The multiplicative filter now has the more general kernel $h(u,v)\delta(u-u', v-v')$. The kernels of the shift, phase shift, chirp multiplication, and chirp convolution operators may be generalized as

$$\delta(u - u' + \xi_u, v - v' + \xi_v), \tag{1.289}$$

$$\exp[i2\pi(\xi_u u + \xi_v v)]\delta(u-u', v-v'), \tag{1.290}$$

$$\exp[-i\pi(q_u u^2 + q_v v^2)]\delta(u-u', v-v'), \tag{1.291}$$

$$e^{-i\pi/2}\sqrt{\frac{1}{r_u}}\sqrt{\frac{1}{r_v}}\exp[i\pi((u-u')^2/r_u + (v-v')^2/r_v)]. \tag{1.292}$$

The coordinate multiplication and differentiation operators, as already defined, will act along one dimension only. It is possible to define a complementary $\mathcal{V}$ operator with kernel $v\delta(v-v')$ which acts along the $v$ axis and distinguish differentiation operators in the two dimensions as $\mathcal{D}_u$ and $\mathcal{D}_v$.

The two-dimensional Fourier transform is defined as

$$F(\mu,\nu) = \iint f(u,v)e^{-i2\pi(\mu u + \nu v)}\,du\,dv. \tag{1.293}$$

The two-dimensional Fourier transforms of common separable functions of the form $f(u,v) = f_u(u)f_v(v)$ are obtained easily from their one-dimensional counterparts by using the result

$$F(\mu,\nu) = F_\mu(\mu)F_\nu(\nu), \tag{1.294}$$

where $F_\mu(\mu)$ and $F_\nu(\nu)$ are the one-dimensional Fourier transforms of $f_u(u)$ and $f_v(v)$ respectively. (Some common Fourier pairs are not separable and cannot be obtained this way; for instance, the two-dimensional Fourier transform of $\text{rect}(q)$ is $\text{jinc}(\varsigma)$, where $\varsigma$ is the polar coordinate variable in the frequency plane.) Since the Fourier transform kernel is separable, its eigenfunctions are also so. Denoting the two-dimensional Hermite-Gaussian functions by $\psi_{lm}(u,v)$, we have

$$\psi_{lm}(u,v) \equiv \psi_l(u)\psi_m(v). \tag{1.295}$$

Most properties of the two-dimensional Fourier transform are straightforward generalizations of the one-dimensional property. A notable exception is the following. If $f(u,v)$ has a two-dimensional Fourier transform $F(\mu,\nu)$, then $f(au+bv+e, cu+dv+f)$ has the two-dimensional Fourier transform

$$\frac{1}{|\Delta|}\exp\left[\frac{i2\pi[(de-bf)\mu+(af-ec)\nu]}{\Delta}\right] F\left(\frac{d\mu-c\nu}{\Delta}, \frac{-b\mu+a\nu}{\Delta}\right), \tag{1.296}$$

where $\Delta = ac - bd$.

As far as this report is concerned, the reader who has grasped the one-dimensional version of a result or concept should have no difficulty generalizing it to two dimensions. Special discussion of two-dimensional signals and systems may be found in Bracewell 1995 and Dudgeon and Mersereau 1984.

One often encounters rotationally symmetric two-dimensional functions and systems, especially in optics. These depend only on $q = (u^2+v^2)^{1/2}$ but not $\phi$ when expressed in polar coordinates. Referring the reader once again to Bracewell 1995 for a more extensive treatment, we will satisfy ourselves by noting that the two-dimensional Fourier transform $F(\varsigma)$ of a rotationally symmetric function $f(q)$ is also rotationally symmetric and is given by

$$F(\varsigma) = 2\pi\int_0^\infty f(q)J_0(2\pi\varsigma q)q\,dq, \tag{1.297}$$

$$f(q) = 2\pi\int_0^\infty F(\varsigma)J_0(2\pi\varsigma q)\varsigma\,d\varsigma, \tag{1.298}$$

where $J_0(q)$ is the zeroth order Bessel function of the first kind. The relationship between $f(q)$ and $F(\varsigma)$ is known as a Hankel transform.

## 1.10 Some additional definitions and results

### 1.10.1 The Radon transform and projection-slice theorem

The Radon transform $\mathcal{RDN}_\phi[f(u,v)](u')$ of a two-dimensional function $f(u,v)$ is defined as the integral projection of the function onto an axis making angle $\phi$ with the $u$ axis:

$$\mathcal{RDN}_\phi[f(u,v)](u') \equiv \int f(u'\cos\phi - v'\sin\phi, u'\sin\phi + v'\cos\phi)\,dv'. \tag{1.299}$$

(If $h(u,v)$ is the two-dimensional impulse response of a two-dimensional system, then the Radon transform $\mathcal{RDN}_\phi[h(u,v)](u')$ can be interpreted as the line response to the input $\delta(u\cos\phi + v\sin\phi)$. For example, the integral projection at angle $\phi = 0$ is the response to the input $\delta(u)$, and is given by $\int h(u,v)\,dv$.) Sometimes, $\mathcal{RDN}_\phi[f](u')$ is interpreted as a function of plane polar coordinates, with $\phi$ being the common polar angle and $u'$ corresponding to the radial variable.

Let us also define the slice $\mathcal{SLC}_\phi[F(\mu,\nu)](\mu')$ of a two-dimensional function $F(\mu,\nu)$ through the relation

$$\mathcal{SLC}_\phi[F(\mu,\nu)](\mu') \equiv F(\mu'\cos\phi, \mu'\sin\phi). \tag{1.300}$$

The slice of a two-dimensional function $F(\mu,\nu)$ at angle $\phi$ is a one-dimensional function which takes the values of $F(\mu,\nu)$ along the radial line making angle $\phi$ with the $\mu$ axis. This radial line has the parametric form $\mu = \mu'\cos\phi$, $\nu = \mu'\sin\phi$.

Of particular interest is the projection-slice theorem, which states how a function $f(u,v)$ can be recovered from its Radon transform. According to this theorem, the one-dimensional Fourier transform of the integral projection at angle $\phi$ is equal to the slice of the two-dimensional Fourier transform at angle $\phi$:

$$\mathcal{F}\left\{\mathcal{RDN}_\phi[f]\right\}(\mu') = \mathcal{SLC}_\phi[F(\mu,\nu)](\mu'), \tag{1.301}$$

where $F(\mu,\nu)$ is the two-dimensional Fourier transform of $f(u,v)$. If we think of $\phi$ as a parameter, both sides of the last equation are functions of the single variable $\mu'$. The above relation can be written more explicitly as

$$\int \left[\mathcal{RDN}_\phi[f(u,v)](u')\right] e^{-i2\pi\mu' u'}\, du' = \int\!\!\int f(u,v) e^{-i2\pi(\mu'\cos\phi\, u + \mu'\sin\phi\, v)}\, du\, dv. \tag{1.302}$$

The left hand side is the one-dimensional Fourier transform of the integral projection at angle $\phi$. The right hand side is the two-dimensional Fourier transform of $f(u,v)$ expressed in polar coordinates $(\mu',\phi)$. The proof of the theorem follows immediately upon substitution of equation 1.299 in the above relation.

More on the Radon transform and projection-slice theorem can be found, for instance, in Bracewell 1995 or Barrett 1984, among many other references.

### 1.10.2 Complex exponential integrals

Here we list a number of complex Gaussian integrals that will be needed in later chapters:

$$\int e^{-p^2u^2 \pm qu}\, du = \frac{\sqrt{\pi}}{p}\, e^{q^2/4p^2}, \tag{1.303}$$

$$\int u e^{-p^2u^2+2qu}\, du = \sqrt{\frac{\pi}{p}}\,\frac{q}{p}\, e^{q^2/p}. \tag{1.304}$$

The square roots are taken so that the arguments lie in the interval $(-\pi/2, \pi/2]$. These results are, of course, not valid when the integrals do not converge.

With the above, care must be exercised if $p^2$ or $p$ is complex or pure imaginary. Thus it will be more convenient to write a number of special forms. Let us start with the Fourier pair $\exp(i\pi u^2)$ and $\exp(i\pi/4)\exp(-i\pi\mu^2)$, write the Fourier transform relations between these and employ variable substitutions to obtain (McBride and Kerr 1987)

$$\int e^{i\pi(\chi u^2 \pm 2\xi u)}\, du = \frac{1}{\sqrt{\chi}} e^{i\pi/4} e^{-i\pi\xi^2/\chi} \qquad \xi \text{ real}, \chi > 0, \tag{1.305}$$

$$\int e^{-i\pi(\chi u^2 \pm 2\xi u)}\, du = \frac{1}{\sqrt{\chi}} e^{-i\pi/4} e^{i\pi\xi^2/\chi} \qquad \xi \text{ real}, \chi > 0. \tag{1.306}$$

Both formulas are consistent with equation 1.303 if we use $\sqrt{i} = \exp(i\pi/4)$ and $\sqrt{-i} = \exp(-i\pi/4)$ when extracting $p$ from $p^2$. The first of the above is also valid for $\chi < 0$, provided we employ the same square root convention. (This can be shown by writing $\chi = -|\chi|$ and using $\sqrt{-1} = i = \exp(i\pi/2)$.)

### 1.10.3 Stationary-phase integral

If $f(u)$ is continuous, and the derivative of $\kappa(u)$ vanishes at only a single point $u = \xi$ in $(-\infty, \infty)$ such that $\kappa'(\xi) = 0$ and $\kappa''(\xi) \neq 0$, then for sufficiently large $\mu$,

$$\int f(u) e^{i2\pi\mu\kappa(u)}\, du \simeq e^{i2\pi\mu\kappa(\xi)} f(\xi) \sqrt{\frac{i}{\mu\kappa''(\xi)}}\,. \tag{1.307}$$

Further discussion may be found in Papoulis 1968. The applications of the stationary phase integral in optics is particularly well discussed in Lohmann 1986.

### 1.10.4 Schwarz's inequality

The general form of this inequality is given in the appendix to this chapter. A commonly used form for two functions $f(u)$ and $h(u)$ is

$$\left| \int f^*(u) h(u)\, du \right|^2 \leq \left[ \int |f(u)|^2\, du \right] \left[ \int |h(u)|^2\, du \right]. \tag{1.308}$$

## 1.11 Further reading

Relatively elementary texts which may be useful for background reading include Bracewell 1995, 1999, Cohen-Tannoudji, Diu, and Laloë 1977, Papoulis 1968, 1977, and Strang 1988. Texts which may be useful for further study include Dym and McKean 1972, Naylor and Sell 1982, and Wolf 1979.

## 1.12 Appendix: Vector spaces and function spaces

Here we provide a basic review of vector and function spaces, assuming familiarity only with elementary linear algebra and vectors in $\mathbf{R}^N$ (Strang 1988). Our main purpose is to enable the reader to grasp the various parts of our presentation in a deeper and more unified way. More extensive discussions and greater rigor is to be found in, for instance, Wolf 1979, Naylor and Sell 1982, Debnath and Mikusiński 1990, and Roman 1992. Wolf 1979 will particularly suit those with a mathematical physics bent and covers in greater detail many of the other topics discussed in this chapter as well. An excellent exposition to most of the basic concepts used here for the simpler case of finite dimensional vectors and matrices may be found in Strang 1988.

### 1.12.1 Vector spaces

The most familiar example of a vector space is the common three-dimensional space consisting of the set of position vectors denoted as $\mathbf{r} = x\hat{\mathbf{u}}_x + y\hat{\mathbf{u}}_y + z\hat{\mathbf{u}}_z$ or $(x, y, z)$, where $\hat{\mathbf{u}}_x, \hat{\mathbf{u}}_y, \hat{\mathbf{u}}_z$ are unit vectors along the coordinate axes, and addition and multiplication with a scalar number are defined in the obvious way.

A vector space is a set of objects for which addition and scalar multiplication operations have been defined such that, for any vectors $f$, $g$, and $h$ which are members of this vector space, and for any scalars $a$ and $b$:

1. the sum $f + g$ is also a member of the vector space,
2. $f + g = g + f$,
3. $f + (g + h) = (f + g) + h$,
4. there exists a zero vector denoted by 0 such that $f + 0 = f$,
5. there exists the negative of $f$ denoted by $-f$ such that $f + (-f) = 0$,
6. the product $af$ is also a member of the vector space,
7. $(ab)f = a(bf)$,
8. $a(f + g) = af + ag$,
9. $(a + b)f = af + bf$,
10. there exists a scalar unity denoted by 1 such that $1f = f$.

The following properties of a vector space can be derived directly from the above defining axioms:

1. for any set of vectors $f_j$ and any set of scalars $a_j$, labeled by the integer index $j$, the vector $\sum_j a_j f_j$ is also a member of the vector space,
2. for any set of vectors $f_\upsilon$ and any set of scalars $a_\upsilon$, labeled by the real index $\upsilon$, the vector $\int_\upsilon a_\upsilon f_\upsilon$ is also a member of the vector space,
3. $0 + f = f + 0 = f$, $(-f) + f = f + (-f) = 0$, $-0 = 0$, $-(-f) = f$,
4. $f1 = 1f = f$, $-1f = -f$.

It is easy to see that the three-dimensional space of position vectors satisfies the axioms and hence the properties listed above. There are many other vector spaces whose elements or structure may be less familiar, but which still satisfy all of the above axioms and properties. These include vector spaces with infinite but still countably (discretely) many dimensions. The simplest example is the extension of the three-dimensional vector space

considered above to infinite dimensions. Members of such spaces will be represented by functions $f(l)$, $g(l)$, and so on, where each value of the integer variable $l$ corresponds to one of the discretely many dimensions. Of greater interest to us will be spaces which have not only infinite, but also uncountably (continuously) many dimensions. Members of such spaces will be represented by functions $f(u)$, $g(u)$, and so on, where each value of the real variable $u$ corresponds to one of the continuously many dimensions. Such vector spaces are also referred to as *function spaces*.

In this report we will usually deal with complex vector spaces whose members are complex-valued functions and in which the scalars are complex numbers.

### 1.12.2 Inner products and norms

An *inner product* associates a scalar $\langle f, g\rangle$ with any two elements $f$ and $g$ of a vector space, such that, for any vectors $f$, $g$, and $h$, and scalar $a$:

1. $\langle f, f\rangle$ is real and $\geq 0$, with equality if and only if $f = 0$,
2. $\langle f, g\rangle = \langle g, f\rangle^*$,
3. $\langle f, g + h\rangle = \langle f, g\rangle + \langle f, h\rangle$,
4. $\langle f, ag\rangle = a\langle f, g\rangle$.

A vector space for which an inner product is defined is called an inner product space. Two vectors whose inner product is zero are called *orthogonal* to each other. We will mostly employ the inner product definitions

$$\langle f, g\rangle \equiv \sum_l f^*(l)g(l), \tag{1.309}$$

$$\langle f, g\rangle \equiv \int f^*(u)g(u)\,du, \tag{1.310}$$

for the discrete and continuous cases respectively. These can be shown to satisfy the listed axioms.

A *norm* associates a scalar real number $\|f\|$ with every element $f$ of a vector space, such that, for any vectors $f$ and $g$, and any scalar $a$:

1. $\|f\| \geq 0$, with equality if and only if $f = 0$,
2. $\|af\| = |a|\,\|f\|$,
3. $\|f + g\| \leq \|f\| + \|g\|$ (triangle inequality).

A vector space for which a norm is defined is known as a normed vector space. If an inner product has already been defined, a norm can be defined as

$$\|f\| \equiv \sqrt{\langle f, f\rangle}, \tag{1.311}$$

which, in the event that the inner product is defined through equations 1.310 or 1.309, is known as the $L_2$ norm. It is possible to show that this definition satisfies the axioms listed above. The energy of $f$ is defined as $\langle f, f\rangle = \|f\|^2$ so that with the definitions in equations 1.309 and 1.310 we have

$$\langle f, f\rangle = \|f\|^2 = \sum_l |f(l)|^2, \tag{1.312}$$

$$\langle f, f\rangle = \|f\|^2 = \int |f(u)|^2\, du, \tag{1.313}$$

for the discrete and continuous cases respectively.

An inner product satisfying the axioms given above satisfies the following properties:

1. $\langle f, 0\rangle = \langle 0, f\rangle = 0$,
2. $\langle af, g\rangle = a^*\langle f, g\rangle$,
3. $\langle f + g, h\rangle = \langle f, h\rangle + \langle g, h\rangle$,
4. $|\langle f, g\rangle|^2 \leq \langle f, f\rangle\langle g, g\rangle$ (Cauchy-Schwarz inequality),

and with the norm as defined by equation 1.311 the further properties:

1. $|\langle f, g\rangle| \leq \|f\|\,\|g\|$ (Cauchy-Schwarz inequality),
2. $\|f + g\|^2 = \|f\|^2 + \|g\|^2 + 2\Re[\langle f, g\rangle]$,
3. $\|f + g\|^2 \leq \|f\|^2 + \|g\|^2 + 2|\langle f, g\rangle| \leq \|f\|^2 + \|g\|^2 + 2\|f\|\,\|g\|$,

where $\Re[\cdot]$ denotes the real part of a complex entity.

The distance $d(f, g)$ between two vectors $f$ and $g$ can be defined as

$$d(f, g) = \|f - g\|, \tag{1.314}$$

which is always nonzero if $f \neq g$. This relation associating a real number with every pair of vectors $f$ and $g$ in the vector space defines a *metric* for the inner product space. A space with a defined metric is known as a *metric space*. A definition of distance is often expected to satisfy the following axioms:

1. $d(f, g) \geq 0$, with equality if and only if $f = g$,
2. $d(f, g) = d(g, f)$,
3. $d(f, g) + d(g, h) \geq d(f, h)$,

as the definition given by equation 1.314 indeed does.

We will mostly, but not exclusively, deal with vectors whose energies and norms are finite. This is because, in most physical applications, the energy as defined here corresponds to actual physical energy. However, this will not exclude us from employing certain idealized unphysical functions with infinite energy as intermediaries in our calculations.

An inner product space (with norm given by equation 1.311 and all of whose members have finite energy), which satisfies an additional condition known as *completeness* which we do not discuss here, is known as a *Hilbert space* (Naylor and Sell 1982). The Hilbert space of complex-valued functions $f(u)$, $u$ real, for which the inner product is defined by equation 1.310, is known as $L_2$, whereas the Hilbert space of complex-valued functions $f(l)$, $l$ integer, for which the inner product is defined by equation 1.309, is known as $\ell_2$. Both of these spaces have discretely (countably) many dimensions (or degrees of freedom). Self-evident in the case of $\ell_2$, this is also true for $L_2$ whose members can always be represented by discretely (countably) many coefficients, for instance as when expanded in terms of a discrete basis such as the Hermite-Gaussian functions. The space of "physically realizable" signals introduced on page 4 is somewhat more restricted than these spaces.

# Chapter 2

# Wigner Distributions and Linear Canonical Transforms

## 2.1 Time-frequency and space-frequency representations

The Fourier transform of a signal gives the relative weights of the various frequency components that make up the signal. It tells us which frequencies exist in the signal and their strengths. However, since the Fourier transform $F(\mu)$ involves integration of the time-domain representation of the signal $f(u)$ from minus to plus infinity, it is difficult to tell by just looking at $F(\mu)$ where these frequencies are located in $f(u)$. The value of $F(\mu)$ at each frequency $\mu$ depends on the value of $f(u)$ at all values of $u$. This character of the Fourier transform is sometimes found to be at odds with common physical intuition and experience. For instance, music scores tell the musician which frequencies to generate at particular time intervals, embodying the notion of particular frequencies being localized around particular instances. When we change the frequency setting of a sinusoidal signal generator from 1 MHz to 2 MHz, we would be reluctant to say that the output waveform continues to contain a frequency component at 1 MHz; we would rather say that it used to contain this frequency, but that it no longer does. As another example, let us consider a linear FM (frequency modulation) signal of the form $\exp(i\pi u^2)$, whose instantaneous frequency $(2\pi)^{-1}d(\pi u^2)/du = u$ is linearly increasing with time $u$. The most pronounced frequency in this signal increases with passing time. Looking at a picture of a dressed person, we may say that the checkered jacket exhibits high spatial frequencies, whereas the white shirt exhibits low spatial frequencies, and the striped tie exhibits equally spaced discrete frequency components.

Clearly, we are well accustomed to the concept of time- or space-dependent frequency content; we often speak about the frequency content of signals at different times or locations. Time- or space-frequency distributions are functions of time (or space) and temporal (or spatial) frequency which display the frequency content of signals for different times (or locations).

### 2.1.1 Short-time or windowed Fourier transform

One way of obtaining the time-dependent frequency content of a signal is to take the Fourier transform of $f(u')$ over an interval around a point $u$, where $u$ is a variable parameter. This is called the short-time or windowed Fourier transform $WF_f^{(w)}(u,\mu)$ and may be defined as (Hlawatsch and Boudreaux-Bartels 1992)

$$WF_f^{(w)}(u,\mu) \equiv \int [f(u')w^*(u'-u)]e^{-i2\pi\mu u'}\,du', \tag{2.1}$$

where $w(u')$ is a suitably chosen lowpass unit-energy window function centered around the origin, which suppresses $f(u')$ outside an interval centered around $u$. A common choice is the unit-energy Gaussian function $2^{1/4}\Delta_u^{-0.5}\exp(-\pi u^2/\Delta_u^2)$, for which $WF_f^{(w)}(u,\mu)$ is essentially the Fourier transform of the function over the interval $[u-\Delta_u/2, u+\Delta_u/2]$, and thus gives us the distribution of frequencies in $f(u)$ in this interval. (At $u=\Delta_u/2$, the Gaussian function drops to 0.46 of its value at $u=0$, and 92.5% of the energy of the Gaussian is contained in this interval.) Thus, the short-time Fourier transform allows us to be specific about the location of certain frequencies with a time resolution of $\sim\Delta_u$. If we want to be able to specify the distribution of frequencies as a function of time with greater temporal accuracy, we must choose shorter (narrower) windows (smaller $\Delta_u$).

It is possible to show that one can express $WF_f^{(w)}(u,\mu)$ in terms of $F(\mu)$ as well:

$$\begin{aligned} WF_f^{(w)}(u,\mu) &= e^{-i2\pi\mu u}\int [F(\mu')W^*(\mu'-\mu)]e^{i2\pi\mu' u}\,d\mu' \\ &= \int [F(\mu''+\mu)W^*(\mu'')]e^{i2\pi\mu'' u}\,d\mu'', \end{aligned} \tag{2.2}$$

where $W(\mu)$ is the Fourier transform of $w(u)$. For the Gaussian function above, $W(\mu) = 2^{1/4}\Delta_\mu^{-0.5}\exp(-\pi\mu^2/\Delta_\mu^2)$, where $\Delta_\mu = 1/\Delta_u$. Multiplying $F(\mu')$ with the bandpass filter $W^*(\mu'-\mu)$ essentially suppresses all frequencies other than those in the interval $[\mu-\Delta_\mu/2, \mu+\Delta_\mu/2]$, resulting in a bandpass signal with center frequency $\mu$. Multiplying with $\exp(-i2\pi\mu u)$ in the time-domain, amounts to a frequency shift which converts this bandpass signal into a lowpass signal. Thus $WF_f^{(w)}(u,\mu)$, interpreted as a function of time $u$ with frequency $\mu$ as a parameter, is a lowpass signal whose frequency distribution is given by $F(\mu''+\mu)W^*(\mu'')$, which is simply the frequency distribution of $f(u)$ in an interval of width $\sim\Delta_\mu$ around $\mu$, shifted down to zero frequency. If we want to be able to specify $\mu$ with greater accuracy, the width $\sim\Delta_\mu$ of $W(\mu)$ must be made smaller. Since $\Delta_u\Delta_\mu = 1$ for a Gaussian window, we conclude that choosing a shorter window $w(u)$ increases temporal resolution while decreasing frequency resolution, whereas choosing a longer window decreases temporal resolution while increasing frequency resolution. More generally, the product of the temporal extent $\Delta_u$ and spectral extent $\Delta_\mu$ of an arbitrary window must always be greater or equal than (approximately) unity. Thus, this tradeoff between temporal and frequency resolution always exists. The extreme cases of $w(u) = \delta(u)$ and $w(u) = 1$ correspond to perfect time resolution and perfect frequency resolution respectively. (Hlawatsch and Boudreaux-Bartels 1992)

It is possible to recover $f(u)$ from $WF_f^{(w)}(u,\mu)$ by using the easily derived result (Hlawatsch and Boudreaux-Bartels 1992)

$$f(u) = \iint WF_f^{(w)}(u',\mu')w(u-u')e^{i2\pi\mu' u}\,du'\,d\mu', \tag{2.3}$$

which may be expressed as

$$f(u) = \iint WF_f^{(w)}(u',\mu')w_{u',\mu'}(u)\,du'\,d\mu', \tag{2.4}$$
$$w_{u',\mu'}(u) \equiv w(u-u')e^{i2\pi\mu' u},$$

where $w_{u',\mu'}(u)$ is interpreted as a basis signal centered at the time-frequency point $(u',\mu')$, since $w(u)$ is a lowpass signal centered around the origin. If $w(u)$ is taken as the Gaussian function used above, the time-frequency extent of this basis signal is $\sim \Delta_u \times \Delta_\mu$. Thus, we see that we can interpret $WF_f^{(w)}(u',\mu')$ as the weighting coefficient of the basis signal concentrated around the time-frequency point $(u',\mu')$, indicating the relative strength in $f(u)$ of certain frequencies $\mu'$ at certain times $u'$.

The absolute square of the windowed or short-time Fourier transform is known as the spectrogram $SP_f^{(w)}(u,\mu) \equiv |WF_f^{(w)}(u,\mu)|^2$. It can be interpreted as an indicator of the energy of the signal at the time and frequency point $(u,\mu)$ in the sense that $SP_f^{(w)}(u,\mu)\,du\,d\mu$ gives the energy of the signal in the time-frequency region $[u-du/2, u+du/2] \times [\mu - d\mu/2, \mu + d\mu/2]$. We also note that equations 2.1 and 2.2 can be made symmetrical by redefining the windowed Fourier transform as (Almeida 1994)

$$WF_f^{(w)}(u,\mu) = e^{i\pi\mu u}\int f(u')w^*(u'-u)e^{-i2\pi\mu u'}\,du', \tag{2.5}$$

$$WF_f^{(w)}(u,\mu) = e^{-i\pi\mu u}\int F(\mu')W^*(\mu'-\mu)e^{i2\pi\mu' u}\,d\mu'. \tag{2.6}$$

We will mostly remain with the standard definitions however.

### 2.1.2 Gabor expansion

The short-time Fourier transform is closely related to the Gabor expansion, which is an expansion of $f(u)$ in term of discretely many basis functions which are localized in time and frequency (Hlawatsch and Boudreaux-Bartels 1992):

$$f(u) = \sum_l \sum_m G_f^{(w)}(l,m)w_{lm}(u), \tag{2.7}$$
$$w_{lm}(u) \equiv w(u - l\,\delta u)e^{i2\pi(m\,\delta\mu)u},$$

where $w(u)$ is a suitably chosen lowpass unit-energy window function centered around the origin, so that $w_{lm}(u)$ is a function localized in time and frequency around the time-frequency point $(l\,\delta u, m\,\delta\mu)$. The time and frequency spacings $\delta u$ and $\delta\mu$ define a lattice in the time-frequency plane. The coefficients $G_f^{(w)}(l,m)$ are referred to as Gabor coefficients,

and the $w_{lm}(u)$ are known as Gabor logons. Note that this expansion is in terms of basis signals which are discretely spaced in the time-frequency plane, as opposed to the expansion in equation 2.4, which is in terms of basis signals which are continuously spaced in the time-frequency plane. A necessary condition for the discretely many logons to be sufficient to expand an arbitrary finite-energy signal $f(u)$ is $\delta u\,\delta\mu \leq 1$ (Daubechies 1990, 1992). This condition indicates the minimum density of the logons that is needed for them to constitute a basis set for finite-energy signals. If the temporal extent of the signal is $\sim \Delta u$ and its bandwidth is $\sim \Delta\mu$, then we would expect the number of Gabor coefficients which are non-negligible to be given by $\sim (\Delta u/\delta u)(\Delta\mu/\delta\mu)$. If $\delta u\,\delta\mu > 1$, the number of these coefficients would be less than $\Delta u\,\Delta\mu$, the time-bandwidth product of the signal. Clearly, we cannot expect the signal to be characterized by fewer coefficients than its time-bandwidth product. On the other hand, if $\delta u\,\delta\mu < 1$, the number of Gabor coefficients will be larger than the time-bandwidth product of the signal, indicating that in this case the logons are not linearly independent, and that the representation is redundant: it has more coefficients than needed. When $\delta u\,\delta\mu = 1$ the logons are linearly independent and the coefficients contain no redundancy; the number of coefficients equals the time-bandwidth product of the signal (Hlawatsch and Boudreaux-Bartels 1992). We will later return to these concepts and see that each coefficient corresponds to one degree of freedom of the signal, and thus that the number of non-negligible Gabor coefficients corresponds to the number of degrees of freedom of the signal.

We now assume $\delta u\,\delta\mu = 1$ and also that $w(u)$ is chosen so that the $w_{lm}(u)$ constitute a basis set for all finite-energy signals (completeness). In general, the $w_{lm}(u)$ will not be orthogonal to each other, so that we cannot find the coefficient $G_f^{(w)}(l,m)$ by taking the inner product of $f(u)$ with $w_{lm}(u)$. However, the coefficients can be found by taking the inner product of $f(u)$ with a new set of signals $v_{lm}(u)$ as follows:

$$G_f^{(w)}(l,m) = \langle v_{lm}(u), f(u)\rangle = \int f(u) v_{lm}^*(u)\,du, \tag{2.8}$$
$$v_{lm}(u) \equiv v(u - l\,\delta u)e^{i2\pi(m\,\delta\mu)u},$$

where the $v_{lm}(u)$ and $w_{lm}(u)$ satisfy the biorthonormality condition

$$\langle v_{l'm'}(u), w_{lm}(u)\rangle = \int v_{l'm'}^*(u) w_{lm}(u)\,du = \delta_{ll'}\delta_{mm'}. \tag{2.9}$$

Determination of an appropriate function $v(u)$ such that the biorthonormality condition is satisfied is discussed in Bastiaans 1994, where it is shown how $v(u)$ may be easily determined by employing the Zak transform.

The biorthonormality condition above ensures that if we start with the coefficients $G_f^{(w)}(l,m)$ and construct the signal $f(u)$ using equation 2.7, then we can obtain the original coefficients from this $f(u)$ by using equation 2.8. It is also possible to show that the above biorthonormality condition implies the dual biorthonormality condition (Bastiaans 1994)

$$\sum_l \sum_m v_{lm}^*(u) w_{lm}(u') = \delta(u - u'). \tag{2.10}$$

This condition ensures that if we start from a signal $f(u)$ and find its Gabor coefficients by using equation 2.8, then we can reconstruct the signal by using equation 2.7.

¿From equation 2.8 we conclude that the Gabor coefficients are in fact samples of the windowed Fourier transform with window $v(u)$:

$$G_f^{(w)}(l,m) = WF_f^{(v)}(l\,\delta u, m\,\delta\mu). \tag{2.11}$$

Thus Gabor's expansion can also be viewed as a way of recovering a signal from the samples of its windowed Fourier transform (rather than the continuous windowed Fourier transform, as in equation 2.3).

Choosing $w(u) = 2^{1/4}\Delta_u^{-0.5}\exp(-\pi u^2/\Delta_u^2)$ ensures the greatest possible simultaneous concentration in time and frequency of the logons. (Actually the choice of a Gaussian function is not compatible with $\delta u\,\delta\mu = 1$ but requires that $\delta u\,\delta\mu$ be smaller than unity, even if only slightly so; see Daubechies 1992, page 107.) A natural choice for the parameters $\Delta_u$ and $\Delta_\mu = 1/\Delta_u$ characterizing the time and frequency extent of this Gaussian is $\Delta_u = 1\,\delta u$ and $\Delta_\mu = 1\,\delta\mu$. (The factor 1 is somewhat arbitrarily chosen; slightly different values may also be used.) In this case each $w_{lm}(u)$ snugly occupies the time-frequency cell on which it is centered, with only little overlap with adjacent cells. In this case the interpretation of the Gabor coefficients as an indicator of the time-frequency content of a signal around given time-frequency points becomes especially transparent. The function $v(u)$ corresponding to a Gaussian $w(u)$ is determined in Bastiaans 1994.

Gabor's expansion is a special case of what are more generally referred to as phase-space expansions (Landau 1993). What makes such expansions of interest is the fact that when appropriately defined, the coefficients $G_f^{(w)}(l,m)$ indicate how the energy of $f(u)$ is distributed over time and frequency; the larger this coefficient, the larger the contribution of that frequency at that time. If $f(u)$ is limited approximately to some time-frequency region such that its energy outside this region is small, then $f(u)$ can be reconstructed to the same degree of approximation from only those components $w_{lm}(u)$ which lie inside that region (Landau 1993). With Gabor's expansion, we have seen that we must have $\delta u\,\delta\mu \leq 1$; that is, at least one sampling point per unit time-frequency area is required. Similar conditions exist for more general classes of expansions (Daubechies 1990, 1992, Landau 1993). In general, at least one coefficient is required per unit time-frequency area. This is consistent with an argument based on the Nyquist sampling theorem, which requires sampling at a rate of $\delta u = 1/\Delta\mu$ over the extent $\Delta u$ of $f(u)$, implying a total of $\Delta u\,\Delta\mu$ samples, one sample per unit time-frequency area. Taken together, these facts support interpreting the minimum number of (non-redundant) expansion coefficients (or the time-bandwidth product) as the number of degrees of freedom of a signal. These concepts will be further discussed in section 2.3.

Original work underlying the Gabor expansion is scattered through many references. The reader may refer to the references found in the above cited works or to the useful papers by Bastiaans (1980, 1981a, 1982a, b, 1985, 1991, 1994). The 1994 paper containing

many references may constitute a useful starting point. A useful exposition of the Zak transform is Janssen 1988.

### 2.1.3 Wavelet transforms

Until now we exclusively took the width of the window function $\Delta_u$ and thus its dimensions in time-frequency $\Delta_u \times \Delta_\mu$ to be the same for all values of $u$ and $\mu$. Thus the absolute resolution obtained for all frequency components, high or low, was the same. However, since relative resolution is sometimes considered to be more relevant than absolute resolution (10 Hz resolution for 10 MHz is as good as 1 Hz resolution for 1 MHz), time-frequency representations involving windows of variable width have been invented. These so-called wavelet transforms are based on a complete set of orthogonal child wavelets $w_{u,\mu}(u')$ which are generated from a parent wavelet $w(u')$ through scaling and shift operations, where $w(u')$ is a unit-energy bandpass function with center frequency $\mu_0$ centered around the origin (Hlawatsch and Boudreaux-Bartels 1992):

$$w_{u,\mu}(u') = |\mu/\mu_0|^{1/2} w\left(\frac{\mu}{\mu_0}(u'-u)\right). \tag{2.12}$$

We note that $w_{u,\mu}(u')$ is centered around $u$ and has center frequency $\mu$. The bandwidth of $w_{u,\mu}(u')$ is $\mu/\mu_0$ times the bandwidth of $w(u')$. Thus we see that the bandwidth of $w_{u,\mu}(u')$ is proportional to its center frequency. This directly translates into frequency-proportionate frequency resolution in the wavelet transform defined as

$$WT_f^{(w)}(u,\mu) = \int f(u') w^*_{u,\mu}(u')\, du', \tag{2.13}$$

which may be compared to equation 2.1. A more detailed comparison of the wavelet transform to the short-time Fourier transform may be found in Hlawatsch and Boudreaux-Bartels 1992. Wavelet transforms are often expressed as time-scale representations, rather than time-frequency representations. Defining $\xi \equiv \mu_0/\mu$, we can rewrite the above definition as

$$TS_f^{(w)}(u,\xi) = \int f(u') w^*_{u,\xi}(u')\, du', \tag{2.14}$$

$$w_{u,\xi}(u') = \frac{1}{|\xi|^{1/2}}\, w\left(\frac{u'-u}{\xi}\right).$$

Introductory sources on wavelet transforms and time-scale representations include Mallat 1989, Daubechies 1990, 1992, Rioul and Vetterli 1991, Akansu and Haddad 1992, Chui 1992, Wavelets: Mathematics and Applications 1993, Walter 1994, Vetterli and Kovacevic 1995, Strang and Nguyen 1996, Suter 1997, and Mallat 1998.

We will be particularly interested in wavelet transforms generated from the parent wavelet $w(u) = \exp(i\pi u^2)$, resulting in the wavelet transform

$$TS_f^{(\text{chirp})}(u,\xi) = \frac{1}{|\xi|^{1/2}} \int f(u') e^{-i\pi(u'-u)^2/\xi^2}\, du', \tag{2.15}$$

which we recognize to be essentially the Fresnel transform. This class of wavelet transforms has been discussed in Onural 1993 and Onural and Kocatepe 1995.

### 2.1.4 Remarks

A very large number of different time-frequency representations have been suggested for their particular properties and suitability for different applications (Cohen 1989, 1995, Hlawatsch and Boudreaux-Bartels 1992). We will mostly concentrate on the Wigner distribution and a number of other closely related representations which are discussed in detail in the following sections.

It is tempting to view time-frequency representations as alternative representations of a signal, just as the time-domain representation, frequency-domain representation, and so on, as we discussed in chapter 1. This is justified by the fact that they often contain the same (or almost the same) information as these other representations. However, there are a number of differences: (i) Time-frequency representations are not always linearly related to other representations, such as the time-domain representation. (ii) It is not always possible to interpret a time-frequency representation as the coefficient of expansion in terms of a basis set, and even when this is the case, the basis may not be orthonormal. (iii) Time-frequency representations of functions of one variable are functions of two variables (time and frequency).

We end by noting that the discussion of this section is not totally precise and far from rigorous. Readers desiring greater rigor and more precise versions of the various statements we have made should consult the references cited.

## 2.2 The Wigner distribution and the ambiguity function

### 2.2.1 The Wigner distribution

The Wigner distribution $W_f(u,\mu)$ of a signal $f$ can be defined in terms of the time-domain representation $f(u)$ of the signal as (Claasen and Mecklenbräuker 1980a, b, c)

$$W_f(u,\mu) \equiv \int f(u+u'/2)f^*(u-u'/2)e^{-i2\pi\mu u'}\,du'. \tag{2.16}$$

Roughly speaking, $W_f(u,\mu)$ is a function which gives the distribution of signal energy over time and frequency, a fact which is not immediately evident from the above definition. However, it is possible to show directly from the above definition that

$$\int W_f(u,\mu)\,d\mu = |f(u)|^2, \tag{2.17}$$

$$\int W_f(u,\mu)\,du = |F(\mu)|^2, \tag{2.18}$$

$$\iint W_f(u,\mu)\,du\,d\mu = \|f\|^2 = \mathrm{En}[f] = \text{signal energy}. \tag{2.19}$$

Note that since the signal energy is given by the integral of either $|f(u)|^2$ over time or the integral of $|F(\mu)|^2$ over frequency, the first two equations are consistent and imply the third. These properties are consistent with, but do not imply, the interpretation of $W_f(u,\mu)$ as the energy density at time-frequency point $(u,\mu)$. Indeed, there are intrinsic difficulties associated with the notion of the energy density of a signal at a specific time and frequency point, stemming from the uncertainty relation (section 1.7). However, we will later justify the interpretation of local averages of $W_f(u,\mu)$ approximately as the time-frequency energy density of a signal. As a consequence, the energy of the signal in any extended time-frequency region can be found by integrating $W_f(u,\mu)$ over that region.

A number of characteristics makes the Wigner distribution a very attractive time-frequency representation. The Wigner distribution is completely symmetric with respect to the time and frequency domains, as evidenced by its expression in terms of the frequency-domain representation $F(\mu)$ of the signal:

$$W_f(u,\mu) = \int F(\mu+\mu'/2)F^*(\mu-\mu'/2)e^{i2\pi\mu' u}\,d\mu', \tag{2.20}$$

which can also be derived from equation 2.16. This symmetry is preserved also with respect to a continuum of domains we will refer to as fractional Fourier domains. Then, it will become particularly apparent that the geometric shape of the Wigner distribution has a reality and significance independent of the particular coordinate system in the time-frequency plane in which it is expressed. The Wigner distribution should be considered as an abstract geometric entity associated with the signal $f$ in the abstract, not being tied to a particular representation of $f$ in a particular domain.

The Wigner distribution of some common signals are given in table 2.1. The Wigner

| | $f(u)$ | $W_f(u,\mu)$ |
|---|---|---|
| 1. | $\exp(i2\pi\xi u)$ | $\delta(\mu-\xi)$ |
| 2. | $\delta(u-\xi)$ | $\delta(u-\xi)$ |
| 3. | $\exp[i\pi(\chi u^2+2\xi u+\zeta)]$ | $\delta(\mu-\chi u-\xi)$ |
| 4. | $(2\chi)^{1/4}\exp(-\pi\chi u^2)$ | $2\exp[-2\pi(\chi u^2+\mu^2/\chi)]$ |
| 5. | $\mathrm{rect}(u)$ | $2(1-\lvert 2u\rvert)\,\mathrm{rect}(u)\,\mathrm{sinc}\,[2(1-\lvert 2u\rvert)\mu]$ |

Table 2.1: Wigner distribution of some common signals. $\xi,\chi,\zeta$ are real.

distribution of the impulse and harmonic functions are easily interpreted in terms of the expected distribution of signal energy of these functions. The Wigner distribution of the chirp function, which includes these two as special cases, is found to be concentrated along the line giving the instantaneous frequency of the chirp: $(2\pi)^{-1}d[\pi(\chi u^2+2\xi u+\zeta)]/du = \chi u+\xi$. The Wigner distribution of a Gaussian signal is a Gaussian in $u$ and $\mu$ whose time and frequency profiles match the time and frequency representations of the Gaussian signal. We observe that the Wigner distribution of the rectangle function is negative for certain values of $u$ and $\mu$. In fact, it is more the norm than the exception for the Wigner

distribution of a signal to exhibit negative values for some values of $u$ and $\mu$, a fact which complicates its interpretation as an energy density. We also give the Wigner distribution of the scaled Hermite-Gaussian function $\psi_n(u/M)$:

$$W_{\psi_n(u/M)}(u,\mu) = 2(-1)^n \exp\left[-2\pi(u^2/M^2 + M^2\mu^2)\right] L_n\left[4\pi(u^2/M^2 + M^2\mu^2)\right], \quad (2.21)$$

where $L_n(\cdot)$ denotes the Laguerre polynomials (Bastiaans 1997). We finally note the Wigner distribution of the delta train or comb function defined as $f(u) = \sum_{n=-\infty}^{\infty} \delta(u-n)$, which is given by

$$W_f(u,\mu) = \frac{1}{2} \sum_{n=-\infty}^{\infty} \sum_{n'=-\infty}^{\infty} (-1)^{nn'} \delta(u - n/2)\delta(\mu - n'/2). \quad (2.22)$$

A pictorial discussion of this Wigner distribution may be found in Testorf and Ojeda-Castañeda 1996.

To recover $f(u)$ from its Wigner distribution $W_f(u,\mu)$, we note that equation 2.16 is a Fourier transform relation which can be inverted as

$$f(u + u'/2)f^*(u - u'/2) = \int W_f(u,\mu) e^{i2\pi\mu u'} \, d\mu. \quad (2.23)$$

If and only if upon evaluation of the right hand side of this equation we arrive at a function expressible in the form indicated by the left hand side, is the given two-dimensional function $W_f(u,\mu)$ a legitimate Wigner distribution of some function $f(u)$. Otherwise, the given two-dimensional function does not correspond to the Wigner distribution of any function, as is the case for most two-dimensional functions. Assuming that $W_f(u,\mu)$ is indeed a legitimate Wigner distribution, we can show

$$f(u) = \frac{1}{f^*(0)} \int W_f(u/2,\mu) e^{i2\pi\mu u} \, d\mu. \quad (2.24)$$

We see that the original function can be recovered only up to a complex constant of unit magnitude. In other words, any function of the form $f(u)\exp(i\pi\zeta)$ where $\zeta$ is a real constant has the same Wigner distribution as $f(u)$.

Important properties of the Wigner distribution are listed in table 2.2. We briefly

| height |
|---|

Table 2.2: Properties of the Wigner distribution (Hlawatsch and Boudreaux-Bartels 1992, table III). $\xi, \xi_1, \xi_2, M$ are real and $n$ is a positive integer.

comment on these properties: (1) The Wigner distribution is everywhere real but not always positive, an issue we will further discuss below. (2,3) It is time and frequency shift-invariant in the sense that shifting the time- and/or frequency-domain representations of a signal results in corresponding shifts in the Wigner distribution. (4,5) Integrating out

time or frequency returns the energy distribution with respect to frequency or time. (6,7) The time and frequency moments can be calculated as weighted averages in the time-frequency plane directly. Here these follow directly from properties 4 and 5, but there are other time-frequency distributions which satisfy 6 and 7 without satisfying 4 and 5. (8) Time and frequency scale inversely. (9,10) The weighted average of frequency for a given time is equal to the instantaneous frequency at that time, and the weighted average of time for a given frequency is equal to the group delay for that frequency. (However, instantaneous frequency and group delay can be meaningfully interpreted for only certain classes of signals.) (11,12) If the time- or frequency-domain representation of a signal is identically zero outside a certain interval, so is its Wigner distribution. However, if the time- or frequency-domain representation is zero inside a finite interval with nonzero values outside this interval, the Wigner distribution will not in general be zero inside that interval (Cohen 1989). (13) This property, known as Moyal's formula, is some kind of Parseval's relation between the time- (or frequency-) domain representation and the Wigner distribution. It basically states that the overlap integral or inner product of two Wigner distributions, is equal to the absolute square of the inner product of the two original signals. (14,15) Convolving $f(u)$ with another function $h(u)$ corresponds to convolving $W_f(u,\mu)$ with $W_h(u,\mu)$ in the time coordinate. Multiplying $f(u)$ with another function $h(u)$ corresponds to convolving $F(\mu)$ with $H(\mu)$, which further corresponds to convolving $W_f(u,\mu)$ with $W_h(u,\mu)$ in the frequency coordinate. (16) The Wigner distribution of the Fourier transform of a function is the Wigner distribution of the original function rotated clockwise by a right angle.

Certain properties of the Wigner distribution are sometimes considered undesirable. Since it is not linear but quadratic in the signal, the Wigner distribution of the sum of two signals will not be equal to the sum of their Wigner distributions, resulting in often undesired cross terms. This has motivated dealing with projections of the Wigner distribution, which we will see correspond to linear representations of the signal in different fractional Fourier domains. From a fundamental viewpoint, since the Wigner distribution is meant to have an energetic interpretation, and since energy is quadratic in the signal, the existence of cross terms should not be considered a defect. However, it does lead to difficulty when visually interpreting signals with multiple components, which we might want to be able to separately identify in a time-frequency plot. Experience with and recognition of the nature of the interference terms (discussed in Hlawatsch and Boudreaux-Bartels 1992), is of great benefit in interpreting such signals. Certain smoothed Wigner distributions, such as the Choi-Williams distribution (Choi and Williams 1989, Cohen 1989), allow the suppression of the interference terms at the expense of time-frequency concentration in a controlled manner by including an adjustable parameter.

The fact that the Wigner distribution can be negative for certain time-frequency values is often considered undesirable because it conflicts with the interpretation of the Wigner distribution as the distribution of signal energy. Such negative values will also tend to disappear with smoothing. We will return to this issue further below.

It is instructive to write the Wigner distribution in terms of the Hermite-Gaussian expansion of a function. The Wigner distribution of a function $f(u) = \sum_{n=0}^{\infty} C_n \psi_n(u)$ becomes

$$\begin{aligned} W_f(u,\mu) &= \int \sum_{n=0}^{\infty} C_n \psi_n(u+u'/2) \sum_{n'=0}^{\infty} C_{n'}^* \psi_{n'}(u-u'/2) e^{-i2\pi\mu u'}\, du' \\ &= \sum_{n=0}^{\infty}\sum_{n'=0}^{\infty} C_n C_{n'}^* \int \psi_n(u+u'/2)\psi_{n'}(u-u'/2) e^{-i2\pi\mu u'}\, du' \\ &= \sum_{n=0}^{\infty}\sum_{n'=0}^{\infty} C_n C_{n'}^* W_{\psi_n,\psi_{n'}}(u,\mu), \end{aligned} \tag{2.25}$$

where $W_{\psi_n,\psi_{n'}}(u,\mu)$ is the *cross Wigner distribution* of $\psi_n(u)$ and $\psi_{n'}(u)$ (Claasen and Mecklenbräuker 1980a). If the final equation is considered as the expansion of a two-dimensional function in terms of the cross Wigner distributions $W_{\psi_n,\psi_{n'}}(u,\mu)$, we see that arbitrary expansion coefficients of the form $C_{nn'}$ are not possible, but that only outer products of the form $C_n C_{n'}^*$ can appear. This is consistent with the fact that the Wigner distribution contains the information of a one-dimensional signal; very few two-dimensional functions are Wigner distributions of some signal.

A useful edited collection on the Wigner distribution and its applications in signal processing is Wigner distribution: Theory and Applications in Signal processing 1997. An extension of the Wigner distribution to signal spaces, rather than a single signal is given in Hlawatsch and Kozek 1993. The original work of Wigner is Wigner 1932.

### 2.2.2 The ambiguity function

Another time-frequency distribution which is closely related to the Wigner distribution is known as the ambiguity function $A(\bar{u},\bar{\mu})$ defined as

$$\begin{aligned} A_f(\bar{u},\bar{\mu}) &\equiv \int f(u'+\bar{u}/2) f^*(u'-\bar{u}/2) e^{-i2\pi\bar{\mu}u'}\, du' \\ &= \int F(\mu'+\bar{\mu}/2) F^*(\mu'-\bar{\mu}/2) e^{i2\pi\mu'\bar{u}}\, d\mu', \end{aligned} \tag{2.26}$$

where the second equality is a consequence of the definition given in the first line. This definition should be carefully compared to that of the Wigner distribution given in equation 2.16. Whereas the Wigner distribution is the prime example of an *energetic* time-frequency representation, the ambiguity function is the prime example of a *correlative* time-frequency representation (Hlawatsch and Boudreaux-Bartels 1992). (The term time-frequency distribution is sometimes used interchangeably with the term time-frequency representation. However, it is more appropriate to reserve the term distribution for those representations which have an energetic interpretation.) The ambiguity function deserves this by virtue of the properties

$$A_f(\bar{u},0) = R_{ff}(\bar{u}) \equiv \int f(u'+\bar{u}) f^*(u')\, du' = \int |F(\mu')|^2 e^{i2\pi\mu'\bar{u}}\, d\mu', \tag{2.27}$$

| | $f(u)$ | $A_f(\bar{u},\bar{\mu})$ |
|---|---|---|
| 1. | $\exp(i2\pi\xi u)$ | $\exp(i2\pi\xi\bar{u})\delta(\bar{\mu})$ |
| 2. | $\delta(u-\xi)$ | $\exp(i2\pi\xi\bar{\mu})\delta(\bar{u})$ |
| 3. | $\exp[i\pi(\chi u^2+2\xi u+\zeta)]$ | $\exp(i2\pi\xi\bar{u})\delta(\bar{\mu}-\chi\bar{u})$ |
| 4. | $(2\chi)^{1/4}\exp(-\pi\chi u^2)$ | $\exp[-\pi(\chi\bar{u}^2+\bar{\mu}^2/\chi)/2]$ |
| 5. | $\operatorname{rect}(u)$ | $(1-\lvert\bar{u}\rvert)\operatorname{rect}(2\bar{u})\operatorname{sinc}[\bar{\mu}(1-\lvert\bar{u}\rvert)]$ |

Table 2.3: Ambiguity functions of some common signals. $\xi,\chi,\zeta$ are real.

$$A_f(0,\bar{\mu}) = R_{FF}(\bar{\mu}) \equiv \int F(\mu'+\bar{\mu})F^*(\mu')\,d\mu' = \int |f(u')|^2 e^{-i2\pi\bar{\mu}u'}\,du', \tag{2.28}$$

$$A_f(\bar{u},\bar{\mu}) \le A_f(0,0) = \|f\|^2 = \mathrm{En}[f] = \text{signal energy}, \tag{2.29}$$

which say that the on-axis profiles of the ambiguity function are equal to the autocorrelation of the signal in the time and frequency domains respectively.

The ambiguity functions of some common signals are given in table 2.3.

Figure 2.1: Relationships between $\gamma(u,\bar{u})$, $\Gamma(\mu,\bar{\mu})$, $A_f(\bar{u},\bar{\mu})$, and $W_f(u,\mu)$. The arrows indicate a Fourier transform with respect to the variables shown. (Bamler and Glünder 1983, figure 1, Hlawatsch and Boudreaux-Bartels 1992, figure 13)

The Wigner distribution and ambiguity function are Fourier transforms of two auxiliary functions defined as (Claasen and Mecklenbräuker 1980c)

$$\gamma(u,\bar{u}) \equiv f(u+\bar{u}/2)f^*(u-\bar{u}/2), \tag{2.30}$$

$$\Gamma(\mu,\bar{\mu}) \equiv F(\mu+\bar{\mu}/2)F^*(\mu-\bar{\mu}/2), \tag{2.31}$$

such that

$$W_f(u,\mu) = \int \gamma(u,\bar{u})e^{-i2\pi\mu\bar{u}}\,d\bar{u} = \int \Gamma(\mu,\bar{\mu})e^{i2\pi\bar{\mu}u}\,d\bar{\mu}, \tag{2.32}$$

$$A_f(\bar{u},\bar{\mu}) = \int \gamma(u,\bar{u})e^{-i2\pi\bar{\mu}u}\,du = \int \Gamma(\mu,\bar{\mu})e^{i2\pi\mu\bar{u}}\,d\mu. \tag{2.33}$$

Combining these relationships, we find that the ambiguity function is related to the Wigner distribution by what is essentially a two-dimensional Fourier transform:

$$A_f(\bar{u},\bar{\mu}) = \iint W_f(u,\mu)e^{-i2\pi(\bar{\mu}u-\bar{u}\mu)}\,du\,d\mu, \tag{2.34}$$

consistent with the energetic nature of the Wigner distribution and the correlative nature of the ambiguity function. It is also possible to show that $\Gamma(\mu,\bar{\mu})$ is the two-dimensional Fourier transform of $\gamma(u,\bar{u})$. These relationships are summarized in figure 2.1. It is also instructive to see the relationship between the (approximate) supports of these four

functions. By the support of a function we mean the region where the value of the function is non-negligible. Let us assume that the signal $f$ has negligible energy outside the time interval $[-\Delta u/2, \Delta u/2]$ and the frequency interval $[\Delta\mu/2, \Delta\mu/2]$. Clearly, the Wigner distribution will (approximately) have a rectangular support defined by these intervals. The supports of the remaining three functions are shown in figure 2.2.

Figure 2.2: Supports of (a) $\gamma(u, \bar{u})$, (b) $\Gamma(\mu, \bar{\mu})$, (c) $A_f(\bar{u}, \bar{\mu})$, and (d) $W_f(u, \mu)$ for an approximately time- and band-limited signal $f$ (Bamler and Glünder 1983, figure 2).

The properties of the ambiguity function are summarized in table 2.4 which has been prepared parallel to table 2.2. The properties of the ambiguity function are easily interpreted by virtue of the fact that the ambiguity function is the Fourier transform of the Wigner distribution.

| height |
|---|

Table 2.4: Properties of the ambiguity function (Hlawatsch and Boudreaux-Bartels 1992, table III). $\xi, \xi_1, \xi_2, M$ are real and $n$ is a positive integer.

The special cases of properties 13 in tables 2.2 and 2.4 when $g = f$ are worth noting:

$$\|f\|^4 = (\mathrm{En}[f])^2 = \iint [W_f(u,\mu)]^2 \, du \, d\mu = \iint |A_f(\bar{u},\bar{\mu})|^2 \, d\bar{u} \, d\bar{\mu}. \tag{2.35}$$

The consistency of these two properties follows from the fact that Parseval's relation implies that the square integrals of the Wigner distribution and ambiguity function, which are essentially a two-dimensional Fourier transform pair, must be equal.

A rather different and useful exposition of the ambiguity function and further discussion of its relation to the second-order moments of a signal may be found in Papoulis 1977, pages 284–295.

### 2.2.3 Cohen's class of shift-invariant distributions

A relatively broad class of energetic time-frequency distributions which includes the Wigner distribution and spectrogram as special cases is known as Cohen's class of shift-invariant time-frequency distributions (Cohen 1966, 1976, 1989, 1995, Hlawatsch and Boudreaux-Bartels 1992). These distributions may be defined in terms of the Wigner distribution through the two-dimensional convolution relation

$$TFE_f(u,\mu) = \psi_{TFE}(u,\mu) ** W_f(u,\mu) = \iint \psi_{TFE}(u-u', \mu-\mu') W_f(u',\mu') \, du' \, d\mu'. \tag{2.36}$$

$\psi_{TFE}(u, \mu)$ is a kernel uniquely corresponding to the distribution $TFE_f(u, \mu)$. A distribution is a member of this class if and only if the distribution corresponding to $f(u -$

$\xi_u)\exp(i2\pi\xi_\mu u)$ is equal to $TFE_f(u-\xi_u,\mu-\xi_\mu)$ (as in properties 2 and 3 in table 2.2), as can be shown from equation 2.36. This is what is meant by shift-invariance. The fact that we have defined this class in terms of the Wigner distribution does not by itself confer a privileged status to the Wigner distribution among other members of the class. It is also possible to define the Cohen class in terms of members other than the Wigner distribution (Cohen 1989).

In analogy with equation 2.34, we can define the correlative dual time-frequency representation $TFC_f(\bar{u},\bar{\mu})$ of the energetic time-frequency distribution $TFE_f(u,\mu)$ as (Hlawatsch and Boudreaux-Bartels 1992)

$$TFC_f(\bar{u},\bar{\mu}) = \iint TFE_f(u,\mu)e^{-i2\pi(\bar{\mu}u-\bar{u}\mu)}\,du\,d\mu, \tag{2.37}$$

from which it follows that equation 2.36 can also be written as

$$TFC_f(\bar{u},\bar{\mu}) = \Psi_{TFE}(\bar{u},\bar{\mu})A_f(\bar{u},\bar{\mu}), \tag{2.38}$$

where $\Psi_{TFE}(\bar{u},\bar{\mu})$ is the two-dimensional Fourier transform of $\psi_{TFE}(u,\mu)$. In this case the representation corresponding to $f(u-\xi_u)\exp(i2\pi\xi_\mu u)$ is equal to $TFC_f(u,\mu)\exp[i2\pi(\xi_\mu\bar{u}-\xi_u\bar{\mu})]$ (as in properties 2 and 3 in table 2.4). Table 2.5 lists the names of a number of time-frequency distributions together with their defining kernels $\Psi_{TFE}(\bar{u},\bar{\mu})$. The correlative duals of these distributions can also be deduced from these kernels.

| Distribution | Kernel $\Psi_{TFE}(\bar{u},\bar{\mu})$ |
|---|---|

Table 2.5: Selected shift-invariant time-frequency distributions and their defining kernels. $\zeta,\chi$ are real parameters and $\eta(\cdot),w(\cdot)$ are suitably selected functions. (Hlawatsch and Boudreaux-Bartels 1992, table IV)

Since the kernels $\psi_{TFE}(u,\mu)$ or $\Psi_{TFE}(\bar{u},\bar{\mu})$ fully characterize a distribution which is a member of the Cohen class, the properties of the distribution can often be determined by examining these kernels. Table 2.6 has been prepared in parallel with table 2.2 and states the constraints that these kernels must satisfy in order for the distribution to exhibit a given property. The kernel $\psi_{TFE}(u,\mu)=\delta(u,\mu)$ or $\Psi_{TFE}(\bar{u},\bar{\mu})=1$ of the Wigner distribution satisfies all of these constraints so that the Wigner distribution exhibits all of these properties.

### 2.2.4 Smoothing of the Wigner distribution

We have already mentioned that the existence of interference terms is usually considered an undesirable property of the Wigner distribution which often makes its visual interpretation difficult. Since such terms are often of oscillatory nature (Hlawatsch and Boudreaux-Bartels 1992), it is possible to attenuate them considerably by smoothing the

| Property | Kernel Constraint |
|---|---|

Table 2.6: Kernel constraints corresponding to various properties of shift-invariant time-frequency distributions. The numbered list of properties correspond to those in table 2.2. The constraints are given in terms of the kernel $\Psi_{TFE}$. $M, \bar{u}, \bar{u}', \bar{\mu}, \bar{\mu}'$ are arbitrary real numbers. (Hlawatsch and Boudreaux-Bartels 1992, table V)

Wigner distribution. This can be achieved by convolving the Wigner distribution with a smooth function. Looking back to equation 2.36, we conclude that such smoothed Wigner distributions are time-frequency distributions belonging to the Cohen class. (However, not all distributions belonging to the Cohen class are smoothed Wigner distributions; this requires that the kernel $\psi_{TFE}(u, \mu)$ be a smooth function). Naturally, such smoothing will result in some loss of time-frequency resolution, and also in a loss of some of the desirable properties of the Wigner distribution (as can be determined from table 2.6).

As we have mentioned before, another undesirable property of the Wigner distribution is that it can be negative for certain time-frequency values, which is troubling because it conflicts with the interpretation of the Wigner distribution as the distribution of signal energy. We will now argue that this is not a fundamental flaw, but rather only an inconvenience. We must first recognize that the value of the Wigner distribution of a signal at a certain time-frequency point, mathematically defined through equation 2.16, does not correspond to a physically measurable quantity. (This is in contrast to the spectrogram which corresponds to a physically measurable quantity, as one can physically window a function and then observe its spectrum.) This has to do with the fact that it is not possible to resolve or isolate a part of a signal which is concentrated at a single time-frequency point. By applying a narrowband filter to the signal in the frequency domain, we can isolate as narrow a band of frequencies as we wish, but application of this filter will also inevitably result in a broadening of the signal in the time domain (since the signal will be convolved with a broad function in the time domain). Alternatively, by multiplying the signal with a narrow window in the time domain, we can isolate as short an interval of the signal as we wish, but this process will also inevitably result in a broadening of the signal in the frequency domain (since the signal will be convolved with a broad function in the frequency domain). According to the uncertainty relation, the product of the duration of the impulse response of a filter and its bandwidth must be greater than (approximately) unity. Thus, the smallest part of the signal we can isolate and subject to an energy measurement has time-frequency area which must also be greater than (approximately) unity. The energy of such an isolated part of the signal is given by an (appropriately weighted) integral of the Wigner distribution over the relevant time-frequency region and it is this quantity that we would expect to be positive.

Measuring the energy of a signal over a unit time-frequency area is closely related to the concept of smoothing the Wigner distribution with a kernel of unit time-frequency

area. If it does not make sense to speak about the energy of the signal in time-frequency areas smaller than unity, then it also does not make sense to specify time-frequency points with joint accuracy exceeding that suggested by the uncertainty relation.

From table 2.5 we see that by smoothing the Wigner distribution of a signal with the coordinate-inverted Wigner distribution of a window function $w(u)$, we obtain the spectrogram of the signal based on the same window function (Hlawatsch and Boudreaux-Bartels 1992):

$$SP_f^{(w)}(u,\mu) = |WF_f^{(w)}(u,\mu)|^2 = \iint \psi_{SP}(u-u',\mu-\mu')W_f(u',\mu')\,du'\,d\mu', \qquad (2.39)$$
$$\psi_{SP}(u,\mu) = W_w(-u,-\mu),$$

where $W_w(u,\mu)$ is the Wigner distribution of the window function $w(u)$. From this result we can conclude that if the smoothing kernel $\psi_{TFE}(u,\mu)$ is the Wigner distribution of some unit-energy function, the resulting distribution will be a spectrogram and hence non-negative (and also a measurable quantity). In particular, let us consider a Gaussian window $2^{1/4}\Delta_u^{-0.5}\exp(-\pi u^2/\Delta_u^2)$ whose Wigner distribution is

$$W_{\mathrm{Ga}}(u,\mu) = 2\exp\left[-2\pi\left(\frac{u^2}{\Delta_u^2} + \mu^2\Delta_u^2\right)\right]. \qquad (2.40)$$

The Wigner distribution of any function extends over a time-frequency region whose area is at least unity. The Wigner distribution of the Gaussian function satisfies this condition with approximate equality, extending roughly over a region of unity area in the time-frequency plane. More generally, it is known that choosing a Gaussian smoothing function $\psi_{\mathrm{Ga}}(u,\mu)$ of the form

$$\psi_{\mathrm{Ga}}(u,\mu) = \frac{2}{\sqrt{\Delta_u\Delta_\mu}}\exp\left[-2\pi\left(\frac{u^2}{\Delta_u^2} + \frac{\mu^2}{\Delta_\mu^2}\right)\right] \qquad (2.41)$$

will result in a positive distribution $TFE_f(u,\mu)$ if $\Delta_u\Delta_\mu \geq 1$ (Cohen 1989). Notice that the Gaussian here is not necessarily the Wigner distribution of anything. The same result generalizes to oblique kernels of the form (Cohen 1989)

$$\psi_{\mathrm{Ga}}(u,\mu) \propto \exp\left[-2\pi\left(\frac{u^2}{\Delta_u^2} + \frac{2u\mu}{\Delta_{u\mu}^2} + \frac{\mu^2}{\Delta_\mu^2}\right)\right], \qquad (2.42)$$

provided $\Delta_u^2\Delta_\mu^2 \geq \Delta_{u\mu}^2/(1+\Delta_{u\mu}^2)$. The quantities $\Delta_u$ and $\Delta_\mu$ appearing in equation 2.41 are approximate measures of the spread of $\psi_{\mathrm{Ga}}(u,\mu)$ in time and frequency respectively. The factor $\Delta_{u\mu}$ appearing in the cross term is related to the obliqueness of the ellipsoidal contours of $\psi_{\mathrm{Ga}}(u,\mu)$. These examples suggest that smoothing the Wigner distribution with a kernel $\psi_{TFE}(u,\mu)$ whose time-frequency area is equal to or greater than unity will result in a positive distribution. Since convolution with this kernel effectively corresponds to a weighted average of the Wigner distribution, one is tempted to state that averages or

integrals of the Wigner distribution over regions of at least unit time-frequency area are always positive. However, despite being common wisdom, these statements are not true in general (Cohen 1989).

While not being true in general, this common wisdom indeed holds for a broad range of interesting cases, particularly when the kernel itself is a smooth localized function. Thus, allowing ourselves to be imprecise, we will take it to be an approximate truth that localized averages of Wigner distributions over time-frequency regions of area greater or equal than unity are always positive.

We emphasize that the above considerations are of theoretical interest, having to do with the interpretation of the negative values of the Wigner distribution. As far as visual interpretation of plots of Wigner distributions are of concern, smoothing with functions which are not the Wigner distribution of anything and/or which have time-frequency area less than unity, may be just as much or more effective if chosen properly. These may result in a display which still exhibits some negative values, but may nevertheless offer an attractive and meaningful visual result. For practical purposes, the choice of a smoothing kernel is often governed by the need to find a compromise between the two goals of maximum interference suppression and maximum time-frequency resolution. Kernels with an adjustable parameter, such as the Choi-Williams distribution already mentioned, are particularly suited to this purpose because of the tuning they allow through their free parameter (Cohen 1989).

### 2.2.5 Effect of linear systems on the Wigner distribution

If $g(u)$ is related to $f(u)$ through the linear relation

$$g(u) = \int h(u,u')f(u')\,du', \tag{2.43}$$

then $W_g(u,\mu)$ is related to $W_f(u,\mu)$ through the relation (Bastiaans 1978, 1979a)

$$W_g(u,\mu) = \iint K_h(u,\mu;u',\mu')W_f(u',\mu')\,du'\,d\mu', \tag{2.44}$$

$$K_h(u,\mu;u',\mu') = \iint h(u+u''/2,u'+u'''/2)h^*(u-u''/2,u'-u'''/2) \times e^{-i2\pi(u''\mu-u'''\mu')}\,du''\,du'''.$$

It is possible to show that the kernel $K_h(u,\mu;u',\mu')$ is always real. If two systems with kernels $K_{h_1}(u,\mu;u',\mu')$ and $K_{h_2}(u,\mu;u',\mu')$ are cascaded, the kernel $K_{h_3}(u,\mu;u',\mu')$ of the resulting system is given by

$$K_{h_3}(u,\mu;u',\mu') = \iint K_{h_2}(u,\mu;u'',\mu'')K_{h_1}(u'',\mu'';u',\mu')\,du''\,d\mu''. \tag{2.45}$$

We recall that unitary systems conserve norm and energy; that is, the energy of the output $g$ is equal to the energy of the input $f$. It is also possible to show that a linear

system which conserves energy is necessarily unitary. (Systems which conserve energy are also referred to as lossless and gainless systems.) We know that the kernel of a unitary system satisfies $h^{-1}(u,u') = h^*(u',u)$. The same condition can be expressed in terms of the kernel $K_h(u,\mu;u',\mu')$ as follows (Bastiaans 1978):

$$\iint K_h(u,\mu;u',\mu')\,du\,d\mu = 1. \tag{2.46}$$

This can be derived either from the condition $h^{-1}(u,u') = h^*(u',u)$ or more instructively as

$$\begin{aligned}\mathrm{En}[g] = \iint W_g(u,\mu)\,du\,d\mu &= \iiiint K_h(u,\mu;u'\mu')W_f(u',\mu')\,du'\,d\mu'\,du\,d\mu \\ &= \iint W_f(u',\mu')\left[\iint K_h(u,\mu;u',\mu')\,du\,d\mu\right] du'\,d\mu'.\end{aligned} \tag{2.47}$$

Clearly, if equation 2.46 is satisfied, then $\mathrm{En}[g] = \mathrm{En}[f]$. The converse is also true; the only way for $\mathrm{En}[g] = \mathrm{En}[f]$ for all $f$ is for equation 2.46 to be satisfied for all $u', \mu'$.

We will now focus our attention to the special linear systems given earlier in table 1.2. In table 2.7 we have summarized the associated kernels $K_h(u,\mu;u'\mu')$ and the Wigner distribution of the output $W_g(u,\mu)$ for these systems.

Most of the items in the table correspond to items appearing in table 2.2; the reader will have no difficulty matching these up. Convolution of $f(u)$ with $h(u)$ corresponds to convolution of their Wigner distributions in $u$. This is denoted as

$$W_g(u,\mu) = W_h(u,\mu) \overset{u}{*} W_f(u,\mu) \equiv \int W_h(u-u',\mu)W_f(u',\mu)\,du'. \tag{2.48}$$

Likewise, multiplication of $f(u)$ with $h(u)$, which corresponds to convolution of $F(\mu)$ with $H(\mu)$, corresponds to convolution of their Wigner distributions in $\mu$. This is denoted as

$$W_g(u,\mu) = W_h(u,\mu) \overset{\mu}{*} W_f(u,\mu) \equiv \int W_h(u,\mu-\mu')W_f(u,\mu')\,d\mu'. \tag{2.49}$$

It is worth examining the geometric distortions in the $u$-$\mu$ plane to which some of the systems appearing in table 2.7 correspond to. The region bounded by the rectangle shown in figure 2.3a represents the Wigner distribution of $f(u)$, within which a large fraction of the energy of $f(u)$ is assumed to be contained. The effect of the scaling operation, of which the identity and parity operations are a special case, is shown in figure 2.3b, which should be interpreted in the light of the fact that the Fourier transform of $f(u/M)$ is $|M|F(M\mu)$ and the fact that the projections of the Wigner distribution of $f(u)$ on the $u$ and $\mu$ axes correspond to $|f(u)|^2$ and $|F(\mu)|^2$ respectively. The effects of the coordinate shift and phase shift operations, shown in parts c and d of the figure, are simply to shift the Wigner distribution in the $u$ and $\mu$ directions respectively, in line with the corresponding

| Symbol | $h(u,u')$<br>$g(u)$ | $K_h(u,\mu;u',\mu')$<br>$W_g(u,\mu)$ |
|---|---|---|
| $\mathcal{I}$ | $\delta(u-u')$ | $\delta(u-u')\delta(\mu-\mu')$ |
| | $f(u)$ | $W_f(u,\mu)$ |
| $\mathcal{P}$ | $\delta(u+u')$ | $\delta(u+u')\delta(\mu+\mu')$ |
| | $f(-u)$ | $W_f(-u,-\mu)$ |
| $\mathcal{M}_M$ | $\sqrt{\lvert M\rvert}\,\delta(u-Mu')$ | $\delta(u-Mu')\delta(\mu-\mu'/M)$ |
| | $(1/\sqrt{\lvert M\rvert}\,)f(u/M)$ | $W_f(u/M,M\mu)$ |
| $\mathcal{SH}_\xi$ | $\delta(u+\xi-u')$ | $\delta(u+\xi-u')\delta(\mu-\mu')$ |
| | $f(u+\xi)$ | $W_f(u+\xi,\mu)$ |
| $\mathcal{PH}_\xi$ | $\exp(i2\pi\xi u)\delta(u-u')$ | $\delta(u-u')\delta(\mu-\xi-\mu')$ |
| | $\exp(i2\pi\xi u)f(u)$ | $W_f(u,\mu-\xi)$ |
| $\Lambda_h$ | $h(u)\delta(u-u')$ | $W_h(u,\mu-\mu')\delta(u-u')$ |
| | $h(u)f(u)$ | $\int W_h(u,\mu-\mu')W_f(u,\mu')\,d\mu'$ |
| $\mathcal{Q}_q$ | $\exp(-i\pi qu^2)\delta(u-u')$ | $\delta(\mu+qu-\mu')\delta(u-u')$ |
| | $\exp(-i\pi qu^2)f(u)$ | $W_f(u,\mu+qu)$ |
| $\Lambda_H$ | $h(u-u')$ | $W_h(u-u',\mu)\delta(\mu-\mu')$ |
| | $\int h(u-u')f(u')\,du'$ | $\int W_h(u-u',\mu)W_f(u',\mu)\,du'$ |
| $\mathcal{R}_r$ | $e^{-i\pi/4}\sqrt{1/r}\,\exp[i\pi(u-u')^2/r]$ | $\delta(u-r\mu-u')\delta(\mu-\mu')$ |
| | $e^{-i\pi/4}\sqrt{1/r}\,\exp(i\pi u^2/r)*f(u)$ | $W_f(u-r\mu,\mu)$ |
| $\mathcal{U}$ | $u\delta(u-u')$ | $u^2\delta(\mu-\mu')+(16\pi^2)^{-1}\delta''(\mu-\mu')$ |
| | $uf(u)$ | $u^2W_f(u,\mu)+(16\pi^2)^{-1}d^2W_f(u,\mu)/d\mu^2$ |
| $\mathcal{D}$ | $(i2\pi)^{-1}\delta'(u-u')$ | $\mu^2\delta(u-u')+(16\pi^2)^{-1}\delta''(u-u')$ |
| | $(i2\pi)^{-1}df(u)/du$ | $\mu^2W_f(u,\mu)+(16\pi^2)^{-1}d^2W_f(u,\mu)/du^2$ |
| $\mathcal{F}$ | $\exp(-i2\pi uu')$ | $\delta(\mu+u')\delta(u-\mu')$ |
| | $F(u)$ | $W_f(-\mu,u)$ |

Table 2.7: The effect of some special linear systems on the Wigner distribution. $\mathcal{I}$: Identity, $\mathcal{P}$: Parity, $\mathcal{M}_M$: Scaling, $\mathcal{SH}_\xi$: Translation, $\mathcal{PH}_\xi$: Phase shift, $\Lambda_h$: Multiplicative filter, $\mathcal{Q}_q$: Chirp multiplication, $\Lambda_H$: Convolutive filter, $\mathcal{R}_r$: Chirp convolution, $\mathcal{U}$: Coordinate multiplication, $\mathcal{D}$: Differentiation, $\mathcal{F}$: Fourier transform. $M,\xi,q,r$ are real parameters, $\delta'(u-u')\equiv d[\delta(u-u')]/du$ and $\delta''(u-u')\equiv d^2[\delta(u-u')]/du^2$.

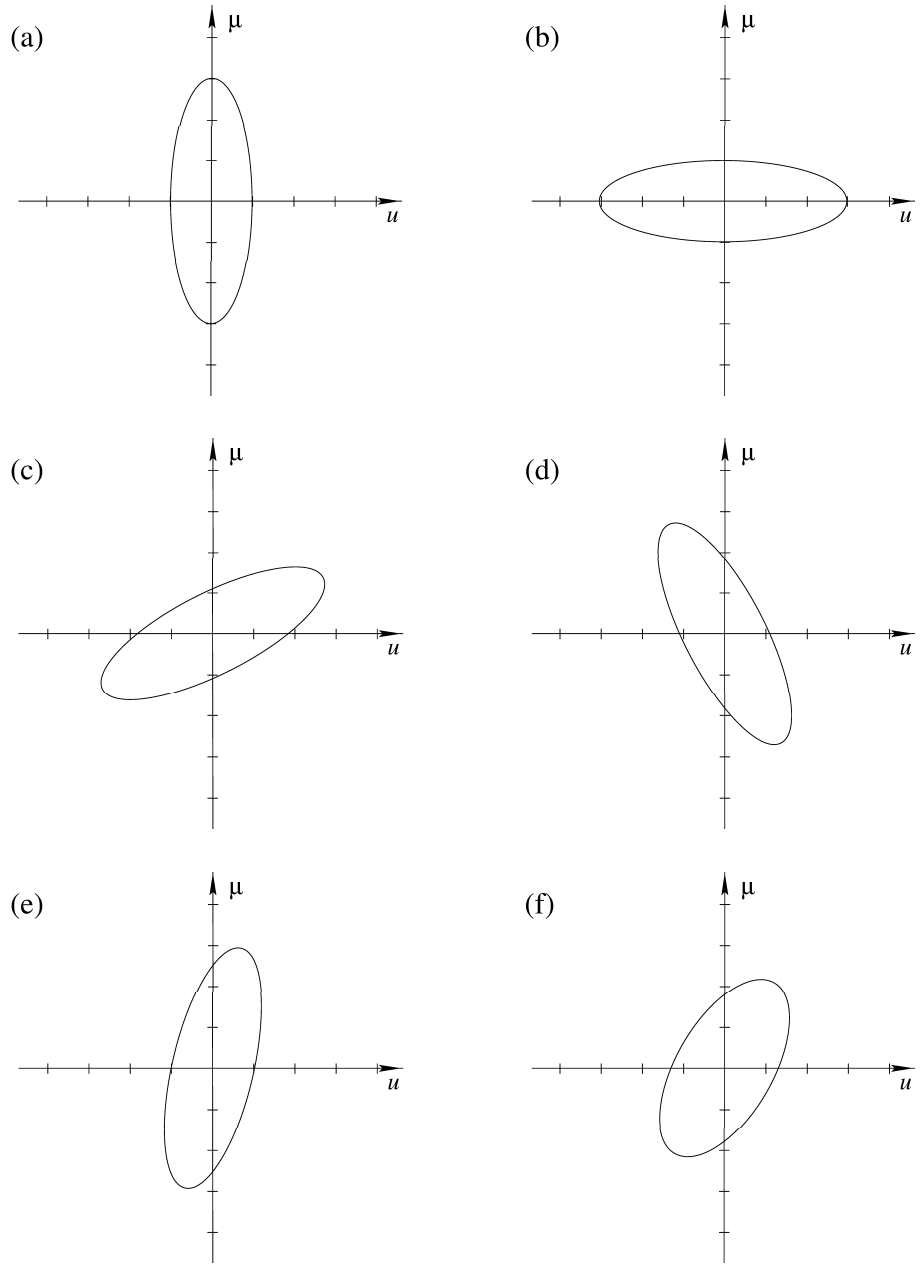


Figure 2.3: (a) Original signal. (b) Scaling with $M = 2$. (c) Translation with $\xi = -2$. (d) Phase shift with $\xi = 2$. (e) Chirp multiplication with $q = 1$. (f) Chirp convolution with $r = 1$.

properties of the Fourier transform (table 1.4.3–4). The effect of the Fourier transform, not shown in the figure, is to rotate the Wigner distribution by a right angle in the clockwise direction, essentially resulting in an interchange of the $u$ and $\mu$ axes. Chirp convolution results in shearing of the Wigner distribution in the $u$ direction and chirp multiplication results in shearing of the Wigner distribution in the $\mu$ direction. Notice that all of the above geometric transformations, including the shearing operations, are area preserving.

When we represent the Wigner distribution of a signal by such a closed curve or rectangle, we assume that a certain large fraction, say 95%, of the signal energy is contained inside the region bounded by that curve. Thus the geometric transformations illustrated in the figure show how the signal energy is redistributed in the $u$-$\mu$ plane. The fact that the area of the region containing 95% of the signal energy does not change, means that while energy is redistributed and mapped to different time- or space-frequency points, the concentration of energy in the $u$-$\mu$ plane does not change under the action of these systems. We may also finally note that any system composed by concatenating any number of the above systems will also result in an area preserving geometric transformation in the $u$-$\mu$ plane.

### 2.2.6 Time-frequency filtering

Here we briefly mention the concept of filtering in the time-frequency plane. In analogy with conventional time-invariant filtering where we modify the Fourier transform of a function with a multiplicative filter to alter its frequency content in the desired manner, time-frequency filtering is based on the idea of modifying the Wigner distribution (or other time-frequency representation) to alter the time-frequency content of the signal in the desired manner, or to construct signals with desired time-frequency content. The procedure is complicated by the fact that even the most reasonable modifications on the Wigner distribution of a signal, such as requiring it to be zero over a certain interval, may result in two-dimensional functions which are not the Wigner distributions of anything. This problem can be remedied in a number of ways. For instance, we may seek the signal which has the distribution closest to the one at hand, where closeness may be defined in the mean-square sense. See Saleh and Subotic 1985 and Cohen 1989 for further discussion.

### 2.2.7 Wigner distribution of random signals

The Wigner distribution of a random signal $f(u)$ is defined as the expectation value of the Wigner distribution defined in equation 2.16:

$$\begin{aligned} W_f(u,\mu) &= \left\langle \int f(u+u'/2) f^*(u-u'/2) e^{-i2\pi\mu u'}\, du' \right\rangle \\ &= \int \left\langle f(u+u'/2) f^*(u-u'/2) \right\rangle e^{-i2\pi\mu u'}\, du' \\ &= \int R_{ff}(u+u'/2, u-u'/2) e^{-i2\pi\mu u'}\, du', \end{aligned} \tag{2.50}$$

where $R_{ff}(u,u') \equiv \langle f(u)f^*(u')\rangle$ is the ensemble-averaged autocorrelation of $f(u)$. We see that the Wigner distribution of a random signal is essentially the Fourier transform of its autocorrelation with respect to the delay variable. Indeed, a similar interpretation is possible for the deterministic case. If we interpret $\gamma(u,u') = f(u+u'/2)f^*(u-u'/2)$ as some kind of time-dependent autocorrelation function (Cohen 1989), then the Wigner distribution is simply the Fourier transform of this function:

$$W_f(u,\mu) = \int \gamma(u,u')e^{-i2\pi\mu u'}\,du', \tag{2.51}$$

which can be considered to be a time-dependent generalization of the common result

$$|F(\mu)|^2 = \int R_{ff}(u')e^{-i2\pi\mu u'}\,du', \tag{2.52}$$

where $R_{ff}(u) \equiv \int f(u+u')f^*(u')\,du'$ here is the deterministic autocorrelation function.

If the signal is wide-sense stationary, then the ensemble-averaged autocorrelation $R_{ff}(u,u') = R_{ff}(u-u')$ so that $R_{ff}(u+u'/2, u-u'/2) = R_{ff}(u')$ and the Wigner distribution becomes independent of $u$ and reduces to the conventional power spectral density $S_{ff}(\mu)$. Thus, in the general case where the signal is not necessarily stationary, the Wigner distribution is readily interpreted as a time-varying power spectral density. For finite-energy signals, most of the original properties hold:

$$\int W(u,\mu)\,du = \left\langle |F(\mu)|^2 \right\rangle, \tag{2.53}$$

$$\int W(u,\mu)\,d\mu = \left\langle |f(u)|^2 \right\rangle. \tag{2.54}$$

$$\tag{2.55}$$

Readers wishing to learn more may consult the references in, for instance, Hlawatsch and Boudreaux-Bartels 1992 and Cohen 1995.

### 2.2.8 Wigner distribution of analytic signals

All of the results presented so far are valid for complex as well as real signals $f(u)$. In most physical applications, it is more common practice to work with the Wigner distribution of the real signal. However, some authors have argued that it is more meaningful to work with the Wigner distribution of the analytic signal (Boashash 1988, Zhu, Peyrin, and Goutte 1989). The Wigner distribution of the analytic signal of a real signal $f(u)$ is not simply the upper ($\mu > 0$) part of the Wigner distribution $W_f(u,\mu)$ of the real signal (Cohen 1989, page 969). For instance, the Wigner distribution of the signal $\exp(i\pi u^2)$, given by $\delta(\mu-u)$, is concentrated along the line $\mu = u$, with a simple interpretation in terms of the instantaneous frequency. However, the real signal $\cos(\pi u^2)$ will exhibit interference terms in addition to the line deltas $\delta(\mu-u)$ and $\delta(\mu+u)$.

The relationship between the Wigner distribution of a signal $f(u)$ and the Wigner distribution of its analytic signal $f_{\text{as}}(u)$ is given by

$$W_{f_{\text{as}}}(u,\mu) = 16 \int W_f(u-u',\mu)\,\mu\,\text{sinc}(4\mu u')\,du', \tag{2.56}$$

if $\mu \geq 0$, and 0 if $\mu < 0$ (Claasen and Mecklenbräuker 1980a).

### 2.2.9 Other properties

There are many other interesting properties of the Wigner distribution and ambiguity function that we do not discuss here. Of particular interest are properties relating the moments of a signal to the moments of its Wigner distribution and properties involving instantaneous frequencies and group delays (Claasen and Mecklenbräuker 1980a, b, c; Cohen 1989; Bastiaans 1989, 1991b). An inspiring treatment of the ambiguity function is Vakman 1968.

## 2.3 Sampling and the number of degrees of freedom

The support of a function is the subset of the real axis in which the function is not equal to zero. This subset is said to be compact if and only if its members are confined to a finite interval around the origin. A function will be referred to as compact if its support is so. In other words, a function is compact if and only if its nonzero values are confined to a finite interval around the origin. A signal is said to be compact in the $u$ domain if it is zero outside a finite interval around the origin in that domain. For instance, the function $\text{rect}(u)$ is compact, but the function $\exp(-\pi u^2)$ is not (although the latter may be considered to be approximately compact because its values are very small for larger $u$). If the Fourier transform of a function is compact, being zero outside the interval $(-\Delta\mu/2, \Delta\mu/2)$, it is said to be bandlimited with bandwidth $\Delta\mu$. Such a signal can be recovered from its samples taken at intervals $\delta u \leq 1/\Delta\mu$, a result known as Nyquist's sampling theorem. Taking the fewest possible samples ($\delta u = 1/\Delta\mu$), the sampled function $f_{\text{samp}}(u)$ becomes

$$\begin{aligned} f_{\text{samp}}(u) &= f(u)\,\text{comb}(u/\delta u) = \sum_{l=-\infty}^{\infty} f(l\,\delta u)\,\delta(u/\delta u - l) \\ &= \delta u \sum_{l=-\infty}^{\infty} f(l\,\delta u)\,\delta(u - l\,\delta u) = \delta u \sum_{l=-\infty}^{\infty} f(l/\Delta\mu)\,\delta(u - l/\Delta\mu). \end{aligned} \tag{2.57}$$

Taking the Fourier transform of both sides

$$\begin{aligned} \mathcal{F}[f_{\text{samp}}(u)](\mu) &= F(\mu) * \delta u\,\text{comb}(\delta u\,\mu) = F(\mu) * \delta u \sum_{l=-\infty}^{\infty} \delta(\delta u\,\mu - l) \\ &= F(\mu) * \sum_{l=-\infty}^{\infty} \delta(\mu - l\,\Delta\mu) = \sum_{l=-\infty}^{\infty} F(\mu - l\,\Delta\mu). \end{aligned} \tag{2.58}$$

The last expression tells us that sampling in the time domain results in periodic replication in the frequency domain. The original signal can be recovered by multiplying $\mathcal{F}[f_{\text{samp}}(u)](\mu)$ with a rectangular window $\text{rect}(\mu/\Delta\mu)$ which will single out the original spectrum $F(\mu)$. The same operation can be written in the time domain as a convolution of the form

$$f_{\text{samp}}(u) * \Delta\mu \,\text{sinc}(u\,\Delta\mu) = \sum_{l=-\infty}^{\infty} f(l/\Delta\mu)\,\text{sinc}(\Delta\mu\, u - l)$$

$$= \sum_{l=-\infty}^{\infty} f(l\,\delta u)\,\text{sinc}(u/\delta u - l) = \sum_{l=-\infty}^{\infty} f(l\,\delta u)\,\text{sinc}\left(\frac{u - l\,\delta u}{\delta u}\right), \tag{2.59}$$

which is the formula allowing us to reconstruct $f(u)$ from its samples and is known as the interpolation formula. Analogous results hold for sampling of the frequency-domain representation $F(\mu)$. Further discussion on the fundamentals of sampling may be found in Bracewell 1986, Marks 1991, and the edited book Advanced Topics in Shannon Sampling and Interpolation Theory 1993.

We already know that a function and its Fourier transform cannot both be compact (unless they are identically zero). That is, $\Delta u$ and $\Delta\mu$ as defined above, cannot both be finite. In practice however, it seems that we are always working with both a finite time (or space) interval and a finite bandwidth. Thus we will find it useful to abandon the above definitions of $\Delta u$ and $\Delta\mu$ in favor of less well defined yet more meaningful ones. A large percentage of the energy of most finite-energy signals arising in physical applications will be concentrated in a finite interval both in the time domain and in the frequency domain, although neither $f(u)$ nor $F(\mu)$ may be identically zero outside of these intervals. For instance, the Gaussian function $\text{gauss}(u)$ is clearly well concentrated around the origin in both the time and frequency domains, although it is not identically zero anywhere and its tails extend to infinity. Strictly speaking, both the temporal extent and the bandwidth of this signal are infinite. Since this is clearly counterintuitive, other measures of spread are often employed. One of these is to take the temporal extent or bandwidth of the signal as the standard deviation (or a certain number of standard deviations) of the time- and frequency-domain representations of the signal respectively. Another measure of spread, appropriate for certain functions which tend to diminish as we move away from their center of gravity, is the distance between the points at which the function has dropped to a certain fraction of its peak value, say $1/e$ or 0.05. A somewhat more generally applicable measure of spread is the length of the interval which contains a certain fraction, say 0.95, of the total energy of the signal. (All three of these measures are appropriate for the particularly well-behaved Gaussian function.). When we speak of the temporal extent $\Delta u$ and bandwidth (spectral extent) $\Delta\mu$ of a signal, we will usually be speaking of the length of an interval containing a sufficiently large fraction of the total energy of the signal. Having ensured that the signal energy outside this interval is negligible, we may thus assume that the signal is (approximately) confined to that interval. The Nyquist sampling theorem and interpolation formula will hold approximately in this case. This definition of spread

works well especially when we are dealing with signals of large time-bandwidth product, a notion which is defined further below.

We recall that we take $u$ and $\mu$ to be dimensionless variables. Let us assume that the time-domain representations of the signals we are dealing with are approximately confined to the interval $[-\Delta t/2, \Delta t/2]$ and that their frequency-domain representations are approximately confined to the interval $[-\Delta f/2, \Delta f/2]$ in real physical units. With this statement we mean that a sufficiently large percentage of the energies of the signals are confined to these intervals in the respective domains. This can be ensured by choosing $\Delta t$ and $\Delta f$ sufficiently large. (If the time or frequency representations of the signals are confined to intervals which are not centered around the origin, we may simply shift the origin of time and frequency so that this becomes the case.)

Let us now introduce the scaling parameter $s$ with the dimension of time and introduce dimensionless coordinates $u = t/s$ and $\mu = fs$. With these new coordinates, the time and frequency domain representations will be confined to intervals of length $\Delta t/s$ and $s\Delta f$. If we choose $s = \sqrt{\Delta t/\Delta f}$, the lengths of both intervals will now be equal to the dimensionless quantity $\sqrt{\Delta f \Delta t}$ which we may denote by $\Delta u$. It is often convenient to assume that such a dimensional normalization has been performed on the signals we work with so that the spread of the signal in the time and frequency domains are comparable in dimensionless coordinates.

We now define the time-bandwidth product $N$ for a set of signals, whose members we assume are approximately confined to an interval of length $\Delta u$ in the time domain and to an interval of length $\Delta\mu$ in the frequency domain. The time-bandwidth product (or space-bandwidth product) is defined as

$$N \equiv \Delta u \Delta \mu. \tag{2.60}$$

$N$ is always greater or equal to unity by virtue of the uncertainty relation. The time-bandwidth product is the minimum number of samples needed to characterize or identify a signal out of all possible signals whose energies are confined to time and frequency intervals of length $\Delta u$ and $\Delta\mu$. If we sample the time-domain representation of a signal at the Nyquist rate of $\delta u = 1/\Delta\mu$, the total number of samples lying in the interval $\Delta u$ is given by $\Delta u/(1/\Delta\mu) = \Delta u \Delta\mu$, which is simply the time-bandwidth product $N$. Alternatively, if we sample the frequency-domain representation of a signal at the Nyquist rate of $\delta\mu = 1/\Delta u$, the total number of samples lying in the interval $\Delta\mu$ is given by $\Delta\mu/(1/\Delta u) = \Delta u \Delta\mu$, which is again the time-bandwidth product $N$. (With the dimensional normalization above which results in $\Delta u = \Delta\mu$, the number of samples is $N = \Delta u^2$ with the samples being spaced $\Delta u^{-1} = N^{-1/2}$ apart in both domains.)

The time-bandwidth product of the set of signals we are dealing with can often be interpreted as the *number of degrees of freedom* or *dimensionality* of the set of signals. Since signals whose energies are (approximately) confined to intervals of length $\Delta u$ and $\Delta\mu$ in the time and frequency domains can be fully characterized by $N$ numbers, there is a one-to-one correspondence between these signals and $N$-dimensional vectors $\mathbf{r} = [r_1\ r_2\ \ldots\ r_N]^{\mathrm{T}}$.

We saw above that the number of samples needed to fully characterize a signal is the same in both the time and the frequency domain, and is given by the time-bandwidth product $N$. The time-bandwidth product will remain invariant under transformations to other representations as well, as the information content of the signal is not altered under invertible unitary transformations of the type we discussed in chapter 1. Thus if a signal can be uniquely characterized in a particular representation by $N$ complex numbers, this will also be the case in any other representation. Much like the norm and energy, the number of degrees of freedom is a property of the signals in the abstract, and not tied to their representations in a particular domain.

We can summarize by saying that the set of time- and band-limited signals in question has approximately $\Delta u \Delta \mu$ degrees of freedom. (Of course, strictly speaking, $f(u) = 0$ is the only such signal, so that we are continuing to talk about approximate time- and band-limitedness). Our argument has been based on sampling theory. We sample every $1/\Delta\mu$ over $\Delta u$ so that the total number of samples is $\Delta u/(1/\Delta\mu) = \Delta u \Delta \mu$, which we interpret as the number of degrees of freedom. For a more rigorous account, see Dym and McKean 1972, pages 129–131.

Great insight into the concept of the number of degrees of freedom can be gained through time-frequency representations. For instance, let us consider Gabor's expansion whose definition we repeat (equation 2.7)

$$f(u) = \sum_l \sum_m G_f^{(w)}(l, m) w_{lm}(u), \tag{2.61}$$

where $w_{lm}(u)$ are basis signals centered at the time-frequency point $(l\,\delta u, m\,\delta\mu)$, with $\delta u\,\delta\mu = 1$. Here we will assume that the $w_{lm}(u)$ are well concentrated in their respective time-frequency cells (of dimensions $\delta u \times \delta\mu$) centered around $(l\,\delta u, m\,\delta\mu)$. (For instance, this will be the case if we choose $w(u) = 2^{1/4}\Delta_u^{-0.5}\exp(-\pi u^2/\Delta_u^2)$ with $\Delta_u = \delta u$, as on page 63.) The number of degrees of freedom of a set of signals can be defined as the number of Gabor coefficients $G_f^{(w)}(l, m)$ which are not negligibly small for all of the signals in this set, since these non-negligible coefficients are sufficient—to a good degree of approximation—to completely characterize and distinguish a particular signal in this set from the others. The region in the time-frequency plane in which the Gabor coefficients are not negligible may be referred to as the *time-frequency support* of the set of signals. The coefficients corresponding to points lying in this region are, roughly speaking, the time-frequency samples of the signal. The number of these samples corresponds to the number of degrees of freedom of the set of signals.

Notice that since a time-frequency area of unity is associated with each coefficient, the number of degrees of freedom thus defined is also equal to the area of the time-frequency support. If we assume that these non-negligible coefficients lie neatly in a rectangular region of dimensions $\Delta u \times \Delta\mu$, we see that the number of these coefficients is simply $(\Delta u/\delta u)(\Delta\mu/\delta\mu) = \Delta u \Delta\mu$, which is equal to the time-bandwidth product. Thus, when the time-frequency content of a signal is confined to a rectangular region, the number of

degrees of freedom is equal to the time-bandwidth product. However, when the time-frequency content of the signal is not confined to a rectangular region, the actual number of degrees of freedom is less than the time-bandwidth product. In such cases, the two concepts must be clearly distinguished, as will be further discussed below.

Although the above argument has been based on the Gabor expansion, similar arguments are possible with any reasonably well-localized energetic time-frequency distribution. The picture is somewhat more complicated for time-frequency distributions exhibiting interference terms (such as the Wigner distribution); nevertheless similar concepts and arguments are found useful in these cases as well. As above, we may define the time-frequency support of the set of signals we are dealing with, as the time-frequency region outside of which the values of the Wigner distributions are negligible. Alternatively, we may define this region by requiring that the integral of the Wigner distributions over this region should be equal to a certain significant fraction (say 0.95) of the energies of the signals. (Remember that the integral of the Wigner distribution over the whole plane is equal to the energy of the signal.) Throughout this book, when we say that we assume the Wigner distribution of a signal or a set of signals to be confined to a certain region, we will be referring to the time-frequency support and implying that such a significant fraction of the energy or energies are confined to that region (as we already did in figure 2.3). A recent paper dealing in a more rigorous manner with the number of degrees of freedom concept in the context of time-frequency representations is Landau 1993. An older reference dealing with related concepts is Vakman 1968.

Thus in the general case, we define the number of degrees of freedom as the time-frequency support. In general, this will be smaller than the the time-bandwidth product, unless the time-frequency support is a rectangle perpendicular to the time-frequency axes (figure 2.4) (Lohmann and others 1996a). Quite commonly the time-bandwidth product is simply taken to be equal to the number of degrees of freedom of a set of signals, without regard to the shape of the time-frequency support. As we have seen, this may overstate the number of degrees of freedom of the set of signals in question. A simple analogy may be useful. Consider the set of points in three-dimensional space which are confined to some particular plane. The elements of the coordinate vector $\mathbf{r} = (x, y, z)$ will each assume values over the complete interval $[-\infty, \infty]$, but not independently so; the number of degrees of freedom is 2 and not 3. A particularly striking example is the set of chirp signals $A\,\mathrm{chirp}(u)$ whose single degree of freedom is represented by the number $A$. However, both the temporal extent and the spectral extent of these signals are infinite.

It is important to note that the number of degrees of freedom is an approximate concept in the sense that the number of degrees of freedom will depend on the amplitude accuracy we are working with, since this is what determines what is negligible and what is non-negligible. If we are working with greater accuracy, then a larger region in the $u$-$\mu$ plane will have Gabor coefficients which are non-negligible and the signal and its Fourier transform will have non-negligible values over larger intervals. In the case of signals with approximately rectangular time-frequency content, the values of $\Delta u$ and $\Delta\mu$ outside of

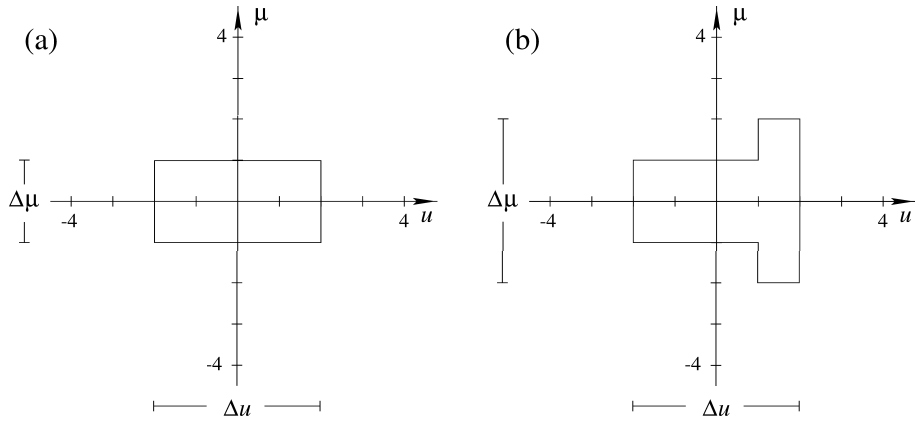


Figure 2.4: (a) Rectangular time-frequency support. The time-frequency area is equal to the time-bandwidth product $\Delta u \Delta \mu = 8$. (b) Irregular time-frequency support. The time-frequency area is 10, which is smaller than the time-bandwidth product $\Delta u \Delta \mu = 16$.

which these signals are negligible in the time and frequency domains respectively, again depend on the accuracy we are working with. The important point is to choose the smallest intervals outside of which the signals are truly negligible with that accuracy. Of course, if the values of the signals are negligible outside an interval of length $\Delta u$, they will also be negligible outside a larger interval. However, unless we choose the smallest possible values of $\Delta u$ and $\Delta \mu$, we will be overstating the number of degrees of freedom of the set of signals.

Until now, we spoke of the temporal and spectral extent and time-bandwidth product of signals without discussing the origin of these finite extents. In the real world a signal is always represented in some physical form in some physical system. These systems which carry or process the signals always limit their temporal duration (or spatial extent) and bandwidths to certain finite values. A physical system cannot allow the existence of frequencies outside a certain band because there is always some limit to the resolution that can be supported. Likewise, since all physical events of interest have a beginning and an end, or because all physical events or systems have a finite extent, the temporal duration or spatial extent of the signal will also be finite. For instance, a computer display with a certain number of pixels cannot represent an image of greater space-bandwidth product. In an optical system the size of the lenses will limit both the spatial extent of the images that can be dealt with and their spatial bandwidths. It is these physical limitations that determine the temporal (or spatial) and spectral extent of the signals and thus their time-bandwidth product. Just as these may be undesirable physical limitations which limit the performance of the system, they may also be deliberate limitations with the purpose of limiting the set of signals we are dealing with. When a signal previously represented by a system with greater time-bandwidth product is input into a system with smaller

time-bandwidth product, an uninvertible process in which information is lost takes place. (It is important to understand that signals have no physical existence outside of a system. Even the "wire" connecting two "systems," is a system itself. Thus signals always move from one system to another.)

The fact that all physical systems support only a finite time-bandwidth product, means that their effect on signals can be simulated with discrete-time systems with the same degree of accuracy that is inherent in the continuous systems or measurement devices from which the signals originate. Further insight on these matters may be obtained by discussing how the discrete Fourier transform provides an approximation to the continuous Fourier transform. The discrete Fourier transform $F(j)$ of $f(l)$ had been defined in equation 1.173 and is repeated here:

$$F(j) \equiv \frac{1}{\sqrt{N}} \sum_{l=0}^{N-1} f(l) e^{-i2\pi jl/N} \qquad j = 0, 1, \ldots, N-1. \tag{2.62}$$

We shall now see that provided $N$ is chosen to be *at least* equal to the time-bandwidth product of the set of signals we are dealing with, the discrete Fourier transform, which can be efficiently computed on a digital computer using the fast Fourier transform (FFT) algorithm, can be used to obtain a good approximation to the continuous Fourier transform. The approximation improves with increasing $N$.

Let us consider a function $f(u)$ and its Fourier transform $F(\mu)$ and define the periodically replicated functions

$$f_{\rm pr}(u) \equiv \sum_{n=-\infty}^{\infty} f(u - n\,\Delta u), \tag{2.63}$$

$$F_{\rm pr}(\mu) \equiv \sum_{n=-\infty}^{\infty} F(\mu - n\,\Delta\mu), \tag{2.64}$$

where $\Delta u$ and $\Delta\mu$ are arbitrary. It is possible to show that samples of these functions constitute a discrete Fourier transform pair as follows (Papoulis 1977, page 74):

$$F_{\rm pr}(j/\Delta u) = \frac{1}{\Delta\mu} \sum_{l=0}^{\Delta u\Delta\mu - 1} f_{\rm pr}(l/\Delta\mu) \exp(-i2\pi jl/\Delta u\Delta\mu). \tag{2.65}$$

Now, let us assume that a significant fraction of the energy of the signal is confined to the intervals $[-\Delta u/2, \Delta u/2]$ and $[-\Delta\mu/2, \Delta\mu/2]$ in the time and frequency domains respectively. In this case, $f(u) \approx f_{\rm pr}(u)$ and $F(\mu) \approx F_{\rm pr}(\mu)$ in the respective intervals. Thus, $f_{\rm pr}(l/\Delta\mu) \approx f(l/\Delta\mu)$ and $F_{\rm pr}(j/\Delta u) \approx F(j/\Delta u)$. As before, $\Delta u\Delta\mu \equiv N$ will denote the time-bandwidth product. To further simplify we may assume scaling such that $\Delta u = \Delta\mu$ so that $\Delta u = \Delta\mu = 1/\sqrt{N}$. Under these circumstances, we see that the DFT of the samples of a function are the samples of the Fourier transform of the function; or, in other words, the DFT maps the samples of $f(u)$ to the samples of $F(\mu)$. The sampling

interval in the time domain is $1/\Delta\mu$ and the sampling interval in the frequency domain is $1/\Delta u$ and the total number of samples in both domains are $N$.

In the interest of easier interpretation with respect to the continuous case, the discrete Fourier transform is sometimes expressed as

$$F(j) \equiv \frac{1}{\sqrt{N}} \sum_{l=N/2-1}^{N/2} f(l) e^{-i2\pi jl/N} \qquad j = N/2-1, N/2, \ldots, N/2, \tag{2.66}$$

where $N$ is assumed to be even. This is easily seen to be identical to equation 2.62 if we think of $f(l)$ and $F(j)$ as periodic functions.

## 2.4 Linear canonical transforms

The class of linear canonical transforms is a three-parameter class of linear integral transforms which includes Fresnel transforms, fractional Fourier transforms, and simple scaling and chirp multiplication operations, as well as certain other transforms among its members.

Linear canonical transforms have been reinvented or reconsidered by many authors under many different names at different times in different contexts, a fact which we consider a tribute to their ubiquity. They have been referred to as quadratic-phase systems (Bastiaans 1979a), generalized Huygens integrals (Siegman 1986), generalized Fresnel transforms (James and Agarwal 1996, Palma and Bagini 1997), special affine Fourier transforms (Abe and Sheridan 1994a, b), extended fractional Fourier transforms (Hua, Liu, and Li 1997c), and Moshinsky-Quesne transforms (Wolf 1979), among other things.

An excellent and alternative exposition to linear canonical transforms may be found in Wolf 1979 (chapter 9: Construction and properties of canonical transforms). This chapter also contains an account of the history of these transforms. Among the important works in this area we may mention Moshinsky and Quesne 1971; Quesne and Moshinsky 1971; Wolf 1974a, b, 1976; García-Calderón and Moshinsky 1980; and Basu and Wolf 1982. Further references may be found in Wolf 1979 (chapter 9).

### 2.4.1 Definition and properties

The linear canonical transform $f_{\mathbf{M}}(u) = (\mathcal{C}_{\mathbf{M}} f)(u)$ of $f(u)$ with parameter $\mathbf{M}$ is most conveniently defined as

$$(\mathcal{C}_{\mathbf{M}} f)(u) = \int C_{\mathbf{M}}(u, u') f(u')\, du', \tag{2.67}$$

$$C_{\mathbf{M}}(u, u') = A_{\mathbf{M}} \exp\left[i\pi(\alpha u^2 - 2\beta u u' + \gamma u'^2)\right],$$

$$A_{\mathbf{M}} = \sqrt{\beta}\, e^{-i\pi/4},$$

where $\alpha$, $\beta$, and $\gamma$ are real parameters independent of $u$ and $u'$. $\mathcal{C}_{\mathbf{M}}$ is the linear canonical transform operator. The label $\mathbf{M}$ represents the three parameters $\alpha$, $\beta$, and $\gamma$ which com-

pletely specify the transform. As with the Fourier transform, a linear canonical transform can be interpreted both as a system and as a transformation to another representation. In the former case, $f(u)$ is the input and $(\mathcal{C}_{\mathbf{M}}f)(u)$, or simply $\mathcal{C}_{\mathbf{M}}f(u)$, is the output. In the latter case, the transform gives us the representation of the signal $f$ in another "domain." Linear canonical transforms are unitary; that is, the inverse transform kernel is the Hermitian conjugate of the original transform kernel: $\mathcal{C}_{\mathbf{M}}^{-1}(u,u') = \mathcal{C}_{\mathbf{M}}^{*}(u',u)$, or more explicitly

$$\begin{aligned}(\mathcal{C}_{\mathbf{M}}^{-1}f)(u) &= \int C_{\mathbf{M}}^{-1}(u,u')f(u')\,du, \qquad (2.68)\\ C_{\mathbf{M}}^{-1}(u,u') &= A_{\mathbf{M}}^{*}\exp\left[-i\pi(\gamma u^2 - 2\beta uu' + \alpha u'^2)\right],\\ A_{\mathbf{M}}^{*} &= (1/\sqrt{1/\beta}\,)e^{i\pi/4} = \sqrt{-\beta}\,e^{-i\pi/4}.\end{aligned}$$

$1/\sqrt{1/\beta}$ is equal to the complex conjugate of $\sqrt{\beta}$ (see the square root convention on page 2). That the transform given in equation 2.68 is indeed the inverse transform can be verified by confirming the identity

$$\int C_{\mathbf{M}}^{-1}(u,u'')C_{\mathbf{M}}(u'',u')\,du'' = \delta(u-u'), \qquad (2.69)$$

with the help of equation 1.8. The inverse of a linear canonical transform is also a linear canonical transform so that we can write $(\mathcal{C}_{\mathbf{M}}^{-1}f)(u) = (\mathcal{C}_{\mathbf{M}^{-1}}f)(u)$, where $\mathbf{M}^{-1}$ denotes the set of parameters of the inverse transform $\alpha_{\text{inv}} = -\gamma$, $\beta_{\text{inv}} = -\beta$, $\gamma_{\text{inv}} = -\alpha$.

We will now examine the consecutive application (also referred to as composition or concatenation) of two linear canonical transforms with arbitrary parameters. We will start with a signal $f(u)$, apply a transform with the parameters $\alpha_1$, $\beta_1$, $\gamma_1$ and then apply to the result a second transform with the parameters $\alpha_2$, $\beta_2$, $\gamma_2$ to obtain finally a signal $g(u)$:

$$\begin{aligned}g(u) &= \int C_{\mathbf{M}_2}(u,u'')\left[\int C_{\mathbf{M}_1}(u'',u')f(u')\,du'\right]du''\\ &= \int\left[\int C_{\mathbf{M}_2}(u,u'')C_{\mathbf{M}_1}(u'',u')\,du''\right]f(u')\,du', \qquad (2.70)\end{aligned}$$

so that the kernel $h(u,u')$ of the composite operation relating $g(u)$ to $f(u)$ is given by

$$\begin{aligned}g(u) &= \int h(u,u')f(u')\,du', \qquad (2.71)\\ h(u,u') &= \int C_{\mathbf{M}_2}(u,u'')C_{\mathbf{M}_1}(u'',u')\,du''.\end{aligned}$$

Upon evaluating the final integral we find that $h(u,u')$ can be expressed in the form (Wolf 1979, page 387)

$$h(u,u') = \operatorname{sgn}(\beta_1\beta_2/\beta_3)\sqrt{\beta_3}\,e^{-i\pi/4}\exp\left[i\pi(\alpha_3u^2 - 2\beta_3uu' + \gamma_3u'^2)\right], \qquad (2.72)$$

where

$$\begin{aligned}\alpha_3 &= \alpha_2 - \frac{\beta_2^2}{\alpha_1+\gamma_2},\\ \beta_3 &= \frac{\beta_1\beta_2}{\alpha_1+\gamma_2},\\ \gamma_3 &= \gamma_1 - \frac{\beta_1^2}{\alpha_1+\gamma_2}.\end{aligned} \tag{2.73}$$

If the factor $\mathrm{sgn}(\beta_1\beta_2/\beta_3)$ did not appear in the result, we could conclude simply that the composition of any two linear canonical transforms is another linear canonical transform. This is not strictly true because of this sign factor. However, we will refer to transforms which differ from equation 2.67 by a minus sign also as linear canonical transforms, so that we can speak of the composition of two linear canonical transforms as being another linear canonical transform. While a more precise formulation is possible, diverting into this technicality would not serve our purpose so that we remain with our present definition to maintain simplicity. (The $\pm 1$ is related to the fact that the class of linear canonical transforms involves a so-called double or twofold cover of the circle, as will be briefly discussed later.)

The composition we have just examined is not in general commutative; that is, in general $\mathcal{C}_{\mathbf{M}_1}\mathcal{C}_{\mathbf{M}_2} \neq \mathcal{C}_{\mathbf{M}_2}\mathcal{C}_{\mathbf{M}_1}$. However, such compositions are associative; that is

$$(\mathcal{C}_{\mathbf{M}_1}\mathcal{C}_{\mathbf{M}_2})\mathcal{C}_{\mathbf{M}_3} = \mathcal{C}_{\mathbf{M}_1}(\mathcal{C}_{\mathbf{M}_2}\mathcal{C}_{\mathbf{M}_3}). \tag{2.74}$$

Until now, we let the symbol $\mathbf{M}$ denote the three parameters $\alpha$, $\beta$, $\gamma$ characterizing a linear canonical transform. Now, we will more specifically define $\mathbf{M}$ as a matrix of the form

$$\mathbf{M} \equiv \begin{bmatrix} A & B \\ C & D \end{bmatrix} \equiv \begin{bmatrix} \gamma/\beta & 1/\beta \\ -\beta+\alpha\gamma/\beta & \alpha/\beta \end{bmatrix} = \begin{bmatrix} \alpha/\beta & -1/\beta \\ \beta-\alpha\gamma/\beta & \gamma/\beta \end{bmatrix}^{-1}, \tag{2.75}$$

with determinant $AD-BC=1$. (Such matrices are called unit-determinant or unimodular matrices.) The matrix elements are fully equivalent to the three independent parameters $\alpha$, $\beta$, and $\gamma$, which can be recovered in terms of the matrix elements as follows

$$\begin{aligned}\alpha &= \frac{D}{B} = \frac{1}{A}\left(\frac{1}{B}+C\right),\\ \beta &= \frac{1}{B},\\ \gamma &= \frac{A}{B} = \frac{1}{D}\left(\frac{1}{B}+C\right).\end{aligned} \tag{2.76}$$

The reason why we define $\mathbf{M}$ in this manner, rather than simply as a parameter vector $[\alpha\ \beta\ \gamma]$, is because of a number of attractive properties of this matrix as defined. The primary rationale behind the definition of this matrix is that the matrix corresponding

to the composition of two systems is the matrix product of the matrices corresponding to the individual systems. That is,

$$\mathbf{M}_3 = \mathbf{M}_2\mathbf{M}_1, \tag{2.77}$$

a result which can be proved by using equations 2.73. (This result is oblivious to the $\pm 1$ that might appear in front of the kernel.) Furthermore, it is easy to show that the matrix corresponding to the inverse of a transform is the inverse of the matrix corresponding to the original transform, as we have already built into our notation:

$$\mathcal{C}_{\mathbf{M}}^{-1} = \mathcal{C}_{\mathbf{M}^{-1}}, \tag{2.78}$$

$$C_{\mathbf{M}}^{-1}(u,u') = C_{\mathbf{M}^{-1}}(u,u') = C_{\mathbf{M}}^{*}(u',u). \tag{2.79}$$

If desired, the defining equation 2.67 can be rewritten in terms of the matrix parameters as follows:

$$f_{\mathbf{M}}(u) = (\mathcal{C}_{\mathbf{M}}f)(u) = \int C_{\mathbf{M}}(u,u')f(u')\,du', \tag{2.80}$$

$$C_{\mathbf{M}}(u,u') = A_{\mathbf{M}} \exp\left[i\pi\left(\frac{D}{B}u^2 - 2\frac{1}{B}uu' + \frac{A}{B}u'^2\right)\right],$$

$$A_{\mathbf{M}} = \sqrt{1/B}\, e^{-i\pi/4}.$$

The set of linear canonical transforms satisfy all the axioms of a noncommutative group (closure, associativity, existence of identity, inverse of each element), just like the set of all unit-determinant $2 \times 2$ matrices. (Again, this is true to the extent that we are willing to be flexible with signs in front of the transform integrals.) Certain subsets (or subclasses) of the set (class) of linear canonical transforms are groups in themselves and thus are subgroups. Several of them will be discussed further below. For example, we will see that the fractional Fourier transform is a subgroup with one real parameter. Integer powers of the Fourier transform are a subgroup with one integer parameter.

A rather trivial extension of linear canonical transforms are transforms which include not only quadratic terms such as $u^2$, $u'^2$, $uu'$, but also linear terms such as $u$, $u'$ in the exponent. Any second-order expression including such terms can be expressed as a quadratic form of $(u-\xi)$ and $(u'-\xi')$, where $\xi$ and $\xi'$ are constants. Thus such transforms can be obtained by shifting the input and output of linear canonical transforms as defined above. We will not develop this extension with the understanding that it can be readily introduced when necessary. A more involved extension is to allow elements of the matrix $\mathbf{M}$ to be complex.

Some of the operational properties of linear canonical transforms are listed in table 2.8. Of course, one must add to these all properties associated with unitarity, such as Parseval's relation: $\|f_{\mathbf{M}}(u)\| = \|f(u)\|$.

We also note here the linear canonical transforms of the eigenfunctions of the Fourier transform (Wolf 1979):

$$(\mathcal{C}_{\mathbf{M}}\psi_n)(u) = \left[\left(2\frac{A+iB}{A-iB}\right)^n n!\pi^{1/2}(A+iB)\right]^{-1/2}$$

| | $f(u)$ | $f_{\mathbf{M}}(u)$ |
|---|---|---|
| 1. | $\sum_j \alpha_j f_j(u)$ | $\sum_j \alpha_j f_{j\mathbf{M}}(u)$ |
| 2. | $f(-u)$ | $f_{\mathbf{M}}(-u)$ |
| 3. | $\lvert M\rvert^{-1} f(u/M)$ | $f_{\mathbf{M}'}(u)$ |
| 4. | $f(u-\xi)$ | $\exp[i\pi(2u\xi C - \xi^2 AC)] f_{\mathbf{M}}(u - A\xi)$ |
| 5. | $\exp(i2\pi\xi u) f(u)$ | $\exp[i\pi\xi D(2u - \xi B)] f_{\mathbf{M}}(u - B\xi)$ |
| 6. | $\exp[i\pi(\chi u^2 + 2\xi u)] f(u)$ | $\exp[i\pi\xi D(2u - \xi B)] f_{\mathbf{M}''}(u - B\xi)$ |
| 7. | $u^n f(u)$ | $[Du - B(i2\pi)^{-1} d/du]^n f_{\mathbf{M}}(u)$ |
| 8. | $(i2\pi)^{-n} d^n f(u)/du^n$ | $[-Cu + A(i2\pi)^{-1} d/du]^n f_{\mathbf{M}}(u)$ |
| 9. | $f^*(u)$ | $f^*_{\mathbf{M}^{-1}}(u)$ |
| 10. | $[f(u) + f(-u)]/2$ | $[f_{\mathbf{M}}(u) + f_{\mathbf{M}}(-u)]/2$ |
| 11. | $[f(u) - f(-u)]/2$ | $[f_{\mathbf{M}}(u) - f_{\mathbf{M}}(-u)]/2$ |

Table 2.8: Properties of linear canonical transforms. The expressions on the right are linear canonical transforms of the expressions on the left. $\alpha_j$ are arbitrary complex constants, $M, \xi, \chi$ are real, and $n$ is a positive integer. $\mathbf{M}'$ is the matrix that corresponds to the parameters $\alpha' = \alpha$, $\beta' = M\beta$, $\gamma' = M^2\gamma$ and $\mathbf{M}''$ is the matrix that corresponds to the parameters $\alpha'' = \alpha$, $\beta'' = \beta$, $\gamma'' = \gamma + \chi$.

$$\times \exp\left(-\frac{D - iC}{A + iB}\,\pi u^2\right) H_n\left[(A^2 + B^2)^{-1/2}\sqrt{2\pi}\,u\right]. \tag{2.81}$$

### 2.4.2 Effect on Wigner distributions

We now discuss the effect of a linear canonical transform on the Wigner distribution or ambiguity function of a signal. We let $f$ denote a signal and $f_{\mathbf{M}}$ its linear canonical transform, where $\mathbf{M}$ is the unit-determinant matrix of coefficients characterizing the transform. Then,

$$W_{f_{\mathbf{M}}}(Au + B\mu, Cu + D\mu) = W_f(u, \mu), \tag{2.82}$$

$$W_{f_{\mathbf{M}}}(u, \mu) = W_f(Du - B\mu, -Cu + A\mu). \tag{2.83}$$

That is, the Wigner distributions of the transformed signal is simply a linearly distorted form of the Wigner distribution of the original signal. This result can be demonstrated directly from the definition of linear canonical transforms and the definition of the Wigner distribution, although the algebra is somewhat involved. This distortion can also be interpreted as a coordinate transformation to the new axes $u_{\text{new}}$, $\mu_{\text{new}}$ defined in terms of the old $u$, $\mu$ as

$$\begin{aligned} u_{\text{new}} &= Au + B\mu, \\ \mu_{\text{new}} &= Cu + D\mu. \end{aligned} \tag{2.84}$$

In this interpretation, we see that the Wigner distribution of the transformed signal is of the same functional form $W_f(\cdot,\cdot)$ with respect to the newly defined axes: $W_f(u_{\text{new}}, \mu_{\text{new}})$. The Jacobian of the above two-dimensional coordinate transformation is simply the determinant of $\mathbf{M}$, which is by definition equal to unity. Thus this transformation is area preserving. It distorts, but does not concentrate or deconcentrate the Wigner distribution. For instance, consider the time-frequency support of a signal, defined as the time-frequency region which contains a certain significant percentage of the signal energy. Then, the similarly defined region corresponding to the transformed signal, containing the same percentage of the signal energy, will have the same support area. That is, the time-frequency area in which this percentage of the signal energy is contained remains invariant under a linear canonical transform. Figure 2.5 illustrates these concepts for a Wigner distribution with approximately rectangular time-frequency support.

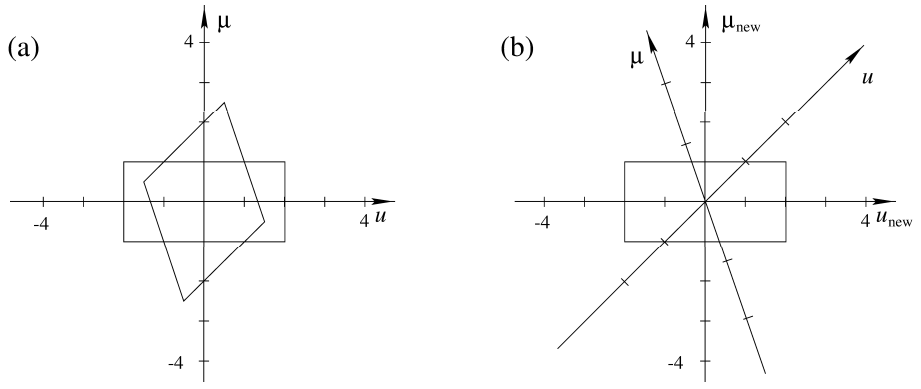


Figure 2.5: (a) Rectangular time-frequency support of the Wigner distribution of $f(u)$ and the time-frequency support of the Wigner distribution of its linear canonical transform $f_{\mathbf{M}}(u)$. (b) Time-frequency support of the Wigner distribution of $f_{\mathbf{M}}(u)$ with respect to the axes $u_{\text{new}}$ and $\mu_{\text{new}}$, which is seen to be of the same rectangular form as the support of the original Wigner distribution with respect to the axes $u$ and $\mu$.

Remembering that the Wigner distribution gives us the distribution of signal energy over time and frequency, the time-frequency area preserving nature of linear canonical transforms means that such transforms do not concentrate or deconcentrate energy in the time-frequency plane. (Since linear canonical transforms are unitary, the total signal energy is conserved to begin with.) This property has many interpretations in physics, some of which we will discuss in chapter 4. For the time being, we note that since the area of the time-frequency support for a set of signals can be interpreted as the number of degrees of freedom, conservation of this area also implies conservation of the number of degrees of freedom and thus information, which is consistent with the fact that linear canonical transforms are invertible.

The results just mentioned are of sufficient importance to warrant an alternative treat-

ment (Bastiaans 1979a). Earlier, we had presented the general form of the kernel transforming the Wigner distribution for any given linear transform (equation 2.44). Specializing this to linear canonical transforms we find that

$$K_{\mathbf{M}}(u,\mu;u',\mu') = \beta\delta[\mu - (\alpha u - \beta u')]\delta[\mu' - (\beta u - \gamma u')]. \tag{2.85}$$

The value of the Wigner distribution $W_f(u,\mu)$ at a certain time-frequency point is mapped to another point which is determined by setting the arguments of the delta functions equal to zero:

$$\begin{bmatrix} \mu \\ \mu' \end{bmatrix} = \begin{bmatrix} \alpha & -\beta \\ \beta & -\gamma \end{bmatrix} \begin{bmatrix} u \\ u' \end{bmatrix}. \tag{2.86}$$

This equation can be algebraically rewritten as

$$\begin{bmatrix} u' \\ \mu' \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} \begin{bmatrix} u \\ \mu \end{bmatrix}, \tag{2.87}$$

where

$$A = \gamma/\beta, \qquad B = 1/\beta, \qquad C = -\beta + \alpha\gamma/\beta, \qquad D = \alpha/\beta, \tag{2.88}$$

just as defined previously in equation 2.75. In the light of these equations, the kernel can be rewritten as

$$K_{\mathbf{M}}(u,\mu;u'\mu') = \delta(u' - Du + B\mu)\,\delta(\mu' + Cu - A\mu). \tag{2.89}$$

This can also be derived directly by substituting equation 2.67 in equation 2.44. This form of the kernel clearly indicates the pointwise mapping involved. The value of the Wigner distribution at time-frequency point $(u,\mu)$ is mapped into the new time-frequency point $(u_{\text{new}},\mu_{\text{new}}) = (Au + B\mu, Cu + D\mu)$ so that

$$W_{f_{\mathbf{M}}}(u,\mu) = W_f(Du - B\mu, -Cu + A\mu), \tag{2.90}$$

$$W_{f_{\mathbf{M}}}(Au + B\mu, Cu + D\mu) = W_f(u,\mu). \tag{2.91}$$

These equations are the same as equations 2.82 through 2.84.

To summarize, when a function undergoes a linear canonical transform, its Wigner distribution undergoes a pointwise geometrical distortion or deformation: the value of the Wigner distribution at each time-frequency point is mapped to another time-frequency point. (We will see in chapter 4 that these points can be interpreted as optical rays.) It is often easier to visualize this by concentrating on the boundary of the region to which the Wigner distribution is approximately confined; the region in which most of the energy of the signal lies. However, the area preserving nature of the distortion is more general. For instance, let us concentrate on any particular closed contour of the Wigner distribution, which does not necessarily contain a large percentage of the signal energy. The region defined by such a contour will be distorted and deformed but its area will remain the same regardless of the parameters of the linear canonical transform. Even more generally,

if we take any region R in $(u, \mu)$ space, and find the image of this region under a linear canonical transform, their areas will always be the same. (If R is a given region in $(u, \mu)$ space, then its image is simply the set of points $(Au+B\mu, Cu+D\mu)$ such that $(u, \mu) \in$ R.)

Until now we limited our attention to the Wigner distribution. The ambiguity function is affected in a similar way. We remember that the ambiguity function is essentially the two-dimensional Fourier transform of the Wigner distribution (equation 2.34). Thus, the two-dimensional Fourier transform property given in equation 1.296 allows us to show that the ambiguity function of the linear canonical transform of a function is related to that of the original function through the relation

$$A_{f_{\mathbf{M}}}(\bar{u}, \bar{\mu}) = A_f(D\bar{u} - B\bar{\mu}, -C\bar{u} + A\bar{\mu}). \tag{2.92}$$

More generally, time-frequency distributions of the Cohen class will exhibit a similar distortion property if the kernel satisfies $\psi_{TFE}(u, \mu) = \psi_{TFE}(Du - B\mu, -Cu + A\mu)$ for all $u, \mu$ and all $A, B, C, D$ satisfying $AD - BC = 1$ (Ozaktas, Erkaya, and Kutay 1996).

### 2.4.3 Special linear canonical transforms

We now consider a number of important special cases of linear canonical transforms. We will see that most of these are operations that we have already defined in table 1.2. We will present the forward kernel $C_{\mathbf{M}}(u, u')$, the inverse kernel $C_{\mathbf{M}}^{-1}(u, u') = C_{\mathbf{M}^{-1}}(u, u')$, the result of the transform $f_{\mathbf{M}}(u)$, the matrix $\mathbf{M}$, and the Wigner distribution of the transformed signal $W_{f_{\mathbf{M}}}(u, \mu)$.

First, we consider simple scaling which is characterized by

$$M_M(u, u') = \sqrt{M}\, \delta(u - Mu'), \tag{2.93}$$

$$M_M^{-1}(u, u') = M_{(1/M)}(u, u') = \sqrt{1/M}\, \delta(u - u'/M), \tag{2.94}$$

$$\mathcal{C}_{\mathbf{M}_M} f(u) = \mathcal{M}_M f(u) = \sqrt{1/M}\, f(u/M), \tag{2.95}$$

$$\mathbf{M}_M = \begin{bmatrix} M & 0 \\ 0 & 1/M \end{bmatrix} = \begin{bmatrix} 1/M & 0 \\ 0 & M \end{bmatrix}^{-1}, \tag{2.96}$$

$$W_{\mathcal{M}_M f}(u, \mu) = W_f(u/M, M\mu), \tag{2.97}$$

where $M > 0$. Here $\mathbf{M}_M$ is the transform matrix corresponding to the scaling operation $\mathcal{M}_M$ with kernel $M_M(u, u')$. (It is unfortunate that we use $\mathbf{M}$ also to denote a generic transform matrix, and that we use $M$ to denote both the functional form of the kernel and its parameter.) The effect on the Wigner distribution is illustrated in figure 2.6b. To see that this is indeed a special case of equation 2.67 requires some care. First notice that $A = M = 1/D$ and $B = 0 = C$ implies $M\alpha = \beta = \gamma/M \to \infty$ from equations 2.76. Now, the desired results can be obtained by virtue of equation 1.4. The identity operation is a further special case of the scaling operation with $M = 1$.

We now turn to chirp multiplication for which

$$Q_q(u,u') = e^{-i\pi q u^2}\delta(u-u'), \tag{2.98}$$

$$Q_q^{-1}(u,u') = Q_{-q}(u,u') = e^{i\pi q u^2}\delta(u-u'), \tag{2.99}$$

$$\mathcal{C}_{\mathbf{Q}_q} f(u) = \mathcal{Q}_q f(u) = e^{-i\pi q u^2} f(u), \tag{2.100}$$

$$\mathbf{Q}_q = \begin{bmatrix} 1 & 0 \\ -q & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ q & 1 \end{bmatrix}^{-1}, \tag{2.101}$$

$$W_{\mathcal{Q}_q f}(u,\mu) = W_f(u,\mu+qu). \tag{2.102}$$

Here $\mathbf{Q}_q$ is the transform matrix corresponding to the chirp multiplication operation $\mathcal{Q}_q$ with kernel $Q_q(u,u')$. Chirp multiplication is characterized by a lower triangular matrix. The effect on the Wigner distribution is a special case of property 15 in table 2.2 (figure 2.6c). It is also of some interest to note that the matrix $\mathbf{Q}_q$ can be written as $\mathbf{Q}_q = \exp(-q\,\mathbf{lower})$ where **lower** is a $2\times 2$ matrix [0 0; 1 0].

The dual of chirp multiplication is chirp convolution for which

$$R_r(u,u') = e^{-i\pi/4}\sqrt{1/r}\,\exp[i\pi(u-u')^2/r], \tag{2.103}$$

$$R_r^{-1}(u,u') = R_{-r}(u,u') = e^{i\pi/4}(1/\sqrt{r}\,)\exp[-i\pi(u-u')^2/r], \tag{2.104}$$

$$\mathcal{C}_{\mathbf{R}_r} f(u) = \mathcal{R}_r f(u) = f(u) * e^{-i\pi/4}\sqrt{1/r}\,\exp(i\pi u^2/r), \tag{2.105}$$

$$\mathbf{R}_r = \begin{bmatrix} 1 & r \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -r \\ 0 & 1 \end{bmatrix}^{-1}, \tag{2.106}$$

$$W_{\mathcal{R}_r f}(u,\mu) = W_f(u-r\mu,\mu). \tag{2.107}$$

Here $\mathbf{R}_r$ is the transform matrix corresponding to the chirp convolution operation $\mathcal{R}_r$ with kernel $R_r(u,u')$. The effect on the Wigner distribution is a special case of property 14 in table 2.2 (figure 2.6d). Chirp convolution is characterized by an upper triangular matrix. It is also of some interest to note that the matrix $\mathbf{R}_q$ can be written as $\mathbf{R}_q = \exp(r\,\mathbf{upper})$ where **upper** is a $2\times 2$ matrix [0 1; 0 0].

The geometrical effect of chirp multiplication and chirp convolution on the Wigner distribution, or more precisely the support of the Wigner distribution, is referred to as "shearing." This geometrical distortion is a special case of that described by equation 2.82, and like all distortions described by this equation, is area preserving. This is easily seen to be the case by considering the effect of shearing on a rectangle. Shearing may be viewed as cutting this rectangle into narrow strips and sliding these with respect to each other (figure 2.7).

We will see later that a general area-preserving distortion, characterized by a unit-determinant matrix, can be decomposed in terms of these two kinds of area-preserving vertical and horizontal shearing operations. Also note that shearing the Wigner distribution in a certain direction will not change the projection of the Wigner distribution (its Radon transform) along that direction, a fact which is also evident from figure 2.7. (As

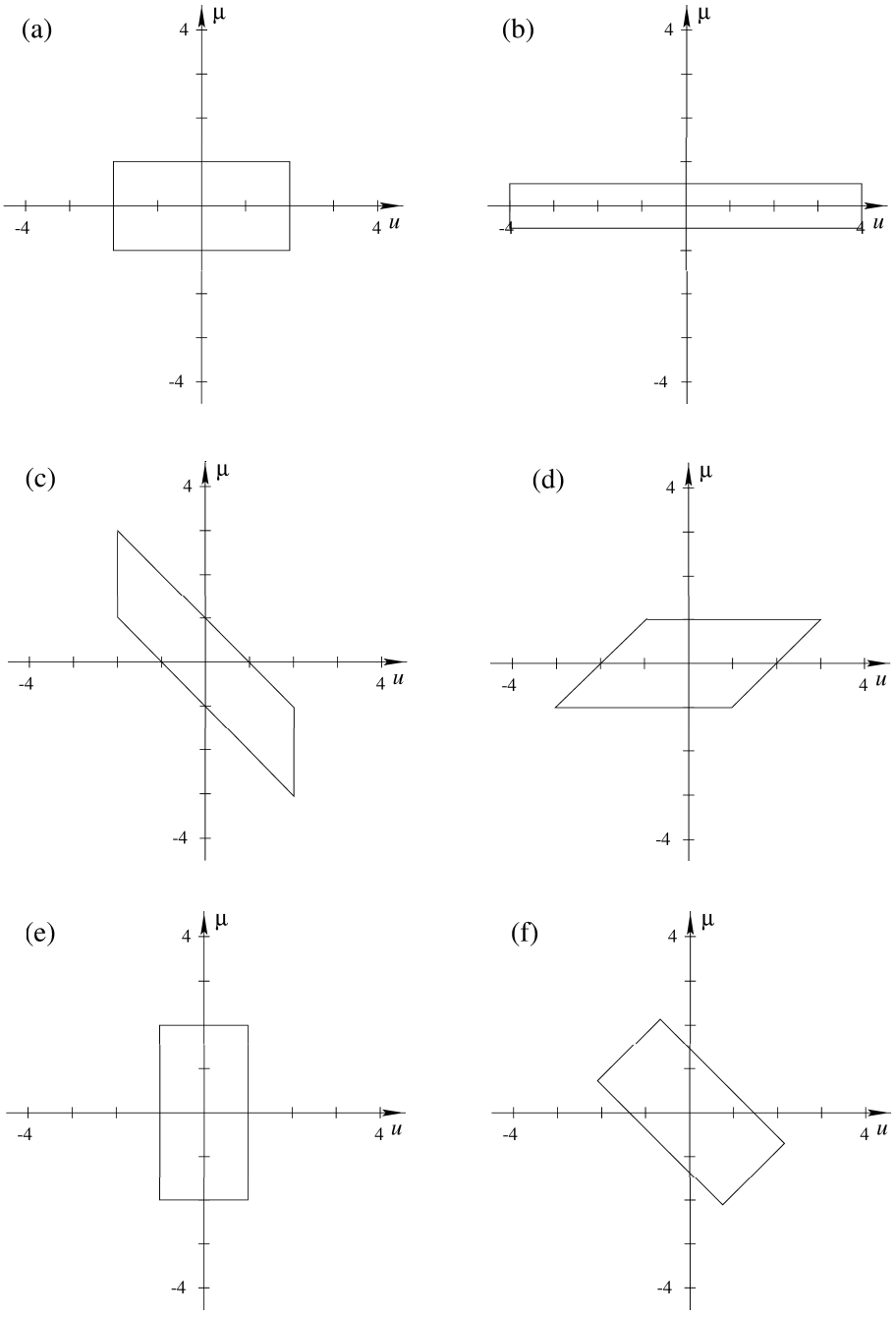


Figure 2.6: (a) Region representing the support of a signal. Effect of (b) $\mathcal{M}_2$, (c) $\mathcal{Q}_1$, (d) $\mathcal{R}_1$, (e) $\mathcal{F}$, (f) $\mathcal{F}^{0.5}$.

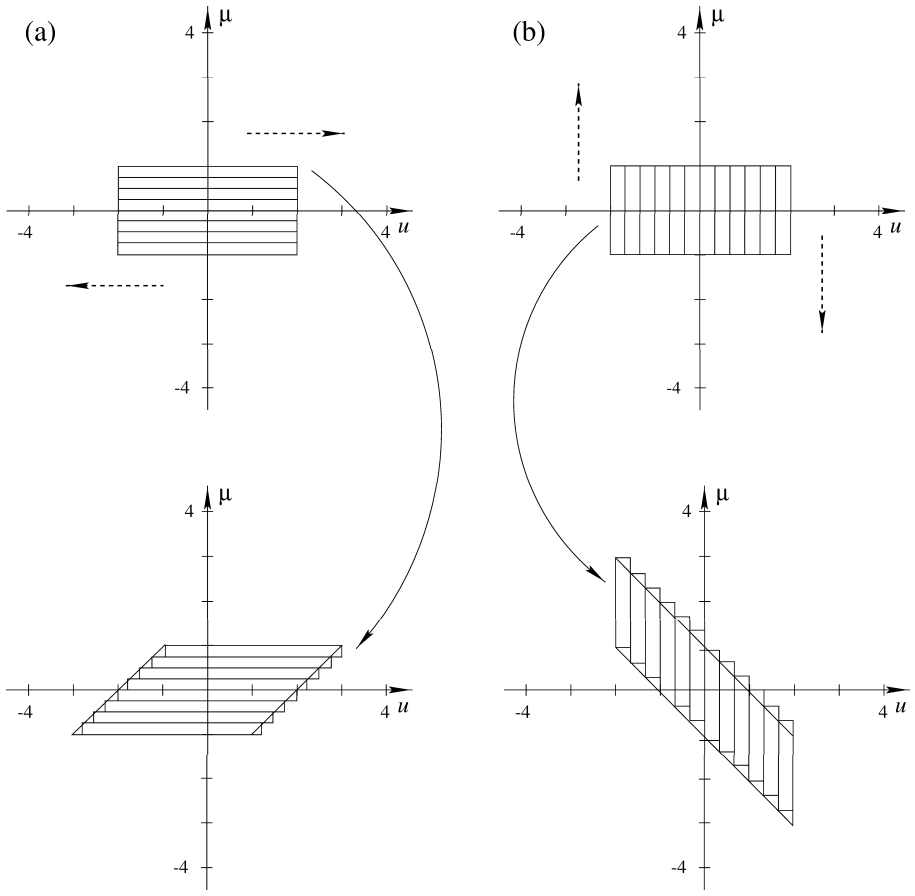


Figure 2.7: Horizontal (a) and vertical (b) shearing as sliding of narrow strips.

an additional exercise, the reader may want to show that such general area-preserving distortions always map a bundle of parallel lines to another bundle of parallel lines, and ellipses to ellipses.)

The Fourier transform is also a special case with

$$F_{\mathrm{lc}}(u,u') = e^{-i\pi/4}e^{-i2\pi uu'}, \tag{2.108}$$

$$F_{\mathrm{lc}}^{-1}(u,u') = F_{\mathrm{lc}}^{*}(u,u') = e^{i\pi/4}e^{i2\pi uu'}, \tag{2.109}$$

$$\mathcal{C}_{\mathbf{F}_{\mathrm{lc}}}f(u) = \mathcal{F}_{\mathrm{lc}}f(u) = e^{-i\pi/4}\int f(u)e^{-i2\pi uu'}\,du', \tag{2.110}$$

$$\mathbf{F}_{\mathrm{lc}} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}^{-1}, \tag{2.111}$$

$$W_{\mathcal{F}_{\mathrm{lc}}f}(u,\mu) = W_f(-\mu,u). \tag{2.112}$$

The effect on the Wigner distribution is seen to be property 16 in table 2.2 (figure 2.6e). The factor $\exp(-i\pi/4)$ is not normally a part of the definition of the Fourier transform.

However, it is this form which is the special case of linear canonical transforms (Wolf 1979). This usually does not cause any trouble since it differs from the standard definition only by a unit-magnitude constant. The reader may also have noticed that $\mathcal{F}_{\text{lc}}$ as defined above satisfies $\mathcal{F}_{\text{lc}}^4 = -\mathcal{I}$ rather than $\mathcal{F}^4 = \mathcal{I}$ satisfied by the standard definition. This fact is related to the sign factor discussed on page 90. (Similar comments also apply to the fractional Fourier transform introduced below.)

Note that all of these special cases, with the exception of the Fourier transform, are one-parameter subgroups of the group of linear canonical transforms. Another further such group is the fractional Fourier transform, characterized by the matrix

$$\mathbf{F}_{\text{lc}}^a = \begin{bmatrix} \cos(a\pi/2) & \sin(a\pi/2) \\ -\sin(a\pi/2) & \cos(a\pi/2) \end{bmatrix} = \begin{bmatrix} \cos(a\pi/2) & -\sin(a\pi/2) \\ \sin(a\pi/2) & \cos(a\pi/2) \end{bmatrix}^{-1}, \tag{2.113}$$

and the rotational time-frequency distortion (figure 2.6f)

$$W_{\mathcal{F}_{\text{lc}}^a f}(u,\mu) = W_f\left[\cos(a\pi/2)\,u - \sin(a\pi/2)\,\mu, \sin(a\pi/2)\,u + \cos(a\pi/2)\,\mu\right]. \tag{2.114}$$

This one-parameter subgroup is also referred to as the *elliptic subgroup* (Wolf 1979). It is also of some interest to note that the matrix $\mathbf{F}_{\text{lc}}^a$ can be written as $\mathbf{F}_{\text{lc}}^a = \exp[(a\pi/2)\mathbf{F}_{\text{lc}}]$ where $\mathbf{F}_{\text{lc}}$ is the $2 \times 2$ matrix $[0\ 1; -1\ 0]$. We emphasize that here $\mathbf{F}$ and $\mathbf{F}^a$ denote the $2 \times 2$ linear canonical transform matrices associated with the continuous Fourier and fractional Fourier transforms; they do *not* denote the discrete Fourier and fractional Fourier transform matrices for which these symbols are used elsewhere in this book.

Another one-parameter subgroup which we will not further discuss has the matrix

$$\mathbf{HYP}_a = \begin{bmatrix} \cosh(a\pi/2) & \sinh(a\pi/2) \\ \sinh(a\pi/2) & \cosh(a\pi/2) \end{bmatrix} = \begin{bmatrix} \cosh(a\pi/2) & -\sinh(a\pi/2) \\ -\sinh(a\pi/2) & \cosh(a\pi/2) \end{bmatrix}^{-1} \tag{2.115}$$

and is referred to as the *hyperbolic subgroup*. We might also add for completeness that the one-parameter subgroup corresponding to the scaling operation is sometimes written with $M = \exp(a\pi/2)$ and referred to as the *parabolic subgroup* (Wolf 1979).

We will briefly mention a number of additional special cases. First we consider the transform characterized by an arbitrary unit-determinant lower triangular matrix

$$\begin{bmatrix} M & 0 \\ -qM & 1/M \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -q & 1 \end{bmatrix} \begin{bmatrix} M & 0 \\ 0 & 1/M \end{bmatrix}, \tag{2.116}$$

which we have written as the composition of a scaling operation followed by a chirp multiplication, so that the effect of the linear canonical transform corresponding to this matrix is to take $f(u)$ to

$$e^{-i\pi q u^2} \sqrt{1/M}\, f(u/M). \tag{2.117}$$

Also, we may consider the transform characterized by an arbitrary unit-determinant upper triangular matrix

$$\begin{bmatrix} M & r/M \\ 0 & 1/M \end{bmatrix} = \begin{bmatrix} 1 & r \\ 0 & 1 \end{bmatrix} \begin{bmatrix} M & 0 \\ 0 & 1/M \end{bmatrix}, \tag{2.118}$$

which we have written as the composition of a scaling operation followed by a chirp convolution, so that the effect of the linear canonical transform corresponding to this matrix is to take $f(u)$ to

$$e^{-i\pi/4}\sqrt{\frac{1}{r}}\,e^{i\pi u^2/r} * \sqrt{1/M}\,f(u/M). \tag{2.119}$$

Matrices of the remaining two triangular forms can be decomposed as follows:

$$\begin{bmatrix} -r/M & M \\ -1/M & 0 \end{bmatrix} = \begin{bmatrix} 1 & r \\ 0 & 1 \end{bmatrix}\begin{bmatrix} M & 0 \\ 0 & 1/M \end{bmatrix}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \tag{2.120}$$

$$\begin{bmatrix} 0 & M \\ -1/M & -qM \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -q & 1 \end{bmatrix}\begin{bmatrix} M & 0 \\ 0 & 1/M \end{bmatrix}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{2.121}$$

The second of these corresponds to a scaled Fourier transform with a residual phase factor.

Lastly we mention that other simple operations, such as coordinate shifting or multiplication by a phase factor, also become special cases of linear canonical transforms if linear terms are allowed in the exponent of the kernel (as discussed on page 91).

### 2.4.4 Decompositions

We have already given above a number of decompositions for triangular matrices. Here we list a number of further decompositions for the general case:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & (A-1)/C \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ C & 1 \end{bmatrix}\begin{bmatrix} 1 & (D-1)/C \\ 0 & 1 \end{bmatrix} \tag{2.122}$$

$$= \begin{bmatrix} 1 & 0 \\ (D-1)/B & 1 \end{bmatrix}\begin{bmatrix} 1 & B \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ (A-1)/B & 1 \end{bmatrix} \tag{2.123}$$

$$= \begin{bmatrix} 1 & 0 \\ C/A & 1 \end{bmatrix}\begin{bmatrix} A & 0 \\ 0 & 1/A \end{bmatrix}\begin{bmatrix} 1 & B/A \\ 0 & 1 \end{bmatrix} \tag{2.124}$$

$$= \begin{bmatrix} 1 & B/D \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1/D & 0 \\ 0 & D \end{bmatrix}\begin{bmatrix} 1 & 0 \\ C/D & 1 \end{bmatrix} \tag{2.125}$$

$$= \begin{bmatrix} B & 0 \\ D & 1/B \end{bmatrix}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ A/B & 1 \end{bmatrix} \tag{2.126}$$

$$= \begin{bmatrix} 1 & 0 \\ D/B & 1 \end{bmatrix}\begin{bmatrix} B & 0 \\ 0 & 1/B \end{bmatrix}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ A/B & 1 \end{bmatrix} \tag{2.127}$$

$$= \begin{bmatrix} -1/C & -A \\ 0 & -C \end{bmatrix}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} 1 & D/C \\ 0 & 1 \end{bmatrix} \tag{2.128}$$

$$= \begin{bmatrix} 1 & A/C \\ 0 & 1 \end{bmatrix}\begin{bmatrix} -1/C & 0 \\ 0 & -C \end{bmatrix}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} 1 & D/C \\ 0 & 1 \end{bmatrix}. \tag{2.129}$$

Care must be exercised when a term appearing in the denominator of any of the matrix elements is zero. The decompositions will remain valid if the limits are evaluated carefully. Some of these decompositions are from Nazarathy and Shamir 1982a.

These decompositions are very easy to demonstrate when stated as matrix products in the above manner. The reader should nevertheless see how these decompositions work out in terms of integral transforms to gain familiarity with them. Remember that lower triangular matrices correspond to chirp multiplication, upper triangular matrices to chirp convolution, diagonal matrices to scaling, and skew diagonal matrices to Fourier transforming. Thus, these decompositions show the many ways in which a general linear canonical transform can be decomposed in terms of these more elementary operations. The decompositions given in equations 2.122 and 2.123 are referred to as canonical decompositions type I and type II, respectively (Papoulis 1974, 1977). (These can be obtained from each other by taking the transpose of both sides of these decompositions.) It is also useful to interpret these decompositions in terms of "shear diagrams," which show how the overall distortion of the Wigner distribution can be broken down into simpler shearing operations. We illustrate this concept with equation 2.122 in figure 2.8.

Equations 2.122 and 2.123 show how any unit-determinant matrix can be written as the product of lower and upper triangular matrices. We have already seen that these lower and upper triangular matrices (which correspond to chirp multiplication and convolution) geometrically correspond to vertical and horizontal shears. Since it is easy to see that these shears preserve area, and since the distortion associated with any unit determinant matrix can be decomposed into such shears, it follows that such distortions are also area preserving. As previously stated, this result also follows more directly from the facts that the determinant of the matrix is the Jacobian of the geometric distortion and that distortions with unity Jacobian are area preserving.

It is also possible to write equations 2.122 through 2.129 in operator form as follows:

$$\mathcal{R}_{(A-1)/C}\ \mathcal{Q}_{-C}\ \mathcal{R}_{(D-1)/C}, \tag{2.130}$$

$$\mathcal{Q}_{(1-D)/B}\ \mathcal{R}_{B}\ \mathcal{Q}_{(1-A)/B}, \tag{2.131}$$

$$\mathcal{Q}_{-C/A}\ \mathcal{M}_{A}\ \mathcal{R}_{B/A}, \tag{2.132}$$

$$\mathcal{R}_{B/D}\ \mathcal{M}_{1/D}\ \mathcal{Q}_{-C/D}, \tag{2.133}$$

$$(\mathcal{Q}_{-D/B}\ \mathcal{M}_{B})\ \mathcal{F}\ \mathcal{Q}_{-A/B}, \tag{2.134}$$

$$\mathcal{Q}_{-D/B}\ \mathcal{M}_{B}\ \mathcal{F}\ \mathcal{Q}_{-A/B}, \tag{2.135}$$

$$(\mathcal{R}_{A/C}\ \mathcal{M}_{-1/C})\ \mathcal{F}\ \mathcal{R}_{D/C}, \tag{2.136}$$

$$\mathcal{R}_{A/C}\ \mathcal{M}_{-1/C}\ \mathcal{F}\ \mathcal{R}_{D/C}. \tag{2.137}$$

The last two pairs have been repeated so as to maintain parallelism with the list of matrix decompositions. These formulas are analogous to what are known as Baker-Campbell-Hausdorff formulas (see the discussion surrounding equation 1.245).

At this point, the reader should be able to interpret any decomposition in all of the following ways and understand the relationships between these interpretations:

- Decomposition into matrices.
- Decomposition into consecutive integral transforms.

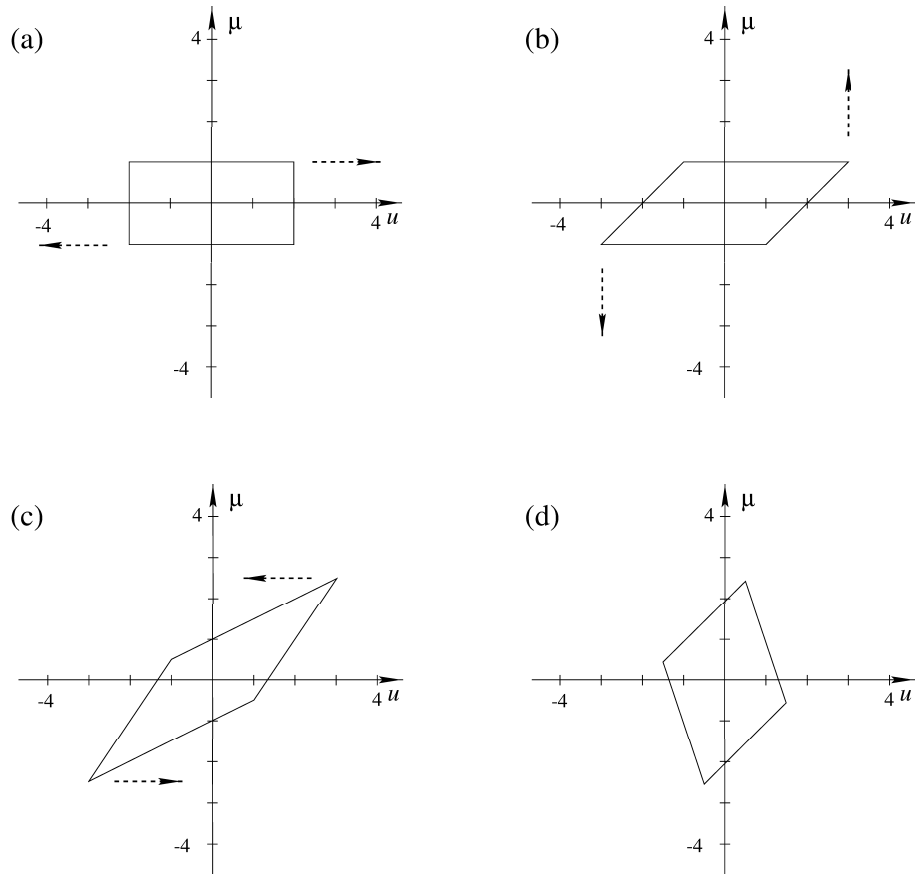


Figure 2.8: (a) Time-frequency support of the Wigner distribution of a signal. (b) Result of first horizontal shear. (c) Result of vertical shear. (d) Result of second horizontal shear, which is the time-frequency support of the linear canonical transform of the signal. $A = 0.5$, $B = -0.5$, $C = 0.5$, $D = 1.5$.

- Decomposition into abstract operators.
- Decomposition into chirp multiplications and chirp convolutions, and scaling operations and Fourier transforms.
- Decomposition of the overall geometric distortion of the time-frequency support of the Wigner distribution into horizontal and vertical shears, and scalings and rotations.

Decomposition into hyperdifferential operators will be added to this list later in this chapter. Further interpretations will be possible in an optical context: decomposition into lenses, sections of free space, imaging systems, and Fourier transforming systems. Any of these interpretations can be used to verify a particular decomposition, but matrix mul-

tiplication is often the simplest. Of course, we remember that these correspondences are valid within a factor of $\pm 1$.

Any linear canonical transform can be obtained from another by properly scaling it and appending chirp multiplications to both ends:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ (DB - D'B')/B^2 & 1 \end{bmatrix} \begin{bmatrix} B/B' & 0 \\ 0 & B'/B \end{bmatrix} \times \begin{bmatrix} A' & B' \\ C' & D' \end{bmatrix} \begin{bmatrix} 1 & 0 \\ (AB' - A'B)/BB' & 1 \end{bmatrix} \tag{2.138}$$

This result can be specialized by replacing either or both linear canonical transforms by special transforms such as Fourier, scaling, fractional Fourier, chirp multiplication, chirp convolution, and so forth. Thus, for example, a Fourier transform can be expressed as a chirp multiplication followed by a scaling operation followed by another chirp multiplication. A dual of this result based on appending chirp convolutions rather than chirp multiplications to both ends also exists. Further variations may be obtained by changing the position of the scaling matrix in the above equation.

It is worth explicitly writing some of the special cases of the above equation for future reference:

$$\mathcal{C}_{\mathbf{M}} = \mathcal{Q}_{q_1} \mathcal{M}_{M_1} \mathcal{F}^a \mathcal{Q}_{q_2}, \tag{2.139}$$

$$\mathcal{C}_{\mathbf{M}} = \mathcal{Q}_{q_3} \mathcal{M}_{M_2} \mathcal{F} \mathcal{Q}_{q_4}, \tag{2.140}$$

which hold for suitable choices of the parameters.

We will most commonly use the canonical decompositions given by equations 2.122 and 2.123. However, if both $B$ and $C$ are zero, it may be more convenient to employ the alternative decomposition (Papoulis 1977):

$$\begin{bmatrix} M & 0 \\ 0 & 1/M \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ (M^{-1} - M^{-2})/X & 1 \end{bmatrix} \begin{bmatrix} 1 & -XM \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ (M^{-1} - 1)/X & 1 \end{bmatrix} \begin{bmatrix} 1 & X \\ 0 & 1 \end{bmatrix}, \tag{2.141}$$

where $X$ is a parameter of our choosing.

We finally note that the operators $\mathcal{Q}_q$ and $\mathcal{R}_r$ can be referred to as being *duals* of each other when $q = r$, since they correspond to each other under Fourier transformation (chirp multiplication corresponds to chirp convolution in the Fourier domain and vice versa):

$$\mathcal{Q}_q = \mathcal{F}^{-1} \mathcal{R}_r \mathcal{F}, \tag{2.142}$$

$$\mathcal{R}_r = \mathcal{F} \mathcal{Q}_q \mathcal{F}^{-1}. \tag{2.143}$$

These equations are nothing but the convolution and product theorems for the Fourier transform.

### 2.4.5 Transformation of moments

The matrix $\mathbf{M}$ also allows one to write simple expressions for the transformation of the first- and second-order moments of a signal, relating the moments of the linear canonical transform to the moments of the original signal. Let $\overline{u}$ and $\overline{\mu}$ denote the first-order moments of the Wigner distribution of $f(u)$:

$$\overline{u} = \frac{\iint u W_f(u,\mu)\,du\,d\mu}{\iint W_f(u,\mu)\,du\,d\mu}, \tag{2.144}$$

$$\overline{\mu} = \frac{\iint \mu W_f(u,\mu)\,du\,d\mu}{\iint W_f(u,\mu)\,du\,d\mu}, \tag{2.145}$$

and define the first-order time-frequency moment vector $\overline{\mathbf{u}} \equiv [\overline{u}\ \overline{\mu}]^{\mathrm{T}}$. Then, it is possible to show that this vector transforms as (Bastiaans 1989, 1991b)

$$\overline{\mathbf{u}}_{f_{\mathbf{M}}} = \mathbf{M}\overline{\mathbf{u}}_f. \tag{2.146}$$

That is, the first-order moment vector of the transformed signal is obtained simply by multiplying the moment vector of the original signal by $\mathbf{M}$. Now, let us define the nonnegative-definite centralized second-order moment matrix (moment of inertia tensor)

$$\mathbf{P} \equiv \begin{bmatrix} \overline{(u-\overline{u})^2} & \overline{(u-\overline{u})(\mu-\overline{\mu})} \\ \overline{(u-\overline{u})(\mu-\overline{\mu})} & \overline{(\mu-\overline{\mu})^2} \end{bmatrix}, \tag{2.147}$$

where the various second-order moments are defined in a manner similar to the first-order moments. It is possible to show that these are transformed as (Bastiaans 1989, 1991b)

$$\mathbf{P}_{f_{\mathbf{M}}} = \mathbf{M}\mathbf{P}_f\mathbf{M}^{\mathrm{T}}, \tag{2.148}$$

a result which preserves the nonnegative-definiteness of the moment matrix.

Since Gaussian signals are completely characterized by their first- and second-order moments, these results are particularly useful when dealing with such signals. We note that the determinant of $\mathbf{P}_{f_{\mathbf{M}}}$ is the same as that of $\mathbf{P}_f$. On the other hand, the trace of $\mathbf{P}_{f_{\mathbf{M}}}$ will be the same as that of $\mathbf{P}_f$ if and only if $\mathbf{M}$ is a rotation matrix. (The square root of the determinant is a measure of the support area whereas the trace corresponds to the moment of inertia of the Wigner distribution.) Further details and discussion with applications in optics may be found in Bastiaans 1989, 1991b.

### 2.4.6 Linear fractional transformations

Let us consider the linear canonical transform $f_{\mathbf{M}}(u)$ of a unit-energy *complex Gaussian function* $f(u)$ with *complex radius* $r_c$:

$$f(u) = (2\Im[1/r_c])^{1/4}\ e^{i\pi u^2/r_c}, \qquad \Im[1/r_c] > 0, \tag{2.149}$$

$$W_f(u,\mu) = 2\exp\left\{-2\pi\left[\Im[1/r_c]u^2 + \frac{(\mu - \Re[1/r_c]u)^2}{\Im[1/r_c]}\right]\right\}, \tag{2.150}$$

$$f_{\mathbf{M}}(u) = (2\Im[1/r_c'])^{1/4}\ e^{i\pi u^2/r_c'}, \tag{2.151}$$

where

$$r_c' = \left[\alpha - \frac{\beta^2 r_c}{\gamma r_c + 1}\right]^{-1}, \tag{2.152}$$

is the complex radius of $f_{\mathbf{M}}(u)$ and $\Re[1/r_c]$ and $\Im[1/r_c]$ are the real and imaginary parts of $1/r_c$. (The term complex radius comes from optics, where $r_c$ is interpreted as the radius of a wavefront.) We see that the linear canonical transform of a complex Gaussian function with given complex radius $r_c$ is always another complex Gaussian function with complex radius $r_c'$. The relationship between the complex radius of the transformed function and that of the original function can be cast in a simpler form if we use the $A, B, C, D$ parameters instead of $\alpha, \beta, \gamma$. Using equation 2.76 we obtain (Bastiaans 1989, 1991b)

$$r_c' = \frac{Ar_c + B}{Cr_c + D}. \tag{2.153}$$

Such functional relationships are referred to as linear fractional transformations. (When $\Im[1/r_c] = 0$, we expect to recover the simple chirp function $\propto \exp(i\pi u^2/r)$ with $\Re[1/r_c] = 1/r_c = 1/r$. However, since the chirp function does not have unit energy, this does not follow as a special case of equation 2.149 by letting $\Im[1/r_c] \to 0$. Nevertheless, the result $r' = (Ar + B)/(Cr + D)$ still holds for such chirp functions.)

Linear fractional transformations constitute an alternative to the matrix formulation of linear canonical transforms. Equation 2.153 can be generalized to arbitrary finite-energy functions (Bastiaans 1989, 1991b). Let us define the complex quantities $Z$ and $Y = 1/Z$ as

$$Z = \frac{\overline{u\mu}}{\overline{\mu^2}} + i\,\frac{-\sqrt{\overline{u^2}\,\overline{\mu^2} - \overline{u\mu}^2}}{\overline{\mu^2}}, \tag{2.154}$$

$$Y = \frac{\overline{u\mu}}{\overline{u^2}} + i\,\frac{\sqrt{\overline{u^2}\,\overline{\mu^2} - \overline{u\mu}^2}}{\overline{u^2}}, \tag{2.155}$$

where the moments are those of the Wigner distribution of the arbitrary function (see page 104). (If desired, $Z$ and $Y$ can also be expressed in terms of moments defined in the time and frequency domains.) Now, let $Z$ denote the complex parameter associated with a function $f(u)$. The complex parameter $Z'$ associated with the linear canonical transform of $f(u)$ is given by

$$Z' = \frac{AZ + B}{CZ + D}. \tag{2.156}$$

Likewise, the complex parameters $Y$ and $Y'$ are related through

$$Y' = \frac{DY + C}{BY + A}. \tag{2.157}$$

We see that the $Z$ parameter transforms in the same way as the complex radius and is thus a generalization of this concept for arbitrary signals. Let us consider some special

cases. First, consider chirp convolution. In this case $A = D = 1$, $B = r$, and $C = 0$. Then

$$Z' = Z + r, \tag{2.158}$$

$$\frac{1}{Y'} = \frac{1}{Y} + r. \tag{2.159}$$

Second, consider chirp multiplication for which $A = D = 1$, $C = -q$, and $B = 0$. Then

$$Y' = Y - q, \tag{2.160}$$

$$\frac{1}{Z'} = \frac{1}{Z} - q. \tag{2.161}$$

We will see in chapter 3 that chirp convolution corresponds to free-space propagation and that $r$ represents the distance of propagation. Thus, the effect of propagation on the $Z$ parameter is simply to increase it by the distance of propagation. We will also see that chirp multiplication corresponds to passage through a thin lens and that $1/q$ corresponds to the focal length of the lens. This leads to the well-recognized lens formula $1/Z' = 1/Z - 1/(\text{focal length})$. If we make an analogy with electrical circuits, the $Z$ parameter is like an impedance and propagation through free space is like the addition of a series element; the $Y$ parameter is like an admittance and passage through a lens is like the addition of a parallel element.

We finally add that it is also possible to generalize these results to non-centered Gaussian or quadratic-phase signals passing through systems which distort the Wigner distribution in an affine manner (that is, including time/space and frequency shifts).

### 2.4.7 Coordinate multiplication and differentiation operators

The coordinate multiplication operator $\mathcal{U}$ and the differentiation operator $\mathcal{D}$ were seen to switch roles when we transform to the Fourier domain. Functions related to each other through linear canonical transforms may be considered representations of the same abstract signal in different linear canonical transform domains. We now seek the relationship between the coordinate multiplication and differentiation operators associated with these different domains. $\mathcal{U}_{\mathbf{M}}$ will denote the operator which multiplies by the coordinate variable and $\mathcal{D}_{\mathbf{M}}$ will denote the operator which differentiates with respect to the coordinate variable, both in the domain represented by the matrix $\mathbf{M}$. Since the effect of $\mathcal{U}_{\mathbf{M}}$ (or $\mathcal{D}_{\mathbf{M}}$) amounts to first taking the linear canonical transform, then coordinate multiplying (or differentiating), and then going back to the original domain, we can write the relationship between these operators as

$$\begin{aligned} \mathcal{U}_{\mathbf{M}} &= \mathcal{C}_{\mathbf{M}}^{-1} \mathcal{U} \mathcal{C}_{\mathbf{M}}, \\ \mathcal{D}_{\mathbf{M}} &= \mathcal{C}_{\mathbf{M}}^{-1} \mathcal{D} \mathcal{C}_{\mathbf{M}}. \end{aligned} \tag{2.162}$$

By using properties 7 and 8 in table 2.8, it is possible to show that

$$\begin{bmatrix} \mathcal{U}_{\mathbf{M}} \\ \mathcal{D}_{\mathbf{M}} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \mathcal{U} \\ \mathcal{D} \end{bmatrix}. \tag{2.163}$$

An alternative development of linear canonical transforms may be found in Wolf 1979, where these equations are taken as the defining characteristics of linear canonical transforms and the integral form is subsequently derived. (It should be noted that Wolf poses the same idea in a different form by concentrating on the transformation between the different representations of the same operator $\mathcal{U}$, rather than the relationship between the different operators $\mathcal{U}_{\mathbf{M}}$ and $\mathcal{U}$ which both coordinate multiply in their respective domains.)

### 2.4.8 Uncertainty relation

The standard deviations of $f(u)$ and $f_{\mathbf{M}}(u)$, denoted by $\Delta u$ and $\Delta u_{\mathbf{M}}$ respectively, satisfy (Wolf 1979)

$$\Delta u \Delta u_{\mathbf{M}} \geq |B|/4\pi, \tag{2.164}$$

which reduces to the ordinary Fourier uncertainty relation when the linear canonical transform in question is the ordinary Fourier transform.

### 2.4.9 Invariants and hyperdifferential forms

A number of Hermitian operators are invariant under some of the special canonical transforms we have discussed. An operator $\mathcal{H}$ is invariant under $\mathcal{C}_{\mathbf{M}}$ if and only if $\mathcal{C}_{\mathbf{M}}^{-1}\mathcal{H}\mathcal{C}_{\mathbf{M}} = \mathcal{H}$, so that $g = \mathcal{H}f$ implies $(\mathcal{C}_{\mathbf{M}}g) = \mathcal{H}(\mathcal{C}_{\mathbf{M}}f)$. These invariant operators are (Wolf 1979, page 391):

$$\begin{array}{lcllll}
\mathcal{H}_M & \equiv & 2\pi\frac{1}{2}(\mathcal{U}\mathcal{D}+\mathcal{D}\mathcal{U}) & \equiv & 2\mathcal{J}_2, & \text{under scaling or parabolic,} \\
\mathcal{H}_R & \equiv & 2\pi\frac{1}{2}\mathcal{D}^2, & & & \text{under chirp convolution,} \\
\mathcal{H}_Q & \equiv & 2\pi\frac{1}{2}\mathcal{U}^2, & & & \text{under chirp multiplication,} \\
\mathcal{H}_{HYP} & \equiv & 2\pi\frac{1}{2}(\mathcal{D}^2-\mathcal{U}^2) & \equiv & 2\mathcal{J}_1, & \text{under hyperbolic,} \\
\mathcal{H}_{fF} & \equiv & 2\pi\frac{1}{2}(\mathcal{D}^2+\mathcal{U}^2) & \equiv & 2\mathcal{J}_0, & \text{under fractional Fourier transform.}
\end{array}$$

The invariance of these operators can be demonstrated directly by using equations 2.162 and equation 2.163. Invariance can equivalently be stated as a commutation relation since $\mathcal{C}_{\mathbf{M}}^{-1}\mathcal{H}\mathcal{C}_{\mathbf{M}} = \mathcal{H}$ implies $\mathcal{H}\mathcal{C}_{\mathbf{M}} = \mathcal{C}_{\mathbf{M}}\mathcal{H}$. For instance, we can write $\mathcal{Q}_q^{-1}\mathcal{U}^2\mathcal{Q}_q = \mathcal{U}^2$ which means that $\mathcal{U}^2$ commutes with $\mathcal{Q}_q$. Thus, eigenfunctions of these operators are also eigenfunctions of the indicated special linear canonical transforms. We also note that $\mathcal{H}_{fF} = \mathcal{A}^{\mathrm{H}}\mathcal{A} + 1/2 = \mathcal{A}\mathcal{A}^{\mathrm{H}} - 1/2$, where $\mathcal{A}$ was defined in equation 1.249.

We have already discussed several one-parameter subgroups of the group of linear canonical transforms. Noting that most of these have the property that when their parameter is zero, they reduce to the identity transform, we will now seek hyperdifferential forms for these operators of the form $\exp(-ip\mathcal{H})$, where $p$ is the relevant parameter. We know that $\exp(-ip\mathcal{H})$ is unitary if and only if $\mathcal{H}$ is Hermitian. The hyperdifferential forms of the various special linear canonical transform operators we have discussed can be given in terms of the above defined Hermitian operators as follows (Wolf 1979, page 408):

$$\mathcal{M}_M = \exp(-i\,\ln M\,\mathcal{H}_M) = \exp[-i2\pi\ln M(\mathcal{U}\mathcal{D}+\mathcal{D}\mathcal{U})/2] = M^{[-i\pi(\mathcal{U}\mathcal{D}+\mathcal{D}\mathcal{U})]}, \tag{2.165}$$

$$\mathcal{R}_r = \exp(-ir\mathcal{H}_R) = \exp(-i\pi r\mathcal{D}^2), \quad (2.166)$$

$$\mathcal{Q}_q = \exp(-iq\mathcal{H}_Q) = \exp(-i\pi q\mathcal{U}^2), \quad (2.167)$$

$$\mathcal{HYP}_a = \exp\left[-i(a\pi/2)\mathcal{H}_{HYP}\right], \quad (2.168)$$

$$\mathcal{F}_{\mathrm{lc}}^a = \exp\left[-i(a\pi/2)\mathcal{H}_{fF}\right] = \exp[-i(a\pi^2/2)(\mathcal{U}^2 + \mathcal{D}^2)]. \quad (2.169)$$

The unitary operators given above are sufficient to construct any linear canonical transform operator by concatenation. This follows from the matrix decompositions given in section 2.4.4. By using the hyperdifferential forms given above, we can also obtain a Baker-Campbell-Hausdorff formula corresponding to each matrix decomposition appearing in section 2.4.4. These abstract operator formulas, can then be specialized to any domain or representation. For example, using equation 2.124 for the matrix given in equation 2.113, we obtain

$$\begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\tan\alpha & 1 \end{bmatrix} \begin{bmatrix} \cos\alpha & 0 \\ 0 & 1/\cos\alpha \end{bmatrix} \begin{bmatrix} 1 & \tan\alpha \\ 0 & 1 \end{bmatrix}. \quad (2.170)$$

Now, we recognize this decomposition as a chirp convolution followed by scaling followed by chirp multiplication (see page 95 onwards). Identifying $r = q = \tan\alpha$ and $M = \cos\alpha$ and using the hyperdifferential forms given above, one immediately obtains equation 1.245.

We should perhaps also note that if $\mathcal{C}_{\mathbf{M}} = \exp(-ip\mathcal{H})$ is the hyperdifferential form of a particular linear canonical transform $\mathcal{C}_{\mathbf{M}}$, it is *necessarily* the case that $\mathcal{H}$ is invariant under that transform. That is,

$$e^{ip\mathcal{H}}\mathcal{H}e^{-ip\mathcal{H}} = \mathcal{H}, \quad (2.171)$$

which is true since $e^{ip\mathcal{H}}$ and $\mathcal{H}$ commute.

Thus, to the list of alternative interpretations of decompositions given on page 101, we may add decomposition into hyperdifferential operators. Each interpretation is associated with a different set of objects. These sets of objects are: matrices; integral transforms; abstract operators; chirp multiplication, convolution and other operations; geometric distortion operations; and hyperdifferential operators. In later chapters, we will add optical components to this list.

### 2.4.10 Differential equations

We now consider differential equations of the form

$$\mathcal{H}f_p(u) = i\,\frac{\partial f_p(u)}{\partial p}, \quad (2.172)$$

where $\mathcal{H}$ is a quadratic Hermitian operator in $\mathcal{U}$ and $\mathcal{D}$. Since the Hermitian conjugate of $\mathcal{DU}$ is $\mathcal{UD}$ and vice versa, such an operator may only contain terms proportional to $\mathcal{U}^2$, $\mathcal{D}^2$, and $\mathcal{UD} + \mathcal{DU}$. The solution of this equation is

$$f_p(u) = e^{-ip\mathcal{H}}f_0(u), \quad (2.173)$$

where $f_0(u)$ serves as the initial or boundary condition. (To prove that this satisfies equation 2.172, it is sufficient to expand the exponential into a series.) Alternatively, the solution can be expressed as a one-parameter canonical transform (Wolf 1979, page 410):

$$\mathcal{C}_p = e^{-ip\mathcal{H}}, \tag{2.174}$$

$$f_p(u) = (\mathcal{C}_p[f_0(u)])(u) = \int C_p(u,u')f_0(u')\,du'. \tag{2.175}$$

The operator $\mathcal{C}_p$ may be referred to as the time-evolution or Green's operator for the system governed by the differential equation in question. The associated kernel is known as the Green's function $C_p(u,u')$, which is simply the response of the system to $f_0(u) = \delta(u-u')$. (Wolf 1979)

Certain functions will preserve their forms under the one-parameter linear canonical transform described above. Let $\psi_\lambda(u)$ be an eigenfunction of $\mathcal{H}$ with eigenvalue $\lambda$. Then

$$\mathcal{C}_p[\psi_\lambda(u)](u) = \int C_p(u,u')\psi_\lambda(u')\,du' = e^{-ip\mathcal{H}}\psi_\lambda(u) = e^{-ip\lambda}\psi_\lambda(u). \tag{2.176}$$

We see that the dependence of $\mathcal{C}_p[\psi_\lambda(u)](u)$ on $u$ and $p$ is separable. The eigenfunctions corresponding to some of the operators we have discussed are

$$\begin{array}{rcl} \psi_{R\lambda}(u) = e^{\pm i2\pi\sqrt{\lambda/\pi}\,u} & \lambda \geq 0 & \text{chirp conv.,} \\ \psi_{Q\lambda}(u) = \delta(u \pm \sqrt{\lambda/\pi}\,) & \lambda \geq 0 & \text{chirp mult.,} \\ \psi_{fF\lambda}(u) = \psi_{\lambda-1/2}(u) & \lambda = n+1/2, \quad n = 0,1,\ldots & \text{frac. Four. trans.,} \end{array}$$

where $\psi_{\lambda-1/2}(u)$ is the $(\lambda-1/2)$th order Hermite-Gaussian function (also see page 44). Further discussion of the eigenfunctions of linear canonical transforms may be found in Wolf 1977.

### 2.4.11 Symplectic systems

We will begin by mentioning the concept of a *form*. The most common example of a form is the scalar inner product of two signals or vectors. It is often of interest to inquire into the nature of the set of systems under which such a form is preserved (is invariant). We know that the inner product of two signals remains the same under passage through a unitary system. We also know that physically this corresponds to a system conserving power or energy (since the norm is a special case of the inner product and signal energy is the square of the norm, and since signal energy usually corresponds to physical energy or power). The inner product is an example of a *symmetric form*.

An example of an *antisymmetric form* is the *symplectic form*. Here we will consider the symplectic form $\prec[u_1,\mu_1]^{\mathrm{T}},[u_2,\mu_2]^{\mathrm{T}}\succ$ of two space-frequency (or time-frequency) vectors $[u_1,\mu_1]^{\mathrm{T}}$ and $[u_2,\mu_2]^{\mathrm{T}}$, which we define as (Folland 1989)

$$\prec[u_1,\mu_1]^{\mathrm{T}},[u_2,\mu_2]^{\mathrm{T}}\succ \equiv [u_1,\mu_1]\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} u_2 \\ \mu_2 \end{bmatrix} = u_1\mu_2 - u_2\mu_1. \tag{2.177}$$

The antisymmetric $2 \times 2$ matrix appearing in the definition will be denoted by the symbol $\mathbf{J}$ and satisfies $\mathbf{J}^{\mathrm{T}} = -\mathbf{J} = \mathbf{J}^{-1}$. We see that the symplectic form of two space-frequency vectors gives the (signed) area of the parallelogram defined by those two vectors. (To see this, recall that the signed area of a parallelogram is given by the vector cross product of the vectors defining it, and that $u_1\mu_2 - u_2\mu_1$ is nothing but the value of this cross product. The "sign" here refers to the direction of the resultant cross product vector.) We also note that the symplectic form can be interpreted as the space-frequency hypervolume for higher-dimensional spaces as well. Thus physically, preservation of the symplectic form corresponds to invariance of space-frequency area or volume, a fact that will become important in chapter 4.

Now, let us consider a matrix $\mathbf{M} = [A\ B;\ C\ D]$ mapping the space-frequency vectors $[u_1, \mu_1]^{\mathrm{T}}$ and $[u_2, \mu_2]^{\mathrm{T}}$ into $\mathbf{M}\,[u_1, \mu_1]^{\mathrm{T}}$ and $\mathbf{M}\,[u_2, \mu_2]^{\mathrm{T}}$. Equating the symplectic form of these new vectors to the original symplectic form as

$$\prec \mathbf{M}[u_1, \mu_1]^{\mathrm{T}}, \mathbf{M}[u_2, \mu_2]^{\mathrm{T}} \succ = \prec [u_1, \mu_1]^{\mathrm{T}}, [u_2, \mu_2]^{\mathrm{T}} \succ, \tag{2.178}$$

$$\left( \mathbf{M} \begin{bmatrix} u_1 \\ \mu_1 \end{bmatrix} \right)^{\mathrm{T}} \mathbf{J}\, \mathbf{M} \begin{bmatrix} u_2 \\ \mu_2 \end{bmatrix} = [u_1, \mu_1]\, \mathbf{J} \begin{bmatrix} u_2 \\ \mu_2 \end{bmatrix}, \tag{2.179}$$

leads us to the condition

$$\mathbf{M}^{\mathrm{T}} \mathbf{J} \mathbf{M} = \mathbf{J}, \tag{2.180}$$

for the preservation of the symplectic form under the mapping represented by $\mathbf{M}$. Matrices $\mathbf{M}$ satisfying this relation are called symplectic matrices; they preserve the symplectic form. In the one-dimensional case we are considering, it is easy to show that this condition is fully equivalent to the unit-determinant condition $AD - BC = 1$. Thus, the matrices characterizing linear canonical transforms introduced in equation 2.75 are symplectic. Although trivial in the one-dimensional case, similar results hold for higher dimensions as well (Folland 1989). In general, matrices are called symplectic if they preserve the symplectic form (which corresponds to preservation of space-frequency area or volume), and the condition for this turns out to be of the same form as equation 2.180, which is not simply equivalent to a unit-determinant condition in higher dimensions. The matrices characterizing multi-dimensional linear canonical transforms are always symplectic.

Moving further in this direction is beyond what we can achieve in this book. The interested reader is referred to Folland 1989 (particularly chapter 1: The Heisenberg group and its representations, and chapter 4: The metaplectic representation) and Guillemin and Sternberg 1984 (chapter 1: Introduction). Another interesting work is Turski 1998. We will revisit these issues in an optical context in section 4.8.

We finally note that the symplecticity of matrices characterizing linear canonical transforms can be alternatively shown as follows: First, we can easily show that the matrices for chirp multiplication and chirp convolution satisfy the symplecticity definition given above. Then, we can show that if two matrices satisfy this definition, all their products and inverses also do. Since any linear canonical transform can be expressed as the prod-

uct of chirp multiplications and chirp convolutions, we can then conclude that the matrix associated with any linear canonical transform is symplectic.

### 2.4.12 Connections to group theory

By now we have talked about quite a number of different sets of objects in association with linear canonical transforms. These sets of objects are: matrices; integral transforms; abstract operators; chirp multiplication, convolution and other operations; geometric distortion operations; hyperdifferential operators; and optical operators (to be introduced later).

Each of these sets of objects together with a rule of composition satisfy the axioms of a group: (i) The composition of any two elements of the set is also an element of the set (closure). (ii) There exists an identity element such that the composition of any element with that identity is again the same element. (iii) Each element has an inverse such that the composition of any element with its inverse is the identity element. (iv) Composition is associative.

It is not difficult to show that any of the sets of objects we considered and their composition rules constitute a group. Furthermore, for every element in one of these groups, there are corresponding elements in each of the other groups, and these correspondences are preserved under composition (save for a possible sign) (Wolf 1979). If we ignore differences caused by $\pm$ signs, these correspondences become one-to-one. Such groups are called isomorphic to each other.

Some groups are given special names. In particular, the group of real $2 \times 2$ matrices of determinant $+1$ is referred to as the $SL(2, \mathbf{R})$ group (Dym and McKean 1972, page 273). All of our groups are isomorphic to this group. The notation and definition offered in Folland 1989, page 171 is slightly different but equivalent: The symplectic group $Sp(2, \mathbf{R})$ is the group of $2 \times 2$ real matrices which preserve the symplectic form. (A matrix "preserves the symplectic form" if and only if it is symplectic.) Dym and McKean (1972, page 275) also define the one parameter subgroup $SO(2)$ of matrices of the form given by equation 2.113. Thus fractional Fourier transforms as a group are isomorphic to the group denoted as $SO(2)$.

As we have noted on several occasions, the above correspondences are true only to the extent that we are willing to be flexible with $\pm$ signs. While it is beyond the scope of this book to present a more precise formulation (see Guillemin and Sternberg 1984, Folland 1989), we mention that linear canonical transforms more precisely constitute a metaplectic group $Mp(2, \mathbf{R})$, rather than a symplectic group. The group $Mp(2, \mathbf{R})$ is what is known as a *double cover* (or *twofold cover*) of the group $Sp(2, \mathbf{R})$. Linear canonical transforms are not strictly isomorphic to the symplectic group, but are rather *locally isomorphic.*

An example of the application of linear canonical transforms and group theoretical methods to optical imaging and image processing may be found in Seger and Lenz 1992 and Seger 1993.

## 2.5 Generalization to two and higher dimensions

Most of the concepts presented in this chapter can be generalized to two and higher dimensions in a straightforward manner. Here we will present generalizations of only some of the more important concepts. The Wigner distribution of $f(u,v)$ is defined as

$$W_f(u,v;\mu,\nu) = \iint f(u+u'/2, v+v'/2) f^*(u-u'/2, v-v'/2) e^{-i2\pi(\mu u' + \nu v')}\, du'\, dv', \quad (2.181)$$

in two dimensions and similarly in higher dimensions. The separable two-dimensional linear canonical transform of $f(u,v)$ is defined as

$$f_{\mathbf{M}}(u,v) = \iint C_{\mathbf{M}}(u,v;u',v') f(u',v')\, du'\, dv', \quad (2.182)$$
$$C_{\mathbf{M}}(u,v;u',v') = C_{\mathbf{M}_u}(u,u') C_{\mathbf{M}_v}(v,v').$$

Here $\mathbf{M}$ is a four-dimensional parameter matrix given by

$$\mathbf{M} = \begin{bmatrix} A_u & 0 & B_u & 0 \\ 0 & A_v & 0 & B_v \\ C_u & 0 & D_u & 0 \\ 0 & C_v & 0 & D_v \end{bmatrix}. \quad (2.183)$$

Thus defined, we may multiply matrices associated with two transforms and find the matrix associated with their concatenation. The Wigner distribution of $f_{\mathbf{M}}(u,v)$ is related to that of $f(u,v)$ according to

$$W_{f_{\mathbf{M}}}(u,v;\mu,\nu) = W_f(D_u u - B_u \mu, D_v v - B_v \nu; -C_u u + A_u \mu, -C_v v + A_v \nu). \quad (2.184)$$

It is also possible to define non-separable linear canonical transforms in which the cross terms are not zero (Folland 1989).

Just as Hankel transforms correspond to two-dimensional Fourier transforms under circular symmetry, it is possible to derive transforms which corresponds to two-dimensional linear canonical transforms under circular symmetry. Such transforms are discussed in Moshinsky, Seligman, and Wolf 1972 and Zalevsky, Mendlovic, and Lohmann 1998.

## 2.6 Further reading

General tutorial references on time-frequency representations with a large number of additional references include Hlawatsch and Boudreaux-Bartels 1992, Cohen 1989, 1995, Flandrin 1993, and Qian and Chen 1996. A classic exposition of the Wigner distribution is Claasen and Mecklenbräuker 1980a, b, c. A general reference which also includes time-frequency transformations is Transforms and Applications Handbook 2000. Time- and space-frequency representations are widely used in physics, and in fact have their origin in physics. For such a perspective, see Hillery and others 1984.

Chapter 9 of Wolf 1979 is an excellent discussion of linear canonical transforms.

# Chapter 3

# Optical Signals and Systems

## 3.1 Introduction

In this book we will restrict our attention to the familiar class of centered optical systems. A typical example of such a system is shown in figure 3.1. It is composed of a number of spherical lenses and spatial filters, separated by sections of free space and centered about the optical axis (customarily chosen as the $z$ axis). Other optical components and features that may appear in such systems include cylindrical and anamorphic lenses, mirrors, prisms, gratings, diffractive optical elements, sections of graded-index media, sections of homogeneous media with arbitrary refractive indices, and planar or spherical interfaces between such media. Axially or rotationally symmetric systems are those whose axially

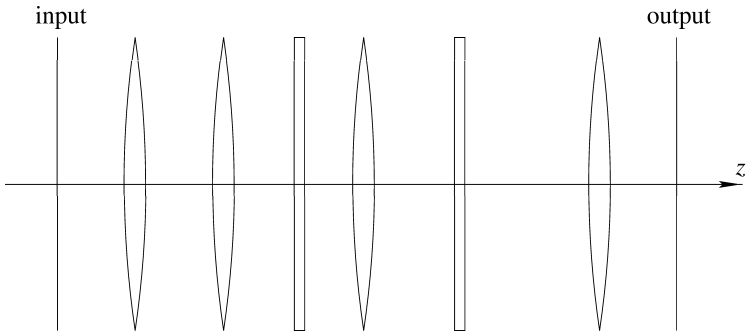


Figure 3.1: Centered optical system consisting of four convex lenses, two spatial filters (shown as thin slabs), and seven sections of free space.

rotated versions are indistinguishable from themselves. A spherical lens is a rotationally symmetric component, whereas a cylindrical lens or prism is not. Centered systems consisting only of spherical lenses and sections of free space are axially symmetric.

The optical components of which optical systems are composed, can be viewed as elementary optical systems themselves (figure 3.2). Each optical component has its own

input and output planes, and alters the distribution of light incident on its input plane in a certain way to produce the distribution of light at its output plane. If the effect of each optical component is known, the overall effect of the optical system can be found.

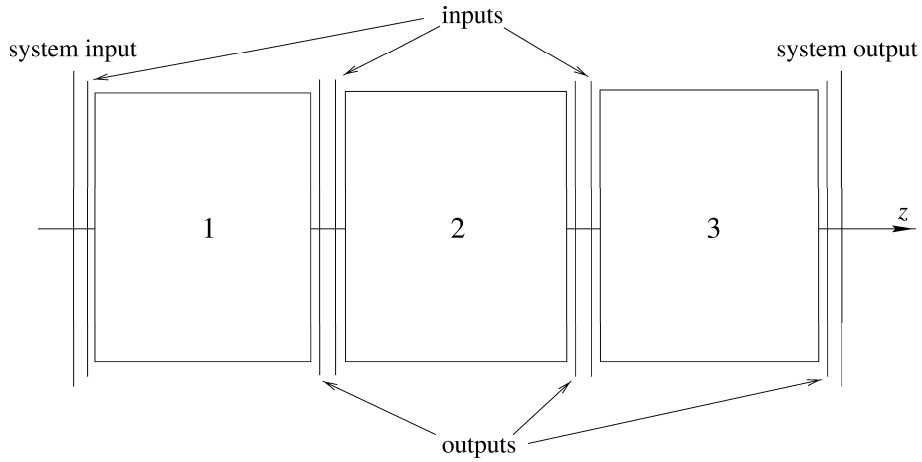


Figure 3.2: An optical system as a sequence of tandem optical components or subsystems. The output of each subsystem is the input of the next subsystem. The input (output) of the first (last) subsystem is the input (output) of the overall system.

## 3.2 Notation and conventions

Dimensionless variables and parameters were employed in the previous chapters for simplicity and purity (see subsection 1.1.2). In this and later chapters dealing with optical signals and systems, we will employ variables with real physical dimensions. We will exercise great care to ensure that the correspondence between the equations, results, and properties presented in both sets of chapters is self-evident, and that translating dimensionless and dimensional equations into each other is straightforward. This will allow the results of dimensionless chapters to be easily employed in dimensional chapters.

The space- and frequency-domain representations of a signal $f$ have their dimensional and dimensionless forms related as follows:

$$\check{f}(x) \equiv \frac{1}{\sqrt{s}} f(x/s) \equiv \frac{1}{\sqrt{s}} f(u), \tag{3.1}$$

$$\check{F}(\sigma_x) \equiv \sqrt{s}\, F(s\sigma_x) \equiv \sqrt{s}\, F(\mu), \tag{3.2}$$

where $u \equiv x/s$, $\mu \equiv s\sigma_x$. The scale parameter $s > 0$ has the same dimension as $x$ and $1/\sigma_x$, usually meters. The circumflex $\check{}$ designates functions taking dimensional space or frequency arguments such as $x, y, z$, $\sigma_x, \sigma_y, \sigma_z$, regardless of domain or representation. As usual, lower case denotes the space-domain representation, and upper case denotes the

frequency-domain representation of a signal. Although $f(u) = f(x/s)$ has a dimensionless argument, nothing prevents it from having a dimension itself (such as V, V/m, or V/m$\Omega^{1/2}$, the latter whose square is W/m$^2$).

$F(\mu) = \int f(u) \exp(-i2\pi\mu u)\, du$ is the Fourier transform of $f(u)$ as defined in chapter 1. With the above conventions, the reader may verify that $\check{F}(\sigma_x)$ is the Fourier transform of $\check{f}(x)$ defined as

$$\check{F}(\sigma_x) = \int \check{f}(x) e^{-i2\pi\sigma_x x}\, dx, \tag{3.3}$$

$$\check{f}(x) = \int \check{F}(\sigma_x) e^{i2\pi\sigma_x x}\, d\sigma_x, \tag{3.4}$$

and that

$$\int |\check{f}(x)|^2\, dx = \int |f(u)|^2\, du = \int |F(\mu)|^2\, d\mu = \int |\check{F}(\sigma_x)|^2\, d\sigma_x, \tag{3.5}$$

so that the signal energy is the same regardless of whether it is evaluated in dimensional or dimensionless coordinates. Since the defining relations for the Fourier transform look the same in dimensional and dimensionless coordinates, all properties of the Fourier transform will also look the same. We also note that the dimension of $\check{F}(\sigma_x)$ is the dimension of $\check{f}(x)$ multiplied by the dimension of $x$.

Now, let us consider the dimensionless and dimensional forms of linear system or transformation integrals where the input and output variables $x'$ and $x$ are of the same dimension:

$$g(u) = \int h(u, u') f(u')\, du', \tag{3.6}$$

$$\check{g}(x) = \int \check{h}(x, x') \check{f}(x')\, dx', \tag{3.7}$$

where $\check{h}(x, x') \equiv s^{-1} h(x/s, x'/s) \equiv s^{-1} h(u, u')$. Our conventions ensure that these equations are consistent and that if $h(u, u')$ is unitary, $\check{h}(x, x')$ will also be unitary. (To see this, the reader may show that the orthonormality relation between the rows/columns of $h(u, u')$, given by $\int h(u, u'') h^*(u'', u')\, du'' = \delta(u - u')$, implies $\int \check{h}(x, x'') \check{h}^*(x'', x')\, dx'' = \delta(x - x')$.) Thus we can translate a dimensionless linear relation into a dimensional one simply by replacing both the input and output functions and the kernel with their dimensional counterparts. $\check{f}(x)$ and $\check{g}(x)$ usually have the same dimension so that $\check{h}(x, x')$ has the inverse of the dimension of $x$.

The dimensionless and dimensional forms of the Wigner distribution can likewise be consistently written in a similar form:

$$W_f(u, \mu) = \int f(u + u'/2) f^*(u - u'/2) e^{-i2\pi\mu u'}\, du', \tag{3.8}$$

$$\check{W}_{\check{f}}(x, \sigma_x) = \int \check{f}(x + x'/2) \check{f}^*(x - x'/2) e^{-i2\pi\sigma_x x'}\, dx'. \tag{3.9}$$

These have the dimension of $|\check{f}(x)|^2$ times the dimension of $x$, which is the same as the dimension of signal energy.

Delta functions require slight care. Remember that $\delta(x-x') = \delta(su-su') = s^{-1}\delta(u-u')$. Thus the sifting property also looks the same whether dimensionless or dimensional:

$$f(u) = \int \delta(u-u')f(u')\,du', \tag{3.10}$$

$$\check{f}(x) = \int \delta(x-x')\check{f}(x')\,dx'. \tag{3.11}$$

Now, let us consider a discrete basis expansion:

$$f(u) = \sum_l C_l\psi_l(u), \qquad C_l = \frac{\int \psi_l^*(u)f(u)\,du}{\int |\psi_l(u)|^2\,du}, \tag{3.12}$$

$$\check{f}(x) = \sum_l C_l\check{\psi}_l(x), \qquad C_l = \frac{\int \check{\psi}_l^*(x)\check{f}(x)\,dx}{\int |\check{\psi}_l(x)|^2\,dx}, \tag{3.13}$$

where $\check{\psi}_l(x) \equiv s^{-1/2}\psi_l(x/s)$. The dimension of $\check{f}(x)$ and $\check{\psi}_l(x)$ (or $f(u)$ and $\psi_l(u)$ for that matter) need not be the same; it is only necessary that the dimension of $\check{\psi}_l(x)$ multiplied by that of $C_l$ be equal to that of $\check{f}(x)$. If the basis is normalized, the denominators in the above expressions would normally not be written, leaving what seems to be a dimensionally inconsistent expression. In such cases, the reader should remember that there is an implicit value of unity in the denominator which carries a dimension. Such implicit unity values which actually carry a dimension appear in other contexts as well, a possibility the reader should remain aware of. For instance, a quadratic-phase optical signal expressed as $\exp[i\pi\sigma(x^2+y^2)/z]$ might not seem to have a dimension, but it actually has the dimension of the scalar amplitude of light hidden in the value of unity in front of the exponential function.

In two dimensions, the corresponding conventions are

$$\check{f}(x,y) \equiv \frac{1}{s}f(x/s, y/s) \equiv \frac{1}{s}f(u,v), \tag{3.14}$$

$$\check{F}(\sigma_x,\sigma_y) \equiv sF(s\sigma_x, s\sigma_y) \equiv sF(\mu,\nu), \tag{3.15}$$

$$\check{h}(x,y;x',y') \equiv s^{-2}h(x/s, y/s; x'/s, y'/s) \equiv s^{-2}h(u,v;u',v'), \tag{3.16}$$

where $u \equiv x/s$, $v \equiv y/s$, $\mu \equiv s\sigma_x$, $\nu \equiv s\sigma_y$.

In general, all of the results and equations in dimensionless chapters will remain to hold true, for the simple reason that the derivations leading to them cannot "know" whether a dimension is attributed to a variable or not. We simply replace $u, v$ and $\mu, \nu$ with $x, y$ and $\sigma_x, \sigma_y$, functions such as $f(u,v)$ and $F(\mu,\nu)$ by $\check{f}(x,y)$ and $\check{F}(\sigma_x,\sigma_y)$, and kernels such as $h(u,v;u',v')$ by $\check{h}(x,y;x',y')$, as shown in the above basic relations. (The circumflex is not used for physical parameter distributions which never appear in dimensionless contexts, such as the refractive index, which we simply write as $n(x,y,z)$ rather than $\check{n}(x,y,z)$.)

We will also replace certain dimensionless parameters with parameters of appropriate physical dimension. For instance, chirp multiplication takes a function $f(u)$ to $\exp(-i\pi qu^2)f(u)$.

Physically, we will see that this models the action of a thin lens, and that the corresponding dimensional relation takes $\check{f}(x)$ to $\exp(-i\pi x^2/\lambda f)\check{f}(x)$, from which we see that the parameter $q$ has been replaced with the physical parameter $1/\lambda f$ whose dimension is equal to the inverse square of the dimension of $x$. Two sets of parameters are important enough that we will use a special convention to stress that they are the dimensional counterparts of the corresponding dimensionless parameters: $\check{A}$, $\check{B}$, $\check{C}$, $\check{D}$ or $\check{\alpha}$, $\check{\beta}$, $\check{\gamma}$ will denote the dimensional counterparts of the dimensionless parameters $A$, $B$, $C$, $D$ or $\alpha$, $\beta$, $\gamma$ characterizing linear canonical transforms.

Finally, we mention some additional notation that will appear transiently at the beginning of the next section and then disappear. We use $\tilde{f}(x,t)$ to denote functions of both space and time and $\check{f}(x,f_{\mathrm{o}})$ to denote functions of space and temporal frequency $f_{\mathrm{o}}$ ($\check{f}(x,f_{\mathrm{o}})$ is the temporal Fourier transform of $\tilde{f}(x,t)$). The spatial Fourier transform of $\check{f}(x,f_{\mathrm{o}})$ is denoted by $\check{F}(\sigma_x,f_{\mathrm{o}})$. Finally, when we concentrate on monochromatic signals and the frequency dependence is dropped, we write $\check{f}(x)$ or $\check{F}(\sigma_x)$, which are then employed throughout the rest of the book.

The conventions we adopt are not new or unusual; they are implicitly employed in many texts without any special discussion and usually go unnoticed by readers, until they stumble upon a dimensionally inconsistent equation or obtain some dimensional paradox. We believe a conscious awareness of dimensions is important enough to warrant an explicit discussion.

## 3.3 Wave optics

Optical signals are most commonly represented by the complex amplitude or intensity of light as a function of space and/or time. We will usually deal with systems in which signals are represented by the amplitude of light as a function of the transverse spatial coordinates $x$ and $y$ over a given plane $z =$ constant. The distribution of light representing the signal propagates from left to right in the positive $z$ direction, being operated on or transformed in the process. The distributions of light on the input and output planes in figure 3.1 represent the input and output of the optical system.

In this book we will mostly restrict our attention to optical systems consisting of linear and time-invariant components, and assume that the behavior of light can be adequately described by a scalar theory. Sections of free space or other homogeneous or inhomogeneous media will also be treated as components; in any event we will assume these to be linear, isotropic, and nondispersive. We will also assume that we are dealing with systems and light sources for which we can assume that the light is *quasi-monochromatic* (effectively temporally coherent). We will however discuss both spatially coherent and spatially incoherent systems.

The output of such a system is related to its input by a relation of the form

$$\tilde{g}(\mathbf{r},t) = \int_{\mathbf{r}}\int_{t} \tilde{h}(\mathbf{r},\mathbf{r}',t-t')\tilde{f}(\mathbf{r}',t')\,dt'\,d\mathbf{r}', \tag{3.17}$$

where $\tilde{g}(\mathbf{r},t)$ and $\tilde{f}(\mathbf{r},t)$ represent the amplitude of light as functions of space and time over the output and input planes.

The wavelength (or center wavelength) of the light used will be denoted by $\lambda$, in the medium of propagation. The speed of light in vacuum will be denoted by $c$ so that the frequency $f_{\mathrm{oc}}$ of a monochromatic optical wave of wavelength $\lambda$ satisfies $c = f_{\mathrm{oc}} n_{\mathrm{A}} \lambda$, where $n_{\mathrm{A}}$ is the refractive index of some medium A. The wavenumber is defined as $\sigma \equiv 1/\lambda$ and is equal to the magnitude of the wavevector $\boldsymbol{\sigma}$. For a plane wave, $\boldsymbol{\sigma}$ points in the direction of propagation of the wave. $\mathbf{r}$ denotes the vector $(x,y)$ or $(x,y,z)$ and $\boldsymbol{\sigma}$ denotes the vector $(\sigma_x,\sigma_y)$ or $(\sigma_x,\sigma_y,\sigma_z)$, depending on the context.

We will further restrict our attention to *first-order* centered systems. These are systems which have the general appearance of the system shown in figure 3.1 and for which a number of simplifying approximations can be employed. For the time being we satisfy ourselves by noting that these are precisely the same approximations employed in the theory of optical systems referred to as Fourier optics (Goodman 1996).

The intensity of a wave at a certain point is defined as the power per unit area at that point. Poynting's theorem (Ramo, Whinnery, and Van Duzer 1994) gives the intensity in terms of the electric field vector $\mathbf{E}$ as $|\mathbf{E}|^2/\eta$ where $\eta$ is the intrinsic impedance of free space. We will assume that the scalar amplitudes we are working with have been normalized so that the intensity is given by (Saleh and Teich 1991, page 44)

$$\tilde{I}_{\tilde{f}}(\mathbf{r},t) \equiv 2\overline{[\tilde{f}(\mathbf{r},t)]^2}, \tag{3.18}$$

where the time average denoted by the overbar is taken over an interval much longer than the optical period but sufficiently shorter than the time over which the envelope of $\tilde{f}(\mathbf{r},t)$ changes appreciably. In the monochromatic case, the real field $\tilde{f}(\mathbf{r},t)$ can be written as

$$\tilde{f}(\mathbf{r},t) = \breve{A}(\mathbf{r}) \cos[2\pi f_{\mathrm{oc}} t + \breve{\varphi}(\mathbf{r})], \tag{3.19}$$

where $\breve{A}(\mathbf{r})$ and $\breve{\varphi}(\mathbf{r})$ are real-valued functions. The corresponding analytic signal $\tilde{f}_{\mathrm{as}}(\mathbf{r},t)$ is $\breve{A}(\mathbf{r}) \exp[-i2\pi f_{\mathrm{oc}} t - i\breve{\varphi}(\mathbf{r})]$, and the complex amplitude which we will denote by $\breve{f}(\mathbf{r})$ is $\breve{A}(\mathbf{r}) \exp[-i\breve{\varphi}(\mathbf{r})]$.

### 3.3.1 The wave equation

The function $\tilde{f}(\mathbf{r},t)$ describing the scalar amplitude distribution of light as a function of the position vector $\mathbf{r} = (x,y,z)$ and time $t$ in a linear isotropic nondispersive medium with time-invariant refractive index distribution $n(\mathbf{r})$ satisfies the wave equation

$$\frac{\partial^2 \tilde{f}}{\partial x^2} + \frac{\partial^2 \tilde{f}}{\partial y^2} + \frac{\partial^2 \tilde{f}}{\partial z^2} - \frac{n^2(\mathbf{r})}{c^2} \frac{\partial^2 \tilde{f}}{\partial t^2} = 0. \tag{3.20}$$

We will assume that the variation of $n(\mathbf{r})$ is small over distances comparable to the wavelengths of light we deal with. Since the wave equation is linear, any linear superposition

of solutions is also a solution. The wave equation can be solved uniquely if the distribution of $\tilde{f}$ over some surface is specified at some time. Furthermore, if new boundary conditions are specified as the linear superposition of some set of boundary conditions for which the solution is already known, the new solution can be written as the same linear superposition.

We take the temporal Fourier transform of both sides of the above equation, with the temporal Fourier transform of $\tilde{f}(\mathbf{r}, t)$ defined as

$$\check{f}(\mathbf{r}, f_\mathrm{o}) = \int \tilde{f}(\mathbf{r}, t) e^{i2\pi f_\mathrm{o} t}\, dt. \tag{3.21}$$

The temporal Fourier transform is defined with a positive sign in the exponent, in contrast to the spatial Fourier transform which is defined in the conventional manner with a negative sign in the exponent. This is consistent with the interpretation of the spatio-temporal Fourier transform as the coefficient of expansion in terms of planes waves of the form $\exp[i2\pi(\boldsymbol{\sigma} \cdot \boldsymbol{r} - f_\mathrm{o} t)]$ (Saleh and Teich 1991, page 925). The temporal Fourier transform of the wave equation is

$$\frac{\partial^2 \check{f}}{\partial x^2} + \frac{\partial^2 \check{f}}{\partial y^2} + \frac{\partial^2 \check{f}}{\partial z^2} + \frac{4\pi^2 n^2 f_\mathrm{o}^2}{c^2} \check{f} = 0. \tag{3.22}$$

This equation is known as the Helmholtz equation. If we solve this equation for $\check{f}(\mathbf{r}, f_\mathrm{o})$ for all $f_\mathrm{o}$, a temporal inverse Fourier transform operation will give us $\tilde{f}(\mathbf{r}, t)$. The analytic signal $\tilde{f}_\mathrm{as}(\mathbf{r}, t)$ corresponding to $\tilde{f}(\mathbf{r}, t)$ is defined as the inverse Fourier transform of $[1 + \mathrm{sgn}(f_\mathrm{o})]\check{f}(\mathbf{r}, f_\mathrm{o})$ and also satisfies the wave equation 3.20. In the event that we are dealing with monochromatic waves of specified frequency $f_\mathrm{oc}$, we have $\check{f}(\mathbf{r}, f_\mathrm{o}) = 0.5\,\check{f}(\mathbf{r})\delta(f_\mathrm{o} - f_\mathrm{oc}) + 0.5\,\check{f}^*(\mathbf{r})\delta(f_\mathrm{o} + f_\mathrm{oc})$. In this case, the Fourier transform of the analytic signal is simply $\check{f}(\mathbf{r})\delta(f_\mathrm{o} - f_\mathrm{oc})$ and the signal can be represented by the complex amplitude or phasor $\check{f}(\mathbf{r})$. That $\check{f}(\mathbf{r})$ also satisfies the Helmholtz equation can be shown most directly by substituting a monochromatic component of the form $\check{f}(\mathbf{r}) \exp(-i2\pi f_\mathrm{oc} t)$ in equation 3.20 to obtain

$$\frac{\partial^2 \check{f}}{\partial x^2} + \frac{\partial^2 \check{f}}{\partial y^2} + \frac{\partial^2 \check{f}}{\partial z^2} + \frac{4\pi^2 n^2 f_\mathrm{oc}^2}{c^2} \check{f} = 0. \tag{3.23}$$

In the monochromatic case, the intensity is simply related to $\check{f}(\mathbf{r})$ as follows:

$$\check{I}_{\check{f}}(\mathbf{r}) = |\check{f}(\mathbf{r})|^2, \tag{3.24}$$

and does not depend on time (Saleh and Teich 1991, page 46).

Two complete sets of solutions of equation 3.23 for a homogeneous medium $n(\mathbf{r}) = n = \text{constant}$, are the set of plane waves and the set of spherical waves respectively given by (Saleh and Teich 1991, pages 47–48):

$$\check{f}(x, y, z) = e^{i2\pi \boldsymbol{\sigma} \cdot \boldsymbol{r}} = e^{i2\pi(\sigma_x x + \sigma_y y + \sigma_z z)}, \qquad \boldsymbol{\sigma} \in \mathbf{R}^3, \tag{3.25}$$

$$\check{f}(x, y, z) = \frac{e^{i2\pi\sigma r}}{i\lambda r}, \qquad \sigma \in \mathbf{R}, \tag{3.26}$$

where $r \equiv |\mathbf{r}|$ and $\sigma^2 \equiv |\boldsymbol{\sigma}|^2 = \sigma_x^2 + \sigma_y^2 + \sigma_z^2$. For each value of $\sigma$, there is a corresponding optical frequency $f_{\text{oc}} = \sigma c/n$. Other solutions can be expressed as linear superpositions of the members of either of these or yet other sets of solutions, as will be elaborated later.

Although the above solutions may be verified by direct substitution, it is instructive to note how they can be directly obtained. First take the three-dimensional spatial Fourier transform of equation 3.22, replacing the derivatives $\partial^2/\partial x^2$, $\partial^2/\partial y^2$, $\partial^2/\partial z^2$ with $(i2\pi\sigma_x)^2$, $(i2\pi\sigma_y)^2$, $(i2\pi\sigma_z)^2$. This results in an equation for the four-dimensional spatio-temporal Fourier transform of $\tilde{f}$:

$$\left[\sigma_x^2 + \sigma_y^2 + \sigma_z^2 - (nf_{\text{o}}/c)^2\right] \check{F}(\sigma_x, \sigma_y, \sigma_z, f_{\text{o}}) = 0, \tag{3.27}$$

where the spatial Fourier transform $\check{F}(\sigma_x, \sigma_y, \sigma_z, f_{\text{o}})$ of $\check{f}(x, y, z, f_{\text{o}})$ is defined as

$$\check{F}(\sigma_x, \sigma_y, \sigma_z, f_{\text{o}}) = \iiint \check{f}(x, y, z, f_{\text{o}}) e^{-i2\pi(\sigma_x x + \sigma_y y + \sigma_z z)}\, dx\, dy\, dz. \tag{3.28}$$

It immediately follows from equation 3.27 that $\check{F}(\sigma_x, \sigma_y, \sigma_z, f_{\text{o}})$ can be nonzero only where $\sigma_x^2 + \sigma_y^2 + \sigma_z^2 - (nf_{\text{o}}/c)^2 = 0$. This implies delta-function-type solutions which can be inverse transformed to obtain the plane and spherical wave solutions presented, a task we leave to the reader. If the light is monochromatic with frequency $f_{\text{oc}}$, then $\sigma_x, \sigma_y, \sigma_z$ must satisfy $\sigma^2 = \sigma_x^2 + \sigma_y^2 + \sigma_z^2 = (nf_{\text{oc}}/c)^2$.

If we know the spatial variation of a plane wave at some plane $z = z_1$, we can easily determine its spatial variation at any other plane $z = z_2 = z_1 + d$. For instance, if the spatial variation of a wave with $\sigma = 10$ at the plane $z = 0$ is given by $K \exp[i2\pi(4x + 6y)]$ where $K$ is some complex constant, we can deduce the complete three-dimensional distribution of the wave as

$$K e^{i2\pi(4x + 6y + \sqrt{10^2 - 4^2 - 6^2}\, z)} \tag{3.29}$$

and thus determine the complex amplitude distribution at any other plane. For instance, at $z = 5.2$, we have $K \exp[i2\pi(4x + 6y + 6.9z)] = K \exp[i2\pi(4x + 6y + 36)] = K \exp[i2\pi(4x + 6y)] \exp[i2\pi(36)]$. We see that the form of the distribution remains unchanged, but is affected by a phase factor of $\exp[i2\pi(36)]$. (In this and similar numerical discussions we assume the dimensions of the numerical factors are implied. For instance, $\sigma = 10$ has the dimensions of inverse length.)

Figure 3.3 shows the wavefronts of a plane wave making angle $\theta_x$ with the $y$-$z$ plane ($\theta_x$ is the complement to $\pi/2$ of the angle made with the $x$ axis). We wish to examine the variation of this wave as a function of $x$ on the $z = 0$ plane. The period of the optical wave is $\lambda$. Along the $x$ axis this translates into a period of $\lambda/\sin\theta_x$, which corresponds to the spatial frequency $\sigma_x$ of the wave along the $x$ direction. Similar considerations apply for the $y$ direction. Thus the following hold:

$$\sin\theta_x = \lambda\sigma_x, \tag{3.30}$$

$$\sin\theta_y = \lambda\sigma_y. \tag{3.31}$$

The $z = 0$ profile of a plane wave $\exp[i2\pi(\sigma_x x + \sigma_y y + \sigma_z z)]$ is a two-dimensional spatial harmonic whose spatial frequencies are related to the direction of propagation of the plane wave as given by equations 3.30 and 3.31.

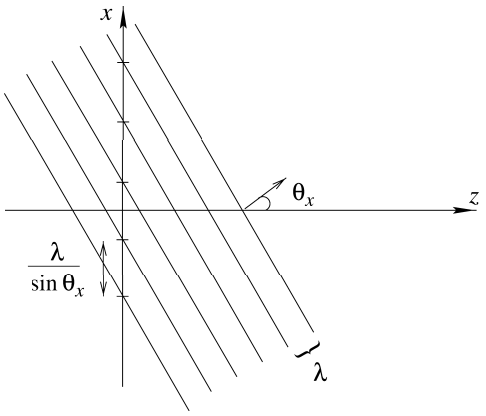


Figure 3.3: Wavefronts of a plane wave making angle $\theta_x$ with the $z$ axis.

The spherical wave solution given in equation 3.26 is the solution corresponding to a point source $\delta(x, y, z)$ centered at the origin; that is, it is the *Green's function* of the wave equation. For future reference, we also note the common *Fresnel approximation* of a spherical wave. With $r^2 = x^2 + y^2 + z^2$ and under the assumption that $z^2 \gg x^2 + y^2$, we obtain

$$\check{f}(x, y, z) = \frac{e^{i2\pi\sigma r}}{i\lambda r} \approx \frac{e^{i2\pi\sigma z}}{i\lambda z} \exp\left[i2\pi\sigma\left(\frac{x^2 + y^2}{2z}\right)\right], \tag{3.32}$$

where we replaced $r \approx z$ in the denominator but $r \approx z + (x^2 + y^2)/2z$ in the exponent because of the greater sensitivity of the imaginary exponent to small changes (Saleh and Teich 1991, page 49). Equation 3.32 is also referred to as a *parabolic wave*.

Each of the following three subsections will address from a different perspective the problem of obtaining the distribution of light on a plane $z = z_2 = z_1 + d$, given the distribution of light on a plane $z = z_1$. Readers willing to take the results for granted may skip to section 3.4.

### 3.3.2 Plane wave decomposition

First, let us assume that the distribution of light at $z = z_1$, which we refer to as the input $\check{f}(x, y)$, is of the form

$$\check{f}(x, y) = Ke^{i2\pi(\sigma_x x + \sigma_y y)}, \tag{3.33}$$

where $K$ is a complex constant and $\sigma_x$ and $\sigma_y$ are the spatial frequencies of this two-dimensional harmonic function. We recognize this as the profile of a plane wave with

wavevector components $\sigma_x$, $\sigma_y$, and $\sigma_z = \sqrt{\sigma^2 - \sigma_x^2 - \sigma_y^2}$:

$$K e^{i2\pi(\sigma_x x + \sigma_y y)} \exp\left[i2\pi\sqrt{\sigma^2 - \sigma_x^2 - \sigma_y^2}\,(z - z_1)\right]. \tag{3.34}$$

The output $\check{g}(x, y)$ observed at the plane $z = z_2 = z_1 + d$ is then given by

$$\check{g}(x, y) = K e^{i2\pi(\sigma_x x + \sigma_y y)} \exp\left(i2\pi\sqrt{\sigma^2 - \sigma_x^2 - \sigma_y^2}\,d\right) = \check{H}(\sigma_x, \sigma_y)\check{f}(x, y), \tag{3.35}$$

where

$$\check{H}(\sigma_x, \sigma_y) = \exp\left(i2\pi\sqrt{\sigma^2 - \sigma_x^2 - \sigma_y^2}\,d\right). \tag{3.36}$$

We see that harmonic functions are eigenfunctions of propagation over a section of free space, with eigenvalue $\check{H}(\sigma_x, \sigma_y)$. Since harmonics are profiles of plane waves, it is sometimes also said that plane waves are eigenfunctions of propagation in free space. (It is also possible to pose the same in terms of temporal evolution by showing that if at any instant in time we observe a plane wave in space, we will observe a plane wave at all consecutive times, which more directly justifies referring to plane waves as eigenfunctions of propagation in free space.)

Now, let us return to the problem of relating the output to the input when the input is an arbitrary distribution of light, and not necessarily a two-dimensional harmonic. An arbitrary distribution of light at the plane $z_1$, denoted by $\check{f}(x, y)$, can be written as a linear superposition of harmonics as follows:

$$\check{f}(x, y) = \iint \check{F}(\sigma_x, \sigma_y)\, e^{i2\pi(\sigma_x x + \sigma_y y)}\, d\sigma_x\, d\sigma_y, \tag{3.37}$$

where $\check{F}(\sigma_x, \sigma_y)$ is the Fourier transform of $\check{f}(x, y)$. Since we know that a linear superposition of inputs will produce the same linear superposition of outputs, the distribution of light at the plane $z_2$ can be obtained easily as

$$\check{g}(x, y) = \iint \check{F}(\sigma_x, \sigma_y)\, e^{i2\pi\sqrt{\sigma^2 - \sigma_x^2 - \sigma_y^2}\,d}\, e^{i2\pi(\sigma_x x + \sigma_y y)}\, d\sigma_x\, d\sigma_y. \tag{3.38}$$

We see that the effect of free-space propagation in the Fourier domain is

$$\check{G}(\sigma_x, \sigma_y) = e^{i2\pi\sqrt{\sigma^2 - \sigma_x^2 - \sigma_y^2}\,d}\, \check{F}(\sigma_x, \sigma_y) = \check{H}(\sigma_x, \sigma_y)\check{F}(\sigma_x, \sigma_y), \tag{3.39}$$

where $\check{H}(\sigma_x, \sigma_y)$ is given by equation 3.36. This result can be written in the space domain as a two-dimensional convolution

$$\check{g}(x, y) = \check{h}(x, y) * * \check{f}(x, y), \tag{3.40}$$

where $\check{h}(x, y)$ is the inverse Fourier transform of $\check{H}(\sigma_x, \sigma_y)$. A simple analytical expression for $\check{h}(x, y)$ is not known. However, the exponent of $\check{H}(\sigma_x, \sigma_y)$ is commonly approximated as

$$\sqrt{\sigma^2 - \sigma_x^2 - \sigma_y^2} \approx \sigma - \frac{\sigma_x^2 + \sigma_y^2}{2\sigma}, \tag{3.41}$$

under the assumption that $\sigma^2 \approx \sigma_z^2 \gg (\sigma_x^2 + \sigma_y^2)$. Then we approximately obtain

$$\check{H}(\sigma_x, \sigma_y) = e^{i2\pi\sigma d} \exp[-i\pi\lambda d(\sigma_x^2 + \sigma_y^2)]. \tag{3.42}$$

For a discussion of the validity of this approximation, known as the Fresnel approximation, see Goodman 1996 or Saleh and Teich 1991. With this approximation, the inverse transform $\check{h}(x, y)$ becomes

$$\check{h}(x, y) = h_0 \exp[i\pi(x^2 + y^2)/\lambda d], \tag{3.43}$$
$$h_0 = \frac{e^{i2\pi\sigma d}}{i\lambda d},$$

which is nothing but the parabolic approximation of the spherical wave given earlier as equation 3.32. The relation between $\check{f}$ and $\check{g}$ takes the form

$$\check{g}(x, y) = \check{h}(x, y) * *\check{f}(x, y) = h_0 \iint e^{\frac{i\pi}{\lambda d}[(x-x')^2+(y-y')^2]} \check{f}(x', y')\, dx'\, dy', \tag{3.44}$$

an expression known as the Fresnel integral or Fresnel transform (see equation 1.29). It gives the amplitude distribution of light at the plane $z = z_2 = z_1 + d$ in terms of that at the plane $z = z_1$. It is the solution of the paraxial Helmholtz equation to be discussed in the next subsection, as can be shown by direct substitution. On the other hand, the exact form with the square root in the exponent is the solution of the exact Helmholtz equation 3.23. The fact that the system represented by the Fresnel integral is space-invariant, is consistent with the fact that the eigenfunctions of this system are harmonic functions (page 9).

Here we have not included the classic derivations through which the Fresnel integral is traditionally arrived at (for instance, see Goodman 1996, Iizuka 1987, Yu 1983, Born and Wolf 1980). We only note that the kernel appearing in equation 3.44 is nothing but the Fresnel approximation of a spherical wave (see equation 3.32). Thus equation 3.44 is essentially an approximation of a weighted superposition of spherical waves. This interpretation is known as the Huygens-Fresnel principle. Each point in the input plane is considered to be a secondary source with amplitude $\check{f}(x, y)$, which gives rise to a spherical wave. Superposing all of these spherical waves gives us the amplitude distribution $\check{g}(x, y)$ at the output plane. The mathematical expression of the Huygens-Fresnel principle is known as the Rayleigh-Sommerfeld diffraction formula (Goodman 1996):

$$\check{g}(x, y) = \frac{1}{i\lambda} \iint \check{f}(x', y') \frac{e^{i2\pi r/\lambda}}{r} \cos\theta\, dx'\, dy', \tag{3.45}$$

where $r = \sqrt{d^2 + (x - x')^2 + (y - y')^2}$ and $\theta$ is the angle between the line joining the input point $(x', y')$ to the output point $(x, y)$, and the $z$ axis (so that $\cos\theta = d/r$). This integral is interpreted as a superposition of diverging spherical waves originating from "secondary sources" located at the input plane. $\cos\theta$ is an "obliquity factor." For an excellent discussion of the developments leading to this equation, see Goodman 1996. It is possible to arrive at equation 3.44 from equation 3.45 by employing the Fresnel approximation of the spherical wave given in equation 3.32.

### 3.3.3 The paraxial wave equation

Here we present a number of approaches closely related to those of the previous subsection, but which nevertheless provide different perspectives. We take the Fourier transform of equation 3.23 with respect to $x$ and $y$ (or equivalently, we consider solutions of the form $\check{f}(x,y,z) = \check{F}(\sigma_x,\sigma_y,z)\exp[i2\pi(\sigma_x x + \sigma_y y)]$) to obtain

$$\frac{\partial^2 \check{F}}{\partial z^2} = -4\pi^2(\sigma^2 - \sigma_x^2 - \sigma_y^2)\check{F}. \tag{3.46}$$

A solution of this equation corresponding to propagation in the positive $z$ direction is

$$\check{F}(\sigma_x,\sigma_y,z) = e^{i2\pi\sqrt{\sigma^2-\sigma_x^2-\sigma_y^2}\,z}\ \check{F}(\sigma_x,\sigma_y,0). \tag{3.47}$$

We use the two-dimensional functions $\check{f}(x,y)$ and $\check{g}(x,y)$ to denote two-dimensional fields on the $z = z_1$ and $z = z_2$ planes respectively, and the three-dimensional function $\check{f}(x,y,z)$ to denote three-dimensional fields. Thus we will write $\check{F}(\sigma_x,\sigma_y,z_1) \equiv \check{F}(\sigma_x,\sigma_y)$ and $\check{F}(\sigma_x,\sigma_y,z_2) \equiv \check{G}(\sigma_x,\sigma_y)$. Now, writing equation 3.47 once for $z_1$ and once for $z_2$ and eliminating $\check{F}(\sigma_x,\sigma_y,0)$, we obtain

$$\check{G}(\sigma_x,\sigma_y) = e^{i2\pi\sqrt{\sigma^2-\sigma_x^2-\sigma_y^2}\,d}\ \check{F}(\sigma_x,\sigma_y), \tag{3.48}$$

which is the same as equation 3.39, from which the same argument leads us to equation 3.44 ($d = z_2 - z_1$).

Another approach is as follows. Restricting ourselves to waves traveling in the positive $z$ direction, the "square root" of equation 3.46 may be written as

$$\frac{\partial \check{F}}{\partial z} = +i2\pi\sqrt{\sigma^2 - \sigma_x^2 - \sigma_y^2}\,\check{F}. \tag{3.49}$$

(Although we are being far from rigorous here, our final result will nevertheless be correct.) Introducing what is essentially the Fresnel approximation at this point we obtain

$$\frac{\partial \check{F}}{\partial z} = +i2\pi(\sigma - \sigma_x^2/2\sigma - \sigma_y^2/2\sigma)\check{F}. \tag{3.50}$$

The solution to equation 3.50 is

$$\check{F}(\sigma_x,\sigma_y,z) = e^{i2\pi(\sigma-\sigma_x^2/2\sigma-\sigma_y^2/2\sigma)z}\ \check{F}(\sigma_x,\sigma_y,0), \tag{3.51}$$

from which one can deduce equation 3.42.

It is also possible to work in the space domain. Starting directly with equation 3.23 and again formally taking the "square root" of the operators yields

$$\frac{\partial \check{f}}{\partial z} = +i\sqrt{4\pi^2\sigma^2 + \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}}\ \check{f}. \tag{3.52}$$

Introducing what is essentially the Fresnel approximation at this point we obtain

$$\frac{\partial \check{f}}{\partial z} = +i\left[2\pi\sigma + \frac{1}{4\pi\sigma}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\right]\check{f}. \tag{3.53}$$

The Fresnel integral given in equation 3.44 is an exact solution of this differential equation. These discussions were inspired by Bastiaans 1979c.

A *paraxial wave* is one whose wavevector(s) make small angles with the optical axis. In other words, the wavefront normals are paraxial rays (Saleh and Teich 1991). Just as a temporally narrowband signal has harmonic components concentrated around a certain center frequency, a paraxial wave has plane-wave components whose wavevectors are concentrated around the optical axis. The major spatial variation is along the $z$ axis, so that we can write $\check{f}(\mathbf{r})$ in the form $\check{f}(\mathbf{r}) = \check{A}(\mathbf{r})\exp(i2\pi\sigma z)$, where $\check{A}(\mathbf{r})$ is a complex envelope. Paraxial waves can be interpreted as spatially narrowband modulated plane waves, just as temporally narrowband signals can be interpreted as temporally narrowband modulated harmonics. Now, substituting this form for $\check{f}(\mathbf{r})$ in equation 3.53 we obtain

$$\frac{\partial^2 \check{A}}{\partial x^2} + \frac{\partial^2 \check{A}}{\partial y^2} + i4\pi\sigma\frac{\partial \check{A}}{\partial z} = 0. \tag{3.54}$$

Equations such as equation 3.53 or 3.54 are known as paraxial wave equations or paraxial Helmholtz equations. A more conventional derivation of equation 3.54 is as follows (Saleh and Teich 1991): We substitute $\check{f}(\mathbf{r}) = \check{A}(\mathbf{r})\exp(i2\pi\sigma z)$ in equation 3.23, and employ $\partial\check{A}/\partial z \ll 2\pi\sigma\check{A}$ and $\partial^2\check{A}/\partial z^2 \ll 4\pi^2\sigma^2\check{A}$, which are mathematical statements of paraxiality, and are also referred to as the slowly varying envelope approximation (since $\check{A}(\mathbf{r})$ varies slowly with $\mathbf{r}$). This derivation again results in equation 3.54.

The parabolic wave is an exact solution of equation 3.54 (Saleh and Teich 1991):

$$\check{A}(\mathbf{r}) = \frac{1}{i\lambda z}e^{i2\pi\sigma(x^2+y^2)/2z}. \tag{3.55}$$

In preparation for the next subsection, we also write the following more general solution of equation 3.54 (Saleh and Teich 1991):

$$\check{A}(\mathbf{r}) = \frac{1}{i\lambda\check{q}(z)}e^{i2\pi\sigma(x^2+y^2)/2\check{q}(z)}, \tag{3.56}$$

where $\check{q}(z) = z - i\check{z}$, $\check{z}$ = constant. $\check{q}$ is *not* a dimensional version of the parameter $q$ appearing in chapter 1, but we nevertheless choose this notation to conform with convention.

### 3.3.4 Hermite-Gaussian beams

As already mentioned, a more general solution of equation 3.54 is

$$\check{A}(\mathbf{r}) \propto \frac{1}{i\lambda\check{q}(z)}\exp[i\pi\sigma(x^2+y^2)/\check{q}(z)], \tag{3.57}$$

where $\check{q}(z) = z - i\check{z}$. Here $\check{z}$ is a constant which is referred to as the Rayleigh range. $\check{q}(z)$ is referred to as the complex radius of curvature or simply as the $\check{q}$-parameter. If we define the beam size $W(z)$ and the wavefront radius of curvature $R(z)$ through

$$\frac{1}{\check{q}(z)} \equiv \frac{1}{R(z)} + i\frac{\lambda}{W^2(z)}, \tag{3.58}$$

it is possible to show that

$$\check{A}(\mathbf{r}) = \frac{2^{1/2}}{W(z)} \exp\left[-\frac{\pi(x^2+y^2)}{W^2(z)}\right] \exp\left[i2\pi\sigma\frac{(x^2+y^2)}{2R(z)} - i\zeta(z)\right], \tag{3.59}$$

where

$$W(z) \equiv W_0\left[1 + (z/\check{z})^2\right]^{1/2}, \tag{3.60}$$

$$R(z) \equiv z\left[1 + (\check{z}/z)^2\right], \tag{3.61}$$

$$\zeta(z) \equiv \arctan(z/\check{z}), \tag{3.62}$$

and $W_0^2 \equiv W^2(0) \equiv \lambda\check{z}$. We have normalized $\check{A}(\mathbf{r})$ so that it has unit energy. We are choosing to employ the parameter $W(z)$ which we refer to as the beam size, rather than the more commonly used $w(z)$ known as the beam radius. These two parameters are simply related by $W^2(z) = \pi w^2(z)$ so that $W_0^2 = \pi w_0^2$. The interpretation of these parameters are discussed in many texts such as Saleh and Teich 1991. The distribution of light represented by equation 3.59 is known as a *Gaussian beam.*

The paraxial approximation of a spherical wave originating at $x = y = z = 0$ can be written in the form $(i\lambda R)^{-1}\exp[i\pi(x^2+y^2)/\lambda R]$ where $R = z$ is the radius of curvature of the wavefronts at $z$. Comparing this with equation 3.57, we see that the Gaussian beam can be interpreted as a spherical wave with complex "radius" $\check{q}$. When $W_0 = \infty$, the beam has infinite transverse extent and the complex radius $\check{q}$ reduces to a real radius (equation 3.58). When $\check{q}$ is complex, the imaginary part manifests itself as the beam size:

$$e^{i\pi(x'^2+y'^2)/\lambda\check{q}} = e^{i\pi(x'^2+y'^2)/\lambda R} e^{-\pi(x'^2+y'^2)/W'^2}. \tag{3.63}$$

An even more general set of solutions of equation 3.54 are the Hermite-Gaussian beams. Unlike the set of plane waves, this is a discrete set with countably many members, enumerated by $(l, m)$:

$$\check{A}(\mathbf{r}) = \frac{1}{W(z)}\psi_l\left(\frac{x}{W(z)}\right)\psi_m\left(\frac{y}{W(z)}\right)\exp\left[i2\pi\sigma\frac{(x^2+y^2)}{2R(z)} - i(l+m+1)\zeta(z)\right], \tag{3.64}$$

where $\psi_l, \psi_m$ are the Hermite-Gaussian functions defined in section 1.5.2. The $l = 0, m = 0$ beam is simply the Gaussian beam given in equation 3.59. The Hermite-Gaussian beams share the parabolic wavefronts of the Gaussian beam, but exhibit different intensity distributions. Further discussion of their physical characteristics can be found in texts such as Saleh and Teich 1991.

Hermite-Gaussian beams are a complete orthonormal set of solutions of the paraxial wave equation. Just like the set of plane waves, they can be used to construct arbitrary solutions. Let us assume that the amplitude distribution of light at the plane $z = 0$ is given by $\check{f}(x, y)$. At $z = 0$, the Hermite-Gaussian beams are equal to $W_0^{-1}\,\psi_l(x/W_0)\,\psi_m(y/W_0)$. We first expand $\check{f}(x, y)$ as

$$\check{f}(x, y, 0) \equiv \check{f}(x, y) = \sum_{l=0}^{\infty}\sum_{m=0}^{\infty} C_{lm}\,\frac{1}{W_0}\,\psi_l\left(\frac{x}{W_0}\right)\,\psi_m\left(\frac{y}{W_0}\right), \tag{3.65}$$

where

$$C_{lm} = \iint \frac{1}{W_0}\,\psi_l\left(\frac{x}{W_0}\right)\,\psi_m\left(\frac{y}{W_0}\right)\check{f}(x, y)\,dx\,dy. \tag{3.66}$$

Now, it is possible to find the amplitude distribution at any $z$ as

$$\begin{aligned}\check{f}(x, y, z) \equiv \check{g}(x, y) = \sum_{l=0}^{\infty}\sum_{m=0}^{\infty} C_{lm}\,e^{i2\pi\sigma z}\,\frac{1}{W(z)}\,\psi_l\left(\frac{x}{W(z)}\right)\,\psi_m\left(\frac{y}{W(z)}\right)\\ \times\exp\left[i2\pi\sigma\frac{(x^2+y^2)}{2R(z)} - i(l+m+1)\zeta(z)\right].\end{aligned} \tag{3.67}$$

This result relating the distribution of light at an arbitrary plane to that at $z = 0$ is equivalent to Fresnel's integral, although it is not straightforward to show so analytically.

There exists many complete sets of solutions of the wave equation. Which set is preferred depends on the situation. Usually, it is best to work with the set that constitutes the eigenfunctions of the system through which light will pass. In free space, plane waves are the natural choice. Hermite-Gaussian functions, on the other hand, are eigenfunctions of spherical mirror resonators and periodic lens waveguides, and are thus useful for such systems. While Hermite-Gaussian functions are not strictly eigenfunctions of free space, they nevertheless retain their general form upon propagation through free space so that they can also be used with relative ease in this case as well. Gaussian beam propagation can also be formulated in cylindrical coordinates, in which case one obtain the so-called Laguerre-Gaussian beams instead of the Hermite-Gaussian beams.

Finally, we discuss how the parameters of Hermite-Gaussian beams change as a result of propagating through an optical system characterized by a linear canonical transform with matrix parameters $\check{A}, \check{B}, \check{C}, \check{D}$ (such optical systems will be discussed in detail in chapter 4). The $\check{q}$-parameter of the output beam can be simply related to the $\check{q}$-parameter of the input beam as follows:

$$(\lambda\check{q}_{\text{out}}) = \frac{\check{A}(\lambda\check{q}_{\text{in}}) + \check{B}}{\check{C}(\lambda\check{q}_{\text{in}}) + \check{D}}, \tag{3.68}$$

a result whose similarity to equation 2.156 is worth pointing out. The relationship between the dimensional parameters $\check{A}, \check{B}, \check{C}, \check{D}$ and the dimensionless $A, B, C, D$ will be discussed on page 162. The most straightforward way of deriving equation 3.68 is to take the linear canonical transform of a Gaussian beam.

Before proceeding any further, we must define a new parameter called the *accumulated Gouy phase shift* (Erden and Ozaktas 1997). The conventional Gouy phase shift $\zeta(z)$ defined in equation 3.62 is the on-axis phase of a Gaussian beam with respect to the beam waist in excess of the phase of a plane wave $\exp(i2\pi\sigma z)$. It is not independent from the beam size $W$ and the wavefront radius of curvature $R$. Of greater interest from an input-output perspective is the phase shift accumulated by the beam as it passes through several lenses and sections of free space, with respect to a single reference point in the system. Thus, we define the accumulated Gouy phase shift $\tilde{\zeta}$ of a Gaussian beam passing through an optical system as the on-axis phase accumulated by the beam in excess of the factor $\exp(i2\pi\sigma z)$. (This latter factor is the on-axis phase that would be accumulated by a plane wave). Mathematically,

$$-\tilde{\zeta} \equiv \angle[\check{A}(0,0,z_{\text{out}})] - \angle[\check{A}(0,0,z_{\text{in}})], \tag{3.69}$$

where $\check{A}(0,0,z_{\text{out}})$ and $\check{A}(0,0,z_{\text{in}})$ denote the on-axis values of the output and input Gaussian beams, and $\angle[\cdot]$ denotes the phase.

Let us now consider a Gaussian beam with parameters $W_{\text{in}}$ and $R_{\text{in}}$ input to an optical system characterized by the parameters $\check{A}, \check{B}, \check{C}, \check{D}$. Also, let the accumulated Gouy phase with respect to some reference point be given as $\tilde{\zeta}_{\text{in}}$. Denoting the corresponding output parameters by $W_{\text{out}}$, $R_{\text{out}}$, and $\tilde{\zeta}_{\text{out}}$, it is possible to show that (Erden and Ozaktas 1997)

$$W_{\text{out}}^2 = \left(\check{A} + \frac{\check{B}}{\lambda R_{\text{in}}}\right)^2 W_{\text{in}}^2 + \frac{\check{B}^2}{W_{\text{in}}^2}, \tag{3.70}$$

$$\frac{1}{\lambda R_{\text{out}}} = \frac{\left(\check{C} + \frac{\check{D}}{\lambda R_{\text{in}}}\right)\left(\check{A} + \frac{\check{B}}{\lambda R_{\text{in}}}\right) + \frac{\check{B}\check{D}}{W_{\text{in}}^4}}{\left(\check{A} + \frac{\check{B}}{\lambda R_{\text{in}}}\right)^2 + \frac{\check{B}^2}{W_{\text{in}}^4}}, \tag{3.71}$$

$$\tilde{\zeta}_{\text{out}} = \tilde{\zeta}_{\text{in}} + \arctan\left[\frac{\check{B}}{\left(\check{A} + \frac{\check{B}}{\lambda R_{\text{in}}}\right) W_{\text{in}}^2}\right]. \tag{3.72}$$

The first two of these equations are a consequence of equation 3.68, whereas the last one is demonstrated in Erden and Ozaktas 1997. The narrowest part of the beam, known as the waist, is observed where the beam size is minimum. When the beam is incident to the system at its waist so that $R_{\text{in}} = \infty$ and $W_{\text{in}} = W_0$, and if we assume $\tilde{\zeta}_{\text{in}} = 0$, the above relations reduce to

$$\frac{1}{\lambda R_{\text{out}}} = \frac{\check{B}/\check{A}}{\check{A}^2 W_0^4 + \check{B}^2} + \frac{\check{C}}{\check{A}}, \tag{3.73}$$

$$W_{\text{out}}^2 = \check{A}^2 W_0^2 + \check{B}^2/W_0^2, \tag{3.74}$$

$$\tilde{\zeta}_{\text{out}} = \arctan\left(\frac{\check{B}}{\check{A} W_{\text{in}}^2}\right). \tag{3.75}$$

The accumulated Gouy phase shift is an independent parameter which complements the beam size and wavefront radius of curvature to constitute three parameters which

uniquely characterize the beam with respect to a reference point in the system. This means that knowledge of these three parameters at any single plane in the system allows them to be calculated at any other plane in the system. Furthermore, measurement of these parameters allows one to uniquely recover the parameters characterizing the first-order system through which the beam propagates (Erden and Ozaktas 1997).

## 3.4 Wave optical characterization of optical components

### 3.4.1 Sections of free space

Despite the fact that they often consist of no more than the stretch of space between two other components, it is common to look upon sections of free space as optical components in their own right. The input and output of this component are the light distributions on the planes $z = z_1$ and $z = z_2 = z_1 + d$ bounding the section of free space from the left and the right. $d$ denotes the length of the section of free space. We have already determined the relation between the input and output in several ways in the preceding section. Here we consolidate the main results.

The output $\check{g}(x, y)$ is related to the input $\check{f}(x, y)$ through the Fresnel integral which is essentially a chirp convolution in the space domain and a chirp multiplication in the frequency domain:

$$\check{g}(x, y) = \check{h}(x, y) * * \check{f}(x, y), \tag{3.76}$$

$$\check{h}(x, y) = e^{i2\pi\sigma d} \frac{1}{i\lambda d} \exp\left[\frac{i\pi(x^2 + y^2)}{\lambda d}\right],$$

$$\check{G}(\sigma_x, \sigma_y) = \check{H}(\sigma_x, \sigma_y)\check{F}(\sigma_x, \sigma_y), \tag{3.77}$$

$$\check{H}(\sigma_x, \sigma_y) = e^{i2\pi\sigma d} \exp\left[-i\pi\lambda d(\sigma_x^2 + \sigma_y^2)\right].$$

Recall that $\check{H}(\sigma_x, \sigma_y)$ is the two-dimensional Fourier transform of $\check{h}(x, y)$ and the eigenvalue associated with the harmonic whose transverse spatial frequencies are $(\sigma_x, \sigma_y)$.

We also write for reference the one-dimensional versions of these results:

$$\check{g}(x) = \check{h}(x) * \check{f}(x), \tag{3.78}$$

$$\check{h}(x) = e^{i2\pi\sigma d} e^{-i\pi/4} \sqrt{\frac{1}{\lambda d}} \exp\left[\frac{i\pi x^2}{\lambda d}\right],$$

$$\check{G}(\sigma_x) = \check{H}(\sigma_x)\check{F}(\sigma_x), \tag{3.79}$$

$$\check{H}(\sigma_x) = e^{i2\pi\sigma d} \exp\left[-i\pi\lambda d\sigma_x^2\right].$$

### 3.4.2 Thin lenses

A thin lens is a special kind of spatial filter that plays an important role in realizing optical systems. Ideally, thin lenses are phase-only multiplicative filters with transmittance

function

$$\check{h}(x,y) = \exp\left[-\frac{i\pi(x^2+y^2)}{\lambda f}\right], \tag{3.80}$$

such that the amplitude distribution $\check{g}(x,y)$ immediately after the lens is related to the distribution $\check{f}(x,y)$ immediately before the lens through

$$\check{g}(x,y) = \check{h}(x,y)\check{f}(x,y). \tag{3.81}$$

Thin lenses are assumed to have no thickness. The parameter $f$ is referred to as the *focal length* of the lens. Lenses are referred to as positive or negative according to the sign of their focal length. The last equation can be written in the frequency domain as

$$\begin{aligned} \check{G}(\sigma_x,\sigma_y) &= \check{H}(\sigma_x,\sigma_y) * *\check{F}(\sigma_x,\sigma_y), \qquad (3.82)\\ \check{H}(\sigma_x,\sigma_y) &= -i\lambda f \exp\left[i\pi\lambda f(\sigma_x^2+\sigma_y^2)\right]. \end{aligned}$$

The one-dimensional versions of the above expressions are

$$\check{g}(x) = \check{h}(x)\check{f}(x), \tag{3.83}$$

$$\check{h}(x) = \exp\left[\frac{-i\pi x^2}{\lambda f}\right],$$

$$\begin{aligned} \check{G}(\sigma_x) &= \check{H}(\sigma_x) * \check{F}(\sigma_x), \qquad (3.84)\\ \check{H}(\sigma_x) &= e^{-i\pi/4}\sqrt{\lambda f}\ \exp\left[i\pi\lambda f\sigma_x^2\right]. \end{aligned}$$

Lenses can be realized by grinding convex or concave spherical surfaces on both sides of a thin slab of glass with refractive index $n_{\rm gl}$. The focal length is related to the radii of curvature of the surfaces by the formula (Saleh and Teich 1991)

$$\frac{1}{f} = \left(\frac{n_{\rm gl}}{n} - 1\right)\left(\frac{1}{R_{\rm right}} - \frac{1}{R_{\rm left}}\right). \tag{3.85}$$

$R_{\rm left}$ is the radius of the left surface and $R_{\rm right}$ is the radius of the right surface. The sign convention is such that surfaces which are convex towards the $+z$ direction are positive. Here $n$ is the refractive index of the medium in which the lens is situated, which in most cases is air so that $n \approx 1$. We also note that $\lambda$ is the wavelength in the same medium, and not in the glass.

Attenuation inside the lens and reflection from its surfaces are usually neglected, but sometimes it is desired to account for the finite size of the lens by defining a pupil function $\check{p}(x,y)$, which is unity within the lens aperture and zero outside:

$$\check{h}(x,y) = \check{p}(x,y)e^{-i\pi(x^2+y^2)/\lambda f}. \tag{3.86}$$

The effects of attenuations, reflections, and aberrations are sometimes handled by absorbing them into the pupil function.

The above results are for spherical lenses which are rotationally symmetric around the $z$ axis. More generally, the transmittance function may take the form

$$\check{h}(x,y) = \exp\left[\frac{-i\pi}{\lambda}\left(\frac{x^2}{f_{xx}} + \frac{2xy}{f_{xy}} + \frac{y^2}{f_{yy}}\right)\right], \tag{3.87}$$

where $f_{xx}$, $f_{xy}$, $f_{yy}$ are the three parameters characterizing the focal characteristics of the lens. Such lenses are often referred to as *anamorphic* lenses. If we express the above transmittance in terms of the rotated coordinates $(x', y')$ (where the $x'$ axis makes a positive angle $(1/2)\text{arccot}[f_{xy}(1/f_{yy} - 1/f_{xx})/2]$ with respect to the $x$ axis), the cross term disappears:

$$\check{h}(x',y') = \exp\left[\frac{-i\pi}{\lambda}\left(\frac{x'^2}{f_{x'x'}} + \frac{y'^2}{f_{y'y'}}\right)\right], \tag{3.88}$$

where $f_{x'x'}$ and $f_{y'y'}$ can be expressed in terms of $f_{xx}$, $f_{xy}$, $f_{yy}$. When $1/f_{x'x'} = 0$ (or $1/f_{y'y'} = 0$) we obtain a cylindrical lens with focal length $f_{y'y'}$ (or $f_{x'x'}$). Any anamorphic lens with given $f_{xx}$, $f_{xy}$, $f_{yy}$ can be simulated by two orthogonally positioned cylindrical lenses with focal lengths $f_{x'x'}$ and $f_{y'y'}$.

### 3.4.3 Quadratic graded-index media

A quadratic graded-index medium is a medium characterized by a refractive index distribution of the form (Yariv 1989)

$$n^2(x,y) = n_0^2[1 - (x/\chi_x)^2 - (y/\chi_y)^2], \tag{3.89}$$

where $\chi_x$, $\chi_y$, and $n_0$ are the medium parameters. More generally these parameters may be functions of $z$, but we will not treat this case. We will further restrict ourselves to the special case $\chi_x = \chi_y = \chi$. The one-dimensional version of this index distribution is taken as

$$n^2(x) = n_0^2[1 - (x/\chi)^2]. \tag{3.90}$$

We start by substituting equation 3.89 in equation 3.23:

$$\frac{\partial^2 \check{f}}{\partial x^2} + \frac{\partial^2 \check{f}}{\partial y^2} + \frac{\partial^2 \check{f}}{\partial z^2} + 4\pi^2\sigma^2[1 - (x^2+y^2)/\chi^2]\check{f} = 0, \tag{3.91}$$

where $\sigma = n_0 f_{\text{oc}}/c$. We seek positive-$z$ traveling eigensolutions of the form $\check{f}(x,y,z) = \check{A}(x,y)\exp(i2\pi\sigma_z z)$. Now, it is possible to show that if the function $\check{A}_x(x)$ satisfies

$$\frac{\partial^2 \check{A}_x}{\partial x^2} + \frac{4\pi^2\sigma^2}{\chi^2}\left(\frac{\sigma_x^2\chi^2}{\sigma^2} - x^2\right)\check{A}_x = 0, \tag{3.92}$$

and if the function $\check{A}_y(y)$ satisfies an identical equation in $y$, then $\check{A}(x,y) = \check{A}_x(x)\check{A}_y(y)$ is a solution of equation 3.91, provided $\sigma_x^2 + \sigma_y^2 + \sigma_z^2 = \sigma^2$.

Introducing the dimensionless variable $u = x/s$ where $s > 0$ is a scaling parameter here taken equal to $s^2 = \chi/\sigma$, the above equation reduces to

$$\frac{\partial^2 A_x}{\partial u^2} + 4\pi^2(s^2\sigma_x^2 - u^2)A_x = 0, \tag{3.93}$$

where $s^{-1/2}A_x(x/s) \equiv \check{A}_x(x)$. Equation 3.93 is precisely the same as equation 1.176 and thus has the following discrete set of solutions for $l = 0, 1, \ldots$:

$$\check{A}_x(x) = s^{-1/2}\psi_l(x/s), \qquad s^2\sigma_x^2 = (2l+1)/2\pi. \tag{3.94}$$

A similar discussion for $y$ leads to the following discrete set of solutions for $m = 0, 1, \ldots$:

$$\check{A}_y(y) = s^{-1/2}\psi_m(y/s), \qquad s^2\sigma_y^2 = (2m+1)/2\pi. \tag{3.95}$$

$\sigma_x$ and $\sigma_y$ can only assume the discrete values dictated by the above equations. For a given $l, m$, the value of $\sigma_z$ is given by

$$\sigma_z = \sqrt{\sigma^2 - \sigma_x^2 - \sigma_y^2} = \sqrt{\sigma^2 - \frac{l+m+1}{\pi s^2}}\,. \tag{3.96}$$

This can be expanded to first order as

$$\sigma_z \approx \sigma - \frac{l+m+1}{2\pi s^2 \sigma} = \sigma - \frac{l+m+1}{2\pi\chi}. \tag{3.97}$$

Thus, each eigensolution (or eigenmode) $s^{-1}\psi_l(x/s)\psi_m(y/s)$ propagates through the graded-index medium with a propagation constant $\sigma - (l+m+1)/2\pi\chi$. In the one-dimensional case, the corresponding result is

$$\sigma_z = \sqrt{\sigma^2 - \frac{l+1/2}{\pi s^2}} \approx \sigma - \frac{l+1/2}{2\pi\chi}. \tag{3.98}$$

The reader may also wish to note the similarity of the eigenmodes of quadratic graded-index media with Hermite-Gaussian beams in free space, with the identification of the beam size $W(z)$ with the scale parameter $s$ appearing here.

Now, let us assume that an arbitrary distribution of light $\check{f}(x, y)$ is incident on such a medium at $z = 0$. This distribution can be expanded in terms of the Hermite-Gaussian functions as

$$\check{f}(x, y, 0) \equiv \check{f}(x, y) = \sum_{l=0}^{\infty}\sum_{m=0}^{\infty} C_{lm}\,\frac{1}{s}\,\psi_l\left(\frac{x}{s}\right)\,\psi_m\left(\frac{y}{s}\right), \tag{3.99}$$

$$C_{lm} = \iint \frac{1}{s}\,\psi_l\left(\frac{x}{s}\right)\,\psi_m\left(\frac{y}{s}\right)\check{f}(x, y)\,dx\,dy.$$

and the amplitude distribution at any $z$ can be written as

$$\check{f}(x, y, z) \equiv \check{g}(x, y) = \sum_{l=0}^{\infty}\sum_{m=0}^{\infty} C_{lm}e^{i2\pi\sigma z}e^{-i(l+m+1)z/\chi}\,\frac{1}{s}\,\psi_l\left(\frac{x}{s}\right)\,\psi_m\left(\frac{y}{s}\right). \tag{3.100}$$

When $z = (\pi/2)\chi$ it is possible to show, using equation 1.184 and substituting for $C_{lm}$, that

$$\check{g}(x,y) = e^{i2\pi\sigma z}\,\frac{e^{-i\pi/2}}{s^2}\iint \check{f}(x',y')e^{-i2\pi(xx'+yy')/s^2}\,dx'\,dy', \tag{3.101}$$

which we recognize as being essentially a Fourier transform relation. Thus, propagation over a distance $z = (\pi/2)\chi$ in such a medium results in Fourier transformation. It also follows from the properties of repeated Fourier transformation that propagation over a distance $z = \pi\chi$ results in an inverted image, and that propagation over a distance $z = 2\pi\chi$ results in an erect image. Now, letting the parameter $a$ denote the *fractional order*, it is possible to show that when $z = a(\pi/2)\chi$ we obtain

$$\check{g}(x,y) = e^{i2\pi\sigma z}\,\frac{e^{-ia\pi/2}}{s^2}\iint K_a(x/s,x'/s)K_a(y/s,y'/s)\check{f}(x',y')\,dx'\,dy', \tag{3.102}$$

where $K_a(u,u')$ is the kernel of the fractional Fourier transform. We see that for arbitrary values of $z$, the effect of propagation in quadratic graded-index media can be interpreted as a fractional Fourier transform. An explicit integral transform relating $\check{g}(x,y)$ to $\check{f}(x,y)$ does appear in some treatments of graded-index media (Ghatak and Thyagarajan 1980, equation 3.25), although it has not been recognized as the fractional Fourier transform until Ozaktas and Mendlovic 1993a, b and Mendlovic and Ozaktas 1993a.

In the one-dimensional case, the corresponding input-output relation is

$$\check{g}(x) = \sum_{l=0}^{\infty} C_l e^{i2\pi\sigma z}e^{-i(l+1/2)z/\chi}\,\frac{1}{\sqrt{s}}\,\psi_l\left(\frac{x}{s}\right). \tag{3.103}$$

When $z = (\pi/2)\chi$,

$$\check{g}(x) = e^{i2\pi\sigma z}\,\frac{e^{-i\pi/4}}{s}\int \check{f}(x')e^{-i2\pi xx'/s^2}\,dx', \tag{3.104}$$

and for arbitrary $z = a(\pi/2)\chi$

$$\check{g}(x) = e^{i2\pi\sigma z}\,\frac{e^{-ia\pi/4}}{s}\int K_a(x/s,x'/s)\check{f}(x')\,dx'. \tag{3.105}$$

We now turn our attention to equations 3.70 and 3.71 which we specialize for graded-index media. Assuming the waist of the beam coincides with the input plane ($1/R_{\text{in}} = 0$) and borrowing the $\check{A},\check{B},\check{C},\check{D}$ parameters from chapter 4 (equation 4.42) we obtain

$$\frac{1}{R_{\text{out}}} = \frac{-\left(1-\frac{s^4}{W_{\text{in}}^4}\right)\frac{\lambda}{s^2}\sin(2d/\chi)}{\left(1-\frac{s^4}{W_{\text{in}}^4}\right)\cos(2d/\chi)+\left(1+\frac{s^4}{W_{\text{in}}^4}\right)}, \tag{3.106}$$

$$\frac{W_{\text{out}}^2}{W_{\text{in}}^2} = \frac{1}{2}\left(1-\frac{s^4}{W_{\text{in}}^4}\right)\cos(2d/\chi)+\frac{1}{2}\left(1+\frac{s^4}{W_{\text{in}}^4}\right). \tag{3.107}$$

We see that in general, the wavefront radius and beam size oscillate periodically with $d$. When the input beam size $W_{\text{in}}$ "matches" the natural scale parameter $s = \sqrt{\chi/\sigma}$ of the medium ($W_{\text{in}} = s$), we obtain $1/R_{\text{out}} = 0$ and $W_{\text{out}} = W_{\text{in}}$ for all $d$.

In the remainder of this subsection, which can be omitted without loss of continuity, we will discuss the number of degrees of freedom a graded-index medium of finite transverse extent $\Delta x$ can support (Ozaktas and Mendlovic 1993b). Our analysis implicitly assumed that the medium is of infinite transverse extent and that equation 3.89 holds for all $x, y$. Of course, this is physically not possible and would result in negative values for the refractive index in equation 3.89. We note that apart from this abstraction, our analysis is fairly exact and does not even employ the slowly varying envelope approximation leading to the paraxial Helmholtz equation, but rather employs the more general Helmholtz equation 3.91. The major assumption behind equation 3.91 (apart from ignoring the vector nature of light) is that the refractive index changes little over distances of the order of a wavelength. This is valid provided $\chi \gg \lambda$, which, as we will see below, must always be satisfied anyway.

The fact that all physical systems are of finite extent has several implications. First, since the amplitude distribution of light and its Fourier transform are both (approximately) confined to finite intervals, a finite number of samples (degrees of freedom) are sufficient to represent both. Second, Hermite-Gaussian functions beyond a certain order will not be relevant because their energy content will mostly lie outside the finite extent of the medium.

Since it is difficult to manufacture large index variations, and more fundamentally since it is necessary that $n(x, y) \geq 1$, we must ensure that $(x^2 + y^2) \ll \chi^2$ if such a system is to be realizable. Thus the extent of the medium must satisfy $\Delta x \ll \chi$. We will now show that if this condition is satisfied, the following consequences hold (Ozaktas and Mendlovic 1993b):

1. The one-dimensional space-bandwidth product, or the number of degrees of freedom the medium can support, is $\approx \Delta x^2/s^2$.

2. The number of Hermite-Gaussian modes whose energies lie predominantly within the medium is $\approx \Delta x^2/s^2$.

3. The first-order approximation for $\sigma_z$ (equation 3.97) is valid for these modes.

To show the first of the above, we note that both the original amplitude distribution of light and its Fourier transform will be confined to an extent of $\Delta x$. This spatial extent corresponds to an interval of length $\Delta x/s^2$ in units of spatial frequency (equation 3.101). Since $\Delta x/s^2$ represents the double-sided spatial bandwidth of the amplitude distribution, it follows that the Nyquist sampling interval is $s^2/\Delta x$. Therefore, $\Delta x/(s^2/\Delta x) = \Delta x^2/s^2$ samples are sufficient to fully represent the amplitude distribution of light propagating through the medium. This quantity is the space-bandwidth product of the medium or the number of degrees of freedom the medium can support. The two-dimensional space-bandwidth product is $(\Delta x^2/s^2)^2$.

To show the second claim, we refer back to page 36 where we had stated that most of the energy of the $l$th Hermite-Gaussian function is concentrated in the interval $[-\sqrt{(l+1/2)/\pi}\,, \sqrt{(l+1/2)/\pi}\,]$.

Examination of higher-order Hermite-Gaussian functions further reveals that most of the energy within this interval tends to be concentrated close to the end points of the interval, rather than around the origin. Thus, Hermite-Gaussian functions $\psi_l(x/s)$ whose energies mostly lie within the interval $[-\Delta x/2, \Delta x/2]$ are those which satisfy $\sqrt{(l+1/2)/\pi} < \Delta x/2s$, or whose orders $l$ are (approximately) less than $\Delta x^2/s^2$. In other words, only these first $\Delta x^2/s^2$ modes are relevant and the summations can be truncated after this mode. In two dimensions the total number of relevant modes are $(\Delta x^2/s^2)^2$.

It is satisfying that the number of relevant modes is equal to the number of spatial degrees of freedom. Thus regardless of whether we prefer to represent the amplitude distribution of light in terms of its samples or in terms of the coefficients $C_{lm}$ of its Hermite-Gaussian expansion, we need the same number of samples or coefficients.

Since the relevant modes have orders $l, m < \Delta x^2/s^2$, it follows that the first-order approximation of $\sigma_z$ given by equation 3.97 is always accurate. The first-order expansion holds when $(l+m+1)/(\pi s^2\sigma^2) \ll 1$. For the relevant modes $(l+m+1)/(\pi s^2\sigma^2) < (\Delta x^2/s^2)/(\pi s^2\sigma^2) = \Delta x^2/\pi\chi^2$. Since we had assumed $\Delta x \ll \chi$ for a realizable medium, it is ensured that this quantity is always $\ll 1$.

That the number of relevant modes should be equal to the space-bandwidth product can also be seen as follows: If the space-bandwidth product is $N$, we can sample both sides of the expansion given in equation 3.99 at $N$ points without loss of information. This gives us $N$ linear equations in the unknown expansion coefficients $C_{lm}$. If the number of modes included in the expansion is less than $N$, the set of equations will be overdetermined, meaning that this number of modes is not sufficient to match the original function at the sample points. If the number of modes is more than $N$, the set of equations will be underdetermined, meaning that there are redundant modes. When the number of modes is equal to $N$, the expansion coefficients can be uniquely solved for to match the original function at the sample points.

An interesting and important interpretation of a quadratic graded-index medium is as the limit of a large number of positive lenses interspersed between short sections of free space. As the number of lenses becomes larger and the sections of free space become shorter, such a lens system approaches a quadratic graded-index medium.

A brief overview of graded-index media in a Fourier optics context with useful references is Gómez-Reino, Bao, and Pérez 1996.

### 3.4.4 Extensions

It is not difficult to also allow for homogeneous regions with refractive indices other than unity, spherical refracting surfaces between such homogeneous regions, and spherical mirrors (Saleh and Teich 1991). We will not explicitly discuss such components, since they can be handled by simple tricks. Homogeneous regions with refractive index $n \neq 1$ can be handled by working with normalized angles, spherical refractive surfaces are treated like lenses by inserting an infinitesimal section of free space on both sides, and spherical

mirrors can be handled by folding the optical axis.

Sections of free space, thin lenses, and quadratic graded-index media belong to the class of *quadratic-phase systems*, which will be further discussed in chapter 4. An example of a common component which is not in this category is a prism (Saleh and Teich 1991). The effect of a prism is similar to the effect of tilting the optical axis and can be analyzed in the same manner. In certain instances, one may also want to be able to deal with transverse displacements of the optical axis. Prisms, as well as tilts and displacements of the optical axis will not be considered in this book.

### 3.4.5 Spatial filters

Spatial filters are optical components whose output $\check{g}(x,y)$ is equal to their input $\check{f}(x,y)$ multiplied by a complex transmittance function $\check{h}(x,y)$:

$$\check{g}(x,y) = \check{h}(x,y)\check{f}(x,y). \tag{3.108}$$

If $\check{h}$ is real and positive, the filter is referred to as a magnitude-only filter. If $|\check{h}| = 1$, the filter is referred to as a phase-only filter. Unless the material exhibits gain the filter function satisfies $|\check{h}| \leq 1$.

Such filters can be realized by thin transmissive elements whose refractive index, attenuation coefficient, or thickness is a function of $(x,y)$:

*Thin plate with variable thickness:* A thin plate of homogeneous refractive index $n_{\mathrm{pl}}$ and variable thickness $d(x,y)$ in a medium with refractive index $n$ will exhibit complex transmittance (Saleh and Teich 1991)

$$\check{h}(x,y) = e^{i2\pi\sigma d_0} \exp[i2\pi\sigma(n_{\mathrm{pl}}/n - 1)d(x,y)], \tag{3.109}$$

where $\sigma = nf_{oc}/c$. This transmittance is defined between two planes separated by a distance $d_0$ between which the variable thickness plate is completely contained. This formula is valid in the paraxial approximation and when the thickness $d_0$ is sufficiently small (Saleh and Teich 1991). The constant factor $\exp(i2\pi\sigma d_0)$ is often dropped. The transmittance function for a thin lens is a special case of the above formula.

*Thin plate with graded index:* Now, let us consider a thin slab of uniform thickness $d_0$ but variable refractive index $n_0[1 + \rho_{\mathrm{pl}}(x,y)]$. In this case the transmittance functions is (Saleh and Teich 1991)

$$\check{h}(x,y) = e^{i2\pi\sigma d_0} \exp[i2\pi\sigma\rho_{\mathrm{pl}}(x,y)d_0], \tag{3.110}$$

where $\sigma = n_0 f_{oc}/c$. When $\rho_{\mathrm{pl}}(x,y) = -(x^2+y^2)/2\chi^2$, the plate behaves like a thin lens. In this case, the plate can also be interpreted as a thin section of quadratic graded-index media (since $n_0[1 + \rho_{\mathrm{pl}}(x,y)]$ is an approximation to equation 3.89).

Amplitude filters can be realized by using thin plates with variable attenuation coefficients. It is also possible to have spatial filters which simultaneously have variable

attenuation, thickness, and refractive index, resulting in more general complex spatial filters. In this book, we will assume that the complex transmittance function is specified and not involve ourselves with its physical realization.

*Apertures* are spatial filters whose transmittance function $\check{h}(x, y)$ takes only two values: 1 or 0. They can be physically realized simply by cutting out the desired shape in an opaque (non-transparent) material. It is assumed that the distribution of light behind the opaque regions is zero, and that the distribution of light behind the transparent (cut out) regions is equal to the incident distribution of light. Although not strictly true, this is usually a sufficiently accurate approximation, known as the Kirchhoff approximation (Goodman 1996, Lohmann 1986).

### 3.4.6 Fourier-domain spatial filters

Fourier-domain filters are optical subsystems whose output $\check{g}(x, y)$ is equal to its input $\check{f}(x, y)$ convolved by $\check{h}(x, y)$:

$$\check{g}(x, y) = \check{h}(x, y) * *\check{f}(x, y), \tag{3.111}$$

$$\check{G}(\sigma_x, \sigma_y) = \check{H}(\sigma_x, \sigma_y)\check{F}(\sigma_x, \sigma_y). \tag{3.112}$$

If $\check{H}$ is real and positive, the filter is referred to as a magnitude-only filter. If $|\check{H}| = 1$, the filter is referred to as a phase-only filter. Unless the material exhibits gain the filter function satisfies $|\check{H}| \leq 1$. Such a filter can be realized by sandwiching a thin spatial filter between a forward and inverse Fourier transform stage (figure 3.4).

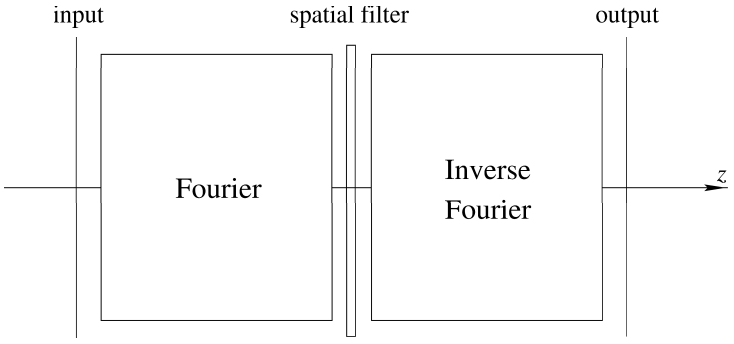


Figure 3.4: A Fourier-domain filtering system consists of a Fourier transform followed by a spatial filter followed by an inverse Fourier transform.

*Fourier transform stages:* Although the Fourier transform can be realized optically in many different ways, the most common configurations are (i) a lens of focal length $f$ sandwiched between two sections of free space of length $d = f$, (ii) a section of free space of length $d$ sandwiched between two lenses of focal length $f = d$, and (iii) a section of quadratic graded-index media of length $(\pi/2)\chi$. Ignoring uninteresting constant phase

factors which do not depend on $(x, y)$, all of these transform an input $\check{f}(x, y)$ into the Fourier transform $\check{F}(\sigma_x, \sigma_y)$ with $\sigma_x = x/s^2, \sigma_y = y/s^2$:

$$\check{F}(x/s^2, y/s^2) \propto \iint \check{f}(x', y') e^{-i2\pi(xx'+yy')/s^2}\, dx'\, dy', \tag{3.113}$$

where $s$ is given by $\sqrt{\lambda d} = \sqrt{\lambda f}$ for (i) and (ii), and $s = \sqrt{\chi/\sigma}$ for (iii). The first two cases can be easily demonstrated by using Fresnel's integral and the transmittance function of a thin lens (Goodman 1996). The third case is demonstrated by equation 3.101. It will be very easy to verify these results using matrix algebra after the matrix representation of such systems are introduced in chapter 4. It is also possible to rewrite equation 3.113 in the purer form

$$F(x/s, y/s) \propto \iint f(x'/s, y'/s)\ \exp\left[-i2\pi\left(\frac{x}{s}\frac{x'}{s} + \frac{y}{s}\frac{y'}{s}\right)\right]\ \frac{dx'}{s}\,\frac{dy'}{s}, \tag{3.114}$$

where $sF(s\sigma_x, s\sigma_y) = \check{F}(\sigma_x, \sigma_y)$ and $s^{-1}f(x'/s, y'/s) = \check{f}(x', y')$. This form can be directly translated to the dimensionless Fourier transform relation $F(\mu, \nu) = \int\int f(u, v) \exp[-i2\pi(\mu u + \nu v)]\, du\, dv$ with $u = x'/s$, $\mu = x/s$ and similarly for $v$ and $\nu$.

Inverse Fourier transforms can be easily obtained by noting that the inverse Fourier transform is simply the forward Fourier transform followed by flipping the coordinate axes. Thus if we choose the output coordinates to be opposite in direction to the input coordinates, the same configuration will give us the inverse transform.

Using the Fourier and inverse Fourier transform as building blocks, it is possible to realize the desired Fourier-domain filtering system by employing a spatial filter of the form

$$\check{H}(x/s^2, y/s^2) \tag{3.115}$$

between the Fourier and inverse Fourier blocks in figure 3.4. When this system is realized by using type (i) Fourier transform stages, it is commonly known as a "$4f$" system. In dimensionless notation $H(s\sigma_x, s\sigma_y) = \check{H}(\sigma_x, \sigma_y)$, the filter is given by $H(x/s, y/s)$.

### 3.4.7 General linear systems

It is also possible to optically realize arbitrary linear systems of the form

$$\check{g}(x, y) = \iint \check{h}(x, y; x', y') \check{f}(x', y')\, dx'\, dy', \tag{3.116}$$

by using systems such as matrix-vector product architectures (Goodman 1996) or a class of systems which may be collectively referred to as multi-facet architectures (Mendlovic and Ozaktas 1993c, Ozaktas, Brenner, and Lohmann 1993, Ozaktas and Mendlovic 1993c). However, the realization of such systems are usually considered inefficient or costly since optical components with space-bandwidth product $O(N^2)$ are needed to realize systems for signals with space-bandwidth product $N$.

### 3.4.8 Spherical reference surfaces

Although not a component in a physical sense, here we also discuss spherical reference surfaces for future reference. Usually we specify the amplitude distribution of light on a planar reference surface. Sometimes however, it is more convenient to specify the amplitude distribution on a spherical reference surface of radius $R$. Referring to figure 3.5, the

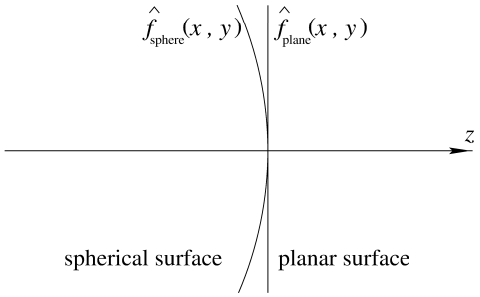


Figure 3.5: Spherical and planar reference surfaces. $R > 0$ as drawn.

relation between the distributions with respect to the planar and spherical surfaces is

$$\check{f}_{\text{plane}}(x,y) = \check{f}_{\text{sphere}}(x,y)\exp[i\pi(x^2+y^2)/\lambda R]. \tag{3.117}$$

In the spherical case, the coordinates $x, y$ are those perpendicularly dropping from the sphere onto the plane. The above relation is valid for small curvatures, so that the surface is "thin" in the same sense that a thin lens is thin. The reader can verify this relation easily by considering the expressions for plane waves and diverging/converging spherical waves.

### 3.4.9 Remarks

We have ignored attenuations and reflections from the components discussed, concentrating mostly on how they modify the phase of the incident optical wave. Furthermore, we have focused our attention to quadratic-phase systems for which higher-order dependences of the phase on $x, y$ can be neglected. (Linear terms—corresponding to shifts and tilts—are also not included for simplicity, though most results can be easily generalized to include them as well.) We will see in chapter 4 that quadratic-phase systems mathematically correspond to linear canonical transforms, and play a central role in the first-order study of optical systems.

## 3.5 Geometrical optics

In geometrical optics, light is represented by light *rays*, which are in general curvilinear paths along which light energy travels. A distribution of light can be represented by a

bundle of rays whose trajectories (and sometimes intensities) are specified. Since we are considering centered optical systems, we will usually represent the distribution of light over a given plane $z$ = constant, by specifying the points $(x, y)$ and angles $(\theta_x, \theta_y)$ at which the rays intersect this plane. $\theta_x$ is the angle the ray makes with the $y$-$z$ plane, and $\theta_y$ is the angle the ray makes with the $x$-$z$ plane. In other words, $\theta_x$ is the complement to $\pi/2$ of the angle the ray makes with the $x$ axis, and $\theta_y$ is the complement to $\pi/2$ of the angle the ray makes with the $y$ axis. The signs of the angles are the same as the signs of the slopes of the rays. In order to make the correspondence with wave optics more transparent, we will often choose to work with $\sigma_x \equiv \sin\theta_x/\lambda$ and $\sigma_y \equiv \sin\theta_y/\lambda$ instead of the angles themselves, where $\lambda$ is the wavelength of light in the medium of propagation. ¿From a purely geometrical optical perspective, $\sigma_x$ and $\sigma_y$ may be interpreted merely as normalized angles. However, it is worth keeping in mind that $\sigma_x, \sigma_y$ can also be interpreted as the transverse spatial frequencies of a plane wave propagating in the direction of the ray (figure 3.3 and equations 3.30 and 3.31).

Optical components are characterized by how they map a ray incident on their input plane to a ray exiting from their output plane. We will again restrict our attention to first-order centered systems for which (i) the paraxial (small-angle) approximation can be employed ($\theta_x, \theta_y \ll 1$), and (ii) in which any lenses are thin lenses. In the paraxial approximation, angles are taken equal to their sines/tangents and the slopes of the rays so that $\sigma_x \approx \theta_x/\lambda$, $\sigma_y \approx \theta_y/\lambda$. In a first-order system, the parameters $(x, y)$ and $(\theta_x, \theta_y)$ characterizing a ray at the output of a system are related to the corresponding parameters characterizing the ray at the input through a linear relation; higher-order dependences are neglected. These approximations are in direct correspondence with those made in section 3.3, which essentially amounted to neglecting terms beyond the quadratic *in the phase* of integral kernels.

The sections on geometrical optics proceed more or less in parallel with the sections on wave optics.

### 3.5.1 The ray equation

The ray equation governs the trajectory of rays $\mathbf{r}(s) \equiv (x(s), y(s), z(s))$ traveling in an inhomogeneous medium (Saleh and Teich 1991):

$$\frac{d}{ds}\left(n\frac{d\mathbf{r}}{ds}\right) = \nabla n, \tag{3.118}$$

where $\nabla n$ is the gradient of the refractive index distribution $n(\mathbf{r})$. Each value of the parameter $s$ corresponds to a point along the ray. Solution of the above equation subject to specified boundary conditions gives a set of trajectories representing a bundle of rays. The boundary conditions may take the form of the positions and directions of a bundle of rays crossing a given plane. (The parameter $s$ used here in accordance with widespread convention should not be confused with the scale parameter introduced in section 3.2.)

In a homogeneous medium, it is easy to show that solutions of the ray equation take the form of linear trajectories; the paths of light are straight lines. Two commonly encountered bundles are those that correspond to a plane wave and a spherical wave. In the former, the bundle consists of a set of rays which are all parallel to each other. In the latter, the rays emanate from a common origin and diverge outwards in all directions. More general bundles of rays correspond to more general waves.

If we know the angles $(\theta_x, \theta_y)$ that the bundle of rays corresponding to a plane wave makes with a particular plane $z = z_1$, then we know that the ray crossing this plane at $(x, y)$ will cross a second plane $z = z_2 = z_1 + d$ at the point $(x + \theta_x d, y + \theta_y d)$, still making the same angles. We have assumed small angles so that $\tan\theta_x \approx \sin\theta_x \approx \theta_x$, and likewise for $\theta_y$.

Turning our attention to the bundle of rays corresponding to a spherical wave originating from the origin $(0, 0, 0)$, it is possible to show that the ray crossing a plane $z =$ constant, at the point $(x, y)$, will be making an angle of $(\theta_x, \theta_y) = (x/z, y/z)$ with that plane (again in the paraxial approximation).

Now, let us consider two homogeneous half spaces which meet at a plane $z =$ constant, such that the medium to the left has a refractive index $n_{\text{left}}$ and the medium to the right has an index $n_{\text{right}}$. Then, the following relations hold for the angles characterizing the ray coming from the left and the ray leaving towards the right:

$$\begin{aligned} n_{\text{left}} \sin\theta_{x\text{left}} &= n_{\text{right}} \sin\theta_{x\text{right}}, \\ n_{\text{left}} \sin\theta_{y_{\text{left}}} &= n_{\text{right}} \sin\theta_{y_{\text{right}}}. \end{aligned} \tag{3.119}$$

These relationships, known as Snell's law, cover both the condition that the incoming and outgoing rays and the normal to the surface all lie in the same plane, and the condition relating the angles made with the normal. Snell's law takes a particularly simple form when stated in terms of the normalized angles $\sigma_x, \sigma_y$:

$$\begin{aligned} \sigma_{x\text{left}} &= \sigma_{x\text{right}}, \\ \sigma_{y_{\text{left}}} &= \sigma_{y_{\text{right}}}, \end{aligned} \tag{3.120}$$

which simply states that $(\sigma_x, \sigma_y)$ is conserved at the boundary. In the paraxial approximation we obtain

$$\begin{aligned} n_{\text{left}}\theta_{x\text{left}} &= n_{\text{right}}\theta_{x\text{right}}, \\ n_{\text{left}}\theta_{y_{\text{left}}} &= n_{\text{right}}\theta_{y_{\text{right}}}. \end{aligned} \tag{3.121}$$

Although the law of refraction at refractive index discontinuities is strongly associated with the name of Snell (and sometimes Descartes), the law appeared in the work of Ibn Sahl some 650 years before Snell. For this reason, it has been suggested that it be referred to as the Ibn Sahl law (Wolf and Krötzsch 1995).

### 3.5.2 Fermat's principle and the eikonal equation

An important concept in geometrical optics is the *optical path length* along a ray. The optical path length from point A to point B is defined as the line integral

$$\int_A^B n(\mathbf{r})\, ds, \tag{3.122}$$

where $ds$ is the differential element along the path of integration (Saleh and Teich 1991). *Fermat's principle* is a statement of the laws of geometrical optics. According to this principle, among all the possible paths connecting the points A and B, only the paths whose optical path length variations with respect to small deviations are zero, correspond to actual light rays. With "small deviations" we mean perturbations of the path in question such that the perturbed path still lies in a narrow tubular neighborhood of the path. The "optical path length variation" is simply the difference between the optical path length of the perturbed path and the original path. Such paths whose variations are zero are known to be either paths with minimum or maximum optical path length, or paths for which the optical path length exhibits an "inflection." Since a minimum is the most commonly encountered case and since the optical path length is proportional to the time it takes light to travel along the path, this principle is also known as the *principle of least time* (Saleh and Teich 1991). Sometimes there are several paths for which the variations are zero. An important special case is that of point imaging when a whole bundle of rays emanating from point A all arrive at point B, in which case all the paths have the same optical path length.

The ray equation can be derived from Fermat's principle by using the calculus of variations (Marcuse 1982, Born and Wolf 1980). (The ray equation is the so-called *Euler equation* of the variational problem.) This variational principle can also be interpreted in terms of wave optical concepts, providing significant insight on the relationship between rays and waves. The path which is the actual optical ray corresponds to the path along which the wave contributions constructively add up, whereas along other paths which are not actual rays the contributions destructively interfere. Constructive interference occurs where the phase varies slowly, particularly where the variation of the phase is zero. Since the optical path length is associated with the phase, this directly corresponds to the variation of the optical path length being zero. These notions find their mathematical expression in the stationary-phase integral (subsection 1.10.3). An excellent discussion of these issues may be found in Lohmann 1986. Further discussion of these relationships and Fermat's principle is beyond the scope of this book. For this we refer the reader to Lohmann 1986, Marcuse 1982, and Born and Wolf 1980.

Another important concept is the *eikonal*. The eikonal $\check{S}(\mathbf{r})$ is a function of position such that (i) its equilevel surfaces are everywhere orthogonal to the optical rays, and (ii) the optical path lengths along all rays from one equilevel surface to another are equal. The rays lie along the gradient of $\check{S}(\mathbf{r})$. An alternative statement of the laws of geometrical

optics is the eikonal equation (Saleh and Teich 1991):

$$|\nabla \check{S}(\mathbf{r})|^2 = n^2(\mathbf{r}). \tag{3.123}$$

The optical path length along a ray between points A and B is simply equal to the difference between the value of the eikonal at these points (Saleh and Teich 1991):

$$\int_A^B n(\mathbf{r})\, ds = \int_A^B |\nabla \check{S}(\mathbf{r})|\, ds = \check{S}(\mathrm{B}) - \check{S}(\mathrm{A}). \tag{3.124}$$

The eikonal equation is equivalent to Fermat's principle and the ray equation can be derived from the eikonal equation as well (Marcuse 1982, Born and Wolf 1980).

The equilevel surfaces of the eikonal are sometimes referred to as *geometrical wavefronts*. They are often close approximations to the physical wavefronts, but deviate to a greater extent in those regions where geometrical optics does not provide a satisfactory description of the behavior of light (such as in those regions where light is tightly focused).

It is possible to substitute an expression of the form $\check{f}(\mathbf{r}) = \check{A}(\mathbf{r}) \exp[i2\pi(f_{\mathrm{oc}}/c)\check{S}(\mathbf{r})]$ in the Helmholtz equation 3.23 and show that in the limit $f_{\mathrm{oc}} \to \infty$ one obtains the eikonal equation. This supports the association between the phase and the optical path length. This classic derivation, which the reader may find in many texts such as Saleh and Teich 1991 and Marcuse 1982, is often used to motivate the fact that geometrical optics is a limiting case of wave optics, which holds when the wavelength is small.

### 3.5.3 Hamilton's equations

We now briefly discuss the Hamiltonian formulation of geometrical optics (Marcuse 1982). We first define the Hamiltonian $\check{H}$ of a system with refractive index distribution $n(x, y, z)$ as

$$\check{H}(x, y, \sigma_x, \sigma_y; z) \equiv -\sqrt{n^2(x, y, z) f_{\mathrm{oc}}^2/c^2 - \sigma_x^2 - \sigma_y^2}\,. \tag{3.125}$$

As before, $\sigma_x \equiv \sin\theta_x/\lambda$ and $\sigma_y \equiv \sin\theta_y/\lambda$, where $\lambda = c/n(x, y, z) f_{\mathrm{oc}}$. We can now write the celebrated Hamilton's equations:

$$\frac{dx}{dz} = \frac{\partial \check{H}}{\partial \sigma_x}, \qquad \frac{dy}{dz} = \frac{\partial \check{H}}{\partial \sigma_y}, \tag{3.126}$$

$$\frac{d\sigma_x}{dz} = -\frac{\partial \check{H}}{\partial x}, \qquad \frac{d\sigma_y}{dz} = -\frac{\partial \check{H}}{\partial y}. \tag{3.127}$$

It is possible to show that Hamilton's equations are equivalent to the ray equation (Marcuse 1982, Goldstein 1980), so that they also constitute an alternative statement of the laws of geometrical optics. Here $x, y, \sigma_x, \sigma_y$ are treated as functions of $z$ and give us the position and angles of a ray propagating through the system in the positive $z$ direction. Noting that $dx/ds = \sin\theta_x$ and $dy/ds = \sin\theta_y$, where $ds = \sqrt{dx^2 + dy^2 + dz^2}$ is the differential

element along the ray, it is possible to write $\sigma_x$ and $\sigma_y$ in terms of the derivatives of $x$ and $y$ with respect to $z$ as follows:

$$\begin{aligned}
\sigma_x = \frac{\sin\theta_x}{\lambda} &= \frac{f_{\rm oc}}{c} n(x,y,z)\sin\theta_x = \frac{f_{\rm oc} n(x,y,z)}{c}\frac{dx}{ds} \\
&= \frac{f_{\rm oc}}{c}\frac{n(x,y,z)dx/dz}{\sqrt{1+(dx/dz)^2+(dy/dz)^2}}, \\
\sigma_y = \frac{\sin\theta_y}{\lambda} &= \frac{f_{\rm oc}}{c} n(x,y,z)\sin\theta_y = \frac{f_{\rm oc} n(x,y,z)}{c}\frac{dy}{ds} \\
&= \frac{f_{\rm oc}}{c}\frac{n(x,y,z)dy/dz}{\sqrt{1+(dx/dz)^2+(dy/dz)^2}},
\end{aligned} \tag{3.128}$$

These equations allow us to express the Hamiltonian in terms of $x, y$ and their derivatives rather than $x, y$ and $\sigma_x, \sigma_y$, a form which is sometimes preferred.

The total derivative of the Hamiltonian with respect to $z$ can be written as

$$\frac{d\check{H}}{dz} = \frac{\partial\check{H}}{\partial x}\frac{dx}{dz} + \frac{\partial\check{H}}{\partial y}\frac{dy}{dz} + \frac{\partial\check{H}}{\partial\sigma_x}\frac{d\sigma_x}{dz} + \frac{\partial\check{H}}{\partial\sigma_y}\frac{d\sigma_y}{dz} + \frac{\partial\check{H}}{\partial z}. \tag{3.129}$$

By using Hamilton's equations, it immediately follows that the total derivative $d\check{H}/dz$ is equal to $\partial\check{H}/\partial z$ since the other terms cancel out. If $n(x,y,z)$ does not depend on $z$, then $d\check{H}/dz = \partial\check{H}/\partial z = 0$ which means that the Hamiltonian remains the same along a ray $(x(z), y(z), \sigma_x(z), \sigma_y(z))$.

The results obtained by solving Hamilton's equations will be the same as those obtained by solving the ray equation. As instructive as it is, the analogy between the Hamiltonian formulation in optics and the Hamiltonian formulation in mechanics will not be pursued in this book. We refer interested readers to Goldstein 1980, Sekiguchi and Wolf 1987, and the other advanced books on mathematical optics that are referred to at the end of chapter 4. However, we note that point particles are to the wavefunctions of quantum mechanics, precisely what rays are to optical waves. Also, the important Hamilton-Jacobi partial differential equation associated with the Hamiltonian given in equation 3.125 is nothing but the eikonal equation 3.123 (Marcuse 1982).

In the paraxial approximation, we take $\sin\theta_x = \theta_x$, $\sin\theta_y = \theta_y$ and $dx/ds = dx/dz$, $dy/ds = dy/dz$ so that

$$\begin{aligned}
\sigma_x = \frac{\theta_x}{\lambda} = \frac{f_{\rm oc}}{c} n(x,y,z)\theta_x = \frac{f_{\rm oc}}{c} n(x,y,z)\frac{dx}{dz}, \\
\sigma_y = \frac{\theta_y}{\lambda} = \frac{f_{\rm oc}}{c} n(x,y,z)\theta_y = \frac{f_{\rm oc}}{c} n(x,y,z)\frac{dy}{dz}.
\end{aligned} \tag{3.130}$$

In this case the Hamiltonian simplifies to

$$\check{H}(x,y,\sigma_x,\sigma_y;z) = \frac{(\sigma_x^2+\sigma_y^2)}{2nf_{\rm oc}/c} - nf_{\rm oc}/c \approx \frac{(\sigma_x^2+\sigma_y^2)}{2n_0 f_{\rm oc}/c} + \Delta n f_{\rm oc}/c - n_0 f_{\rm oc}/c, \tag{3.131}$$

where we let $n = n_0 - \Delta n$ with the assumption that $\Delta n \ll n_0$ and replaced $n \approx n_0$ in the denominator (Marcuse 1982). The constant term $-n_0 f_{\rm oc}/c$ may be dropped since it will disappear in Hamilton's equations anyway.

Although we have not provided a derivation of Hamilton's equations from the ray equation, we mention a crude but simple and instructive derivation which allows us to see their consistency in the paraxial case. Referring ahead to equation 3.136 and using equation 3.130, we can obtain $d\sigma_x/dz \approx (f_{\rm oc}/c)\partial n/\partial x$, which in turn is equal to $-\partial \check{H}/\partial x$ evaluated from equation 3.131. The other equation $\partial \check{H}/\partial \sigma_x = dx/dz$ is also easily seen to carry the same information as equation 3.130. (The equations for $y$ are of course identical.)

Finally, we show that in their paraxial form, Hamilton's equations have a particularly simple interpretation in terms of Snell's law. (The rest of this subsection can be omitted if desired.) First, consider the equation $dx/dz = \partial \check{H}/\partial \sigma_x$. The right hand side is readily evaluated from the paraxial Hamiltonian as $\sigma_x/(n f_{\rm oc}/c) = \sigma_x \lambda = \theta_x$ so that we are left with $dx/dz = \theta_x$: the rate of change of $x$ is simply given by $\theta_x$ which in the paraxial approximation also equals the slope of the ray. Second, let us consider the equation $d\sigma_x/dz = -\partial \check{H}/\partial x$. We will show that this equation corresponds to Snell's law. The right hand side evaluates to $(f_{\rm oc}/c)\partial n/\partial x = -(f_{\rm oc}/c)\partial \Delta n/\partial x$. Thus the incremental change in $\sigma_x$ can be written as

$$\sigma_x(z + dz) = \sigma_x(z) + (f_{\rm oc}/c)\frac{\partial n(x,z)}{\partial x}\, dz. \tag{3.132}$$

Using $\sigma_x = \theta_x/\lambda \approx n_0 \theta_x (f_{\rm oc}/c)$ to change to angles, and noting that the axial and transverse increments in the ray position $dz$ and $dx$ are related by $dx = \theta_x dz$, we get an angular deflection given by

$$d\theta_x \equiv \theta_x(x + dx) - \theta_x(x) = \frac{1}{\theta_x}\frac{dn}{n_0}, \tag{3.133}$$

where $dn = (\partial n/\partial x)dx$. Now, let us write Snell's law for an interface parallel to the optical axis such that the refractive index on one side is $n$ and that on the other side is $n + dn$:

$$\sin[\phi_x(x + dx)]\,(n + dn) = \sin[\phi_x(x)]\, n, \tag{3.134}$$

where the angle $\phi_x(x)$ complements the angle $\theta_x$ to $\pi/2$ and is thus a large angle (figure 3.6). Writing $\phi_x(x + dx) = \phi_x(x) + d\phi_x$ and expanding the sine of a sum, it follows that equation 3.134 implies an angular deflection of

$$d\phi_x \equiv \phi_x(x + dx) - \phi_x(x) = -\tan[\phi_x(x)]\,\frac{dn}{n} \approx -\tan[\phi_x(x)]\,\frac{dn}{n_0}. \tag{3.135}$$

Since the angular deflections $d\theta_x$ and $-d\phi_x$ are equal, and since $\tan\phi_x = \cot\theta_x = 1/\tan\theta_x \approx 1/\theta_x$, it follows that the second Hamilton's equation corresponds to Snell's law. We also note that the derivative of the refractive index with respect to $z$ does not come into play since for paraxial angles, the deflective force of changes with respect to $x$ is dominant.

Chapter 3 of Marcuse 1982 is particularly recommended for a very accessible discussion of these and related topics. Further references are given at the end of chapter 4.

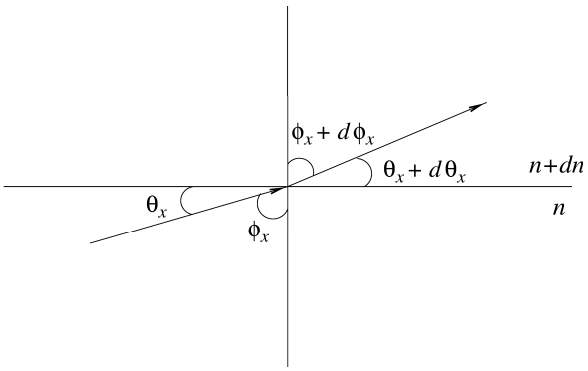


Figure 3.6: Deflection of a ray by an infinitesimal refractive index change.

## 3.6 Geometrical optical characterization of optical components

### 3.6.1 Sections of free space

If we restrict ourselves to paraxial rays which make small angles with the $z$ axis, then $ds \approx dz$ and we can write the ray equation as (Saleh and Teich 1991)

$$\frac{d}{dz}\left(n\frac{dx}{dz}\right) \approx \frac{\partial n}{\partial x}, \qquad \frac{d}{dz}\left(n\frac{dy}{dz}\right) \approx \frac{\partial n}{\partial y}. \tag{3.136}$$

In a homogeneous medium where $n$ is constant we have

$$\frac{d^2x}{dz^2} = 0, \qquad \frac{d^2y}{dz^2} = 0, \tag{3.137}$$

which imply that $x(z)$ and $y(z)$ both increase linearly with $z$, corresponding to the fact that the rays are straight lines. If a ray intercepts the plane $z = z_1$ at $(x_1, y_1)$ making angles $(\theta_{x1}, \theta_{y_1})$ with the $y$-$z$ and $x$-$z$ planes respectively, then this ray will intercept the plane $z = z_2 = z_1 + d$ at $(x_2, y_2)$ making angles $(\theta_{x2}, \theta_{y_2})$, where

$$x_2 = x_1 + \theta_{x1}d, \qquad y_2 = y_1 + \theta_{y_1}d, \tag{3.138}$$

$$\theta_{x2} = \theta_{x1}, \qquad \theta_{y_2} = \theta_{y_1}. \tag{3.139}$$

These equations can also be easily derived from the paraxial form of Hamilton's equations.

### 3.6.2 Thin lenses

Thin spherical lenses are characterized by their focal length $f$ (equation 3.85), which is the distance from the lens along the optical axis at which an incident bundle of parallel rays intersect at a point (figure 3.7). The thickness of a thin lens is assumed to be sufficiently

small so that the point $(x_2, y_2)$ from which a ray leaves the lens is equal to the point $(x_1, y_1)$ at which the ray is incident on the lens. Employing Snell's law twice, it is possible to show—under the paraxial approximation—that for a thin lens the angles $(\theta_{x2}, \theta_{y_2})$ of a ray leaving the lens are related to the angles $(\theta_{x1}, \theta_{y_1})$ of the ray incident on the lens through relations which depend only on the focal length of the lens (and not on the separate surface curvatures). Taken together, these results provide the geometrical optical

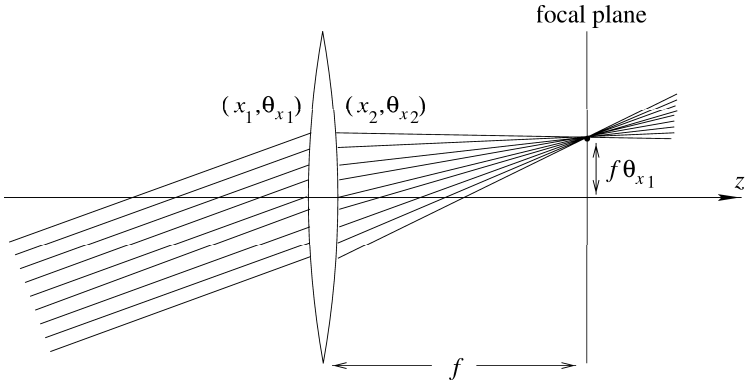


Figure 3.7: In the paraxial approximation, parallel rays incident on a thin lens intersect at a single point lying on the focal plane.

characterization of a thin lens in the paraxial approximation:

$$x_2 = x_1, \qquad y_2 = y_1, \tag{3.140}$$

$$\theta_{x2} = \theta_{x1} - \frac{x_1}{f}, \qquad \theta_{y_2} = \theta_{y_1} - \frac{y_1}{f}. \tag{3.141}$$

### 3.6.3 Quadratic graded-index media

The refractive index distribution of a quadratic graded-index medium was given in equation 3.89. Assuming $\chi_x = \chi_y = \chi$, substituting in equation 3.136, and noting that the refractive index does not depend on $z$ we obtain

$$\frac{d^2x}{dz^2} = \frac{1}{n^2}\frac{-xn_0^2}{\chi^2}, \qquad \frac{d^2y}{dz^2} = \frac{1}{n^2}\frac{-yn_0^2}{\chi^2}. \tag{3.142}$$

Usually $\chi$ is large so that the refractive index distribution $n(x, y)$ appearing in the denominators can be replaced by $n_0$ within the extent of the medium (Saleh and Teich 1991), leading to

$$\frac{d^2x}{dz^2} = -\frac{x}{\chi^2}, \qquad \frac{d^2y}{dz^2} = -\frac{y}{\chi^2}. \tag{3.143}$$

These equations can also be derived from the paraxial form of Hamilton's equations with the Hamiltonian given in equation 3.131. For instance, we can use $dx/dz = \partial\check{H}/\partial\sigma_x =$

$c\sigma_x/n_0 f_{\rm oc}$ and $d\sigma_x/dz = -\partial\check{H}/\partial x = (f_{\rm oc}/c)\partial n/\partial x$, leading to $n_0 d^2x/dz^2 = \partial n/\partial x$. Now, using equation 3.89 we again obtain equation 3.143.

Equation 3.143 has oscillatory harmonic solutions $x(z)$ and $y(z)$. The angles of the ray are given by the derivatives of the position of the ray: $\theta_x(z) = dx(z)/dz$, $\theta_y(z) = dy(z)/dz$. Making use of this fact, it is possible to show that if a ray is incident onto a quadratic graded-index medium at the plane $z = z_1$ at $(x_1, y_1)$ making angles $(\theta_{x1}, \theta_{x2})$, then at the plane $z = z_2 = z_1 + d$, the position $(x_2, y_2) = (x(z_2), y(z_2))$ and angles $(\theta_{x2}, \theta_{y_2}) = (\theta_x(z_2), \theta_y(z_2))$ will be given by

$$x_2 = x_1 \cos(d/\chi) + \theta_{x1}\chi \sin(d/\chi), \tag{3.144}$$

$$y_2 = y_1 \cos(d/\chi) + \theta_{y_1}\chi \sin(d/\chi), \tag{3.145}$$

$$\theta_{x2} = -\frac{x_1}{\chi}\sin(d/\chi) + \theta_{x1}\cos(d/\chi), \tag{3.146}$$

$$\theta_{y_2} = -\frac{y_1}{\chi}\sin(d/\chi) + \theta_{y_1}\cos(d/\chi). \tag{3.147}$$

The period of oscillation is $2\pi\chi$.

### 3.6.4 Extensions

Homogeneous regions with refractive indices other than unity, spherical refracting surfaces between such homogeneous regions, and spherical mirrors are easily handled with Snell's law and the law of reflection (Saleh and Teich 1991). Again we will not explicitly discuss such components, since they can be handled by simple tricks. Homogeneous regions with refractive index $n \neq 1$ can be handled by working with $\sigma_x = \theta_x/\lambda$ and $\sigma_y = \theta_y/\lambda$ instead of the angles themselves, where $\lambda$ is the wavelength in the homogeneous medium. Spherical refractive surfaces can be treated like lenses by inserting an infinitesimal section of free space on both sides, and spherical mirrors can be handled by folding the optical axis.

All of the components we have treated so far were characterized by linear relations between the output position and angles and input position and angles:

$$x_2 = \check{A}_x x_1 + \check{B}_x \frac{\theta_{x1}}{\lambda}, \qquad y_2 = \check{A}_y y_1 + \check{B}_y \frac{\theta_{y_1}}{\lambda}, \tag{3.148}$$

$$\frac{\theta_{x2}}{\lambda} = \check{C}_x x_1 + \check{D}_x \frac{\theta_{x1}}{\lambda}, \qquad \frac{\theta_{y_2}}{\lambda} = \check{C}_y y_1 + \check{D}_y \frac{\theta_{y_1}}{\lambda}, \tag{3.149}$$

where the parameters $\check{A}, \check{B}, \check{C}, \check{D}$ with subscripts $x, y$ are constants. An example of a common component which cannot be characterized by these relations is a prism (Saleh and Teich 1991), which would require the addition of constant terms to the right hand sides of the above equations. The effect of a prism is similar to the effect of tilting the optical axis and can be analyzed in the same manner. In certain instances, one might also want to be able to deal with transverse displacements of the optical axis, which also require the addition of constant terms on the right hand sides of the above equations. Prisms, as well as tilts and displacements of the optical axis will not be considered in this book.

### 3.6.5 Spatial filters

The most common type of spatial filter discussed in geometrical optical terms are opaque apertures. Rays are completely blocked when they intercept the non-transparent parts of an aperture but pass unhindered through the transparent parts.

Systems involving more general spatial filters are usually analyzed with wave optics; we will exclusively do so in this book. Nevertheless, it is possible to assign an amplitude or intensity to each ray in a bundle and thus represent quite general distributions of light. Attenuating or amplifying filters will then modify this amplitude or intensity. Phase filters (including thin plates with variable thickness or graded index) have the effect of bending the rays by an angle determined by the local partial derivatives of the phase function. However, such approaches are less frequently employed.

### 3.6.6 Fourier-domain spatial filters

Similar comments apply to Fourier-domain spatial filters, which are likewise usually analyzed with wave optics. However, it is instructive to examine the behavior of rays in the three Fourier transform stages discussed on page 138 (figure 3.8). It is seen in all parts of the figure that parallel rays converge to a point on the Fourier plane, corresponding to the fact that the Fourier transform of a harmonic function is a delta function. Rays emanating from a point on the input plane appear in the Fourier plane as a bundle of parallel rays, whose angle with respect to the optical axis is determined by the position of the point. The position $x_{\mathrm{F}}$ and angle $\theta_{x\mathrm{F}}$ of a ray at the Fourier plane are related to those at the input plane through the relations:

$$x_{\mathrm{F}} = s^2 \sigma_{x\mathrm{in}}, \tag{3.150}$$

$$\sigma_{x\mathrm{F}} = -\frac{x_{\mathrm{in}}}{s^2}, \tag{3.151}$$

where $\sigma_x = \theta_x/\lambda$. This equation is the geometrical-optical analog of equation 3.113. It is an instructive exercise to verify the above relations using simple geometry or the input-output relations for each component. The intuition gained by carefully examining figure 3.8 from both wave optical and geometrical optical perspectives goes a long way towards understanding the relationship between these two descriptions of light.

### 3.6.7 General linear systems

Once again such systems will be exclusively dealt with in wave optical terms.

### 3.6.8 Spherical reference surfaces

In certain cases, it may be desirable to specify the intercepts and angles of rays with respect to spherical reference surfaces. It is possible to write relations between the ray parameters with respect to the planar and spherical reference surfaces shown in figure 3.5. However, these are not presented since we will not be making use of such results.

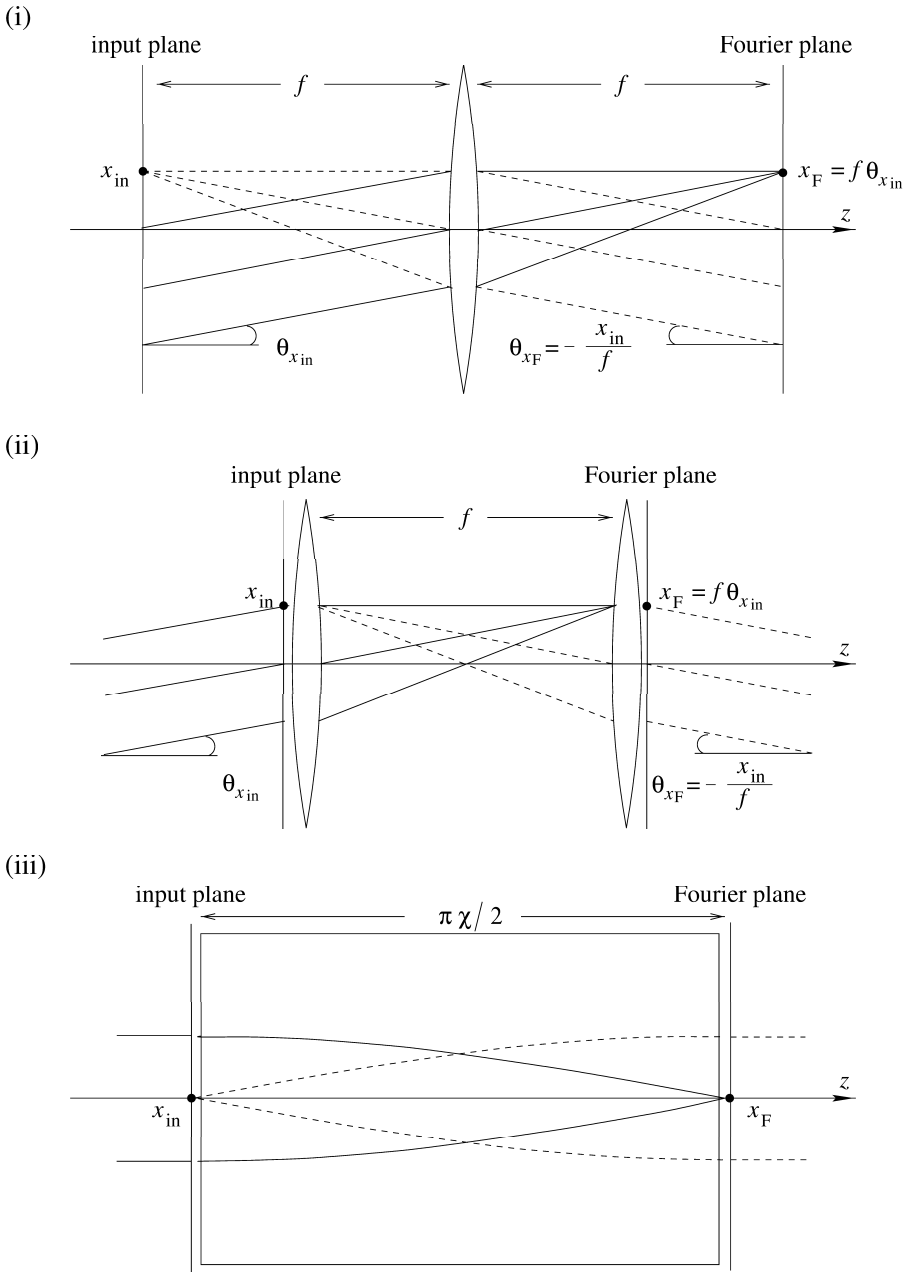


Figure 3.8: Fourier transform stages introduced on page 138. The lenses have focal length $f$. (i) and (ii) $s = \sqrt{\lambda f}$, (iii) $s = \sqrt{\chi/\sigma}$.

### 3.6.9 Remarks

We have ignored attenuations and reflections from the components discussed, concentrating only on how they bend the rays (which corresponds to modification of the phase of the wavefield). Furthermore, we are focusing our attention to systems for which the ray position and angles at the output are linearly related to those at the input, assuming that higher-order dependences can be neglected. This corresponds to neglecting higher than quadratic terms in the phase in wave optics. (Constant terms—corresponding to shifts and tilts—are also not included for simplicity, though most results can be easily generalized to include them as well.) We will see in chapter 4 that systems represented by equations 3.148 and 3.149 play a central role in the first-order study of optical systems.

## 3.7 Partially coherent light

In sections 3.3 and 3.4 we assumed that the distribution of light at any given plane could be represented by deterministic functions $\check{f}(x, y)$. In some cases, either due to the intrinsic random nature of light sources or the random nature of the media which they travel through, it is more appropriate to represent light as random processes.

In this book, we generally assume that the light is temporally stationary and ergodic, and also *quasi-monochromatic*, which means it can effectively be assumed to be *temporally coherent*. The temporal power spectral density of such light at any given point in space, is confined to a narrow band of temporal frequencies. (This assumption requires that the path length differences in the optical system are much smaller than the *coherence length* of the light.) For definitions and further discussion of these concepts, we refer the reader to Goodman 1985 and Saleh and Teich 1991.

Light which is quasi-monochromatic may nevertheless exhibit different degrees of *spatial coherence*. Light with partial spatial coherence is characterized by its autocorrelation $\check{R}_{\check{f}\check{f}}$, better known in optics as its *mutual intensity*, defined by

$$\check{R}_{\check{f}\check{f}}(x_1, y_1; x_2, y_2) \equiv \langle \check{f}(x_1, y_1)\check{f}^*(x_2, y_2)\rangle, \tag{3.152}$$

where the angular brackets denote an ensemble average. When $x_1 = x_2 = x$ and $y_1 = y_2 = y$, we obtain the average intensity

$$\check{R}_{\check{f}\check{f}}(x, y; x, y) = \langle |\check{f}(x, y)|^2\rangle = \check{I}_{\check{f}}(x, y). \tag{3.153}$$

When the light is *spatially coherent*, the average is redundant and $\check{R}_{\check{f}\check{f}}(x_1, y_1; x_2, y_2) = \check{f}(x_1, y_1)\check{f}^*(x_2, y_2)$. This corresponds to the deterministic case, where we usually simply work with the amplitude $\check{f}(x, y)$ rather than the mutual intensity. Light may be treated as spatially coherent if the function $\check{R}_{\check{f}\check{f}}(x_1, y_1; x_2, y_2)$ does not become small for the range of values $x_1, y_1, x_2, y_2$ can take within the aperture of the optical system. When the light is *spatially incoherent*, $\check{R}_{\check{f}\check{f}}(x_1, y_1; x_2, y_2)$ behaves like the delta function $\delta(x_1 - x_2, y_1 - y_2)$. This corresponds to the case where distinct spatial points, even very close ones, do not

exhibit any correlation with each other. Light may be treated as spatially incoherent if the spatial extent over which the value of $\check{R}_{\check{f}\check{f}}(x_1, y_1; x_2, y_2)$ is substantially greater than zero is smaller than the resolution of the optical system.

The definitions and results presented in section 1.8 on random processes directly apply and can be used, for instance, to determine the mutual intensity $\check{R}_{\check{g}\check{g}}$ at the output of an optical system characterized by the kernel $\check{h}(x, y; x', y')$, in terms of the mutual intensity $\check{R}_{\check{f}\check{f}}$ at its input:

$$\check{R}_{\check{g}\check{g}}(x_1, y_1; x_2, y_2) = \iiiint \check{R}_{\check{f}\check{f}}(x'_1, y'_1; x'_2, y'_2) \times \check{h}(x_1, y_1; x'_1, y'_1)\check{h}^*(x_2, y_2; x'_2, y'_2)\, dx'_1\, dy'_1\, dx'_2\, dy'_2. \tag{3.154}$$

If the system is a thin spatial filter with $\check{h}(x, y; x', y') = \check{h}(x, y)\delta(x - x', y - y')$, then $R_{\check{g}\check{g}}(x_1, y_1; x_2, y_2) = \check{h}(x_1, y_1)\check{h}^*(x_2, y_2)R_{\check{f}\check{f}}(x_1, y_1; x_2, y_2)$, and $\check{I}_{\check{g}}(x, y) = |\check{h}(x, y)|^2\check{I}_{\check{f}}(x, y)$. When the light is spatially coherent, equation 3.154 simply reduces to a duplicated form of the coherent relation

$$\check{g}(x, y) = \iint \check{h}(x, y; x', y')\check{f}(x', y')\, dx'\, dy'. \tag{3.155}$$

On the other hand, when the light is spatially incoherent, equation 3.154 leads to the following relation in terms of the intensities:

$$\check{I}_{\check{g}}(x, y) = \kappa \iint |\check{h}(x, y; x', y')|^2\check{I}_{\check{f}}(x', y')\, dx'\, dy', \tag{3.156}$$

where $\kappa$ is a constant (see Goodman 1985, page 206 or Saleh and Teich 1991, page 368).

## 3.8 Fourier optical systems

The purpose of this chapter has been to present the basic tools needed to analyze *Fourier optical systems*. In this book, this term refers to centered optical systems consisting of arbitrary concatenations of sections of free space in the Fresnel approximation, thin spherical lenses, sections of quadratic graded-index media, and thin spatial filters. Although more general systems can also be analyzed in terms of Fourier transforms and linear systems theory, such systems will not be dealt with in this book. We will see in chapter 4 that centered systems consisting of arbitrary concatenations of sections of free space, thin lenses, and sections of graded-index media constitute the class of *first-order optical systems* or *quadratic-phase systems* (Bastiaans 1979a). Thus Fourier optical systems consist of thin spatial filters sandwiched between any number of first-order optical systems.

The key results that will be most frequently used are that describing the effect of a section of free space (equation 3.76), that describing the effect of a thin lens (equation 3.80), that describing the effect of quadratic graded-index media (equation 3.100), and that describing the effect of a thin spatial filter (equation 3.108). Using these results

consecutively, it is possible to analyze any first-order or Fourier optical system and obtain either the overall input-output relation or the distribution of light at any desired plane.

Although we will not present the explicit derivations, here we will list the overall input-output relations for a number of important elementary systems. (The derivations are much simplified by the matrix formalism to be discussed in chapter 4.)

First, we assume that a spatial filter with amplitude transmittance $\check{f}(x,y)$ is situated at $z=0$ and illuminated with a unit-amplitude plane wave. Then it is possible to show that the amplitude distribution $\check{g}(x,y)$ at the plane $z=d$ when $d\to\infty$ is given by (Saleh and Teich 1991)

$$\check{g}(x,y)=\frac{e^{i2\pi\sigma d}}{i\lambda d}\,e^{i\pi(x^2+y^2)/\lambda d}\,\check{F}\left(\frac{x}{\lambda d},\frac{y}{\lambda d}\right), \tag{3.157}$$

where $\check{F}$ represents the Fourier transform of $\check{f}$.

Another important system is the so-called "$2f$" system which consists of a lens of focal length $f$ sandwiched between two sections of free space of length $f$ each. This system was discussed in subsection 3.4.6, where it was shown that it acts as a Fourier transformer. Yet another important system is the "$4f$" system also discussed in the same subsection. This system can be used to convolve an input light distribution with a desired function, or in other words to implement a desired Fourier-domain filter. The necessary filters can be realized holographically, by using computer-aided and/or lithographic techniques, or by using a spatial light modulator. Discussion of these are beyond the scope of this book.

The possibility of realizing desired convolutions and spatial filters has led to a vast and diverse array of applications often referred to as analog optical information processing (or signal processing). Very broadly speaking, these applications can be roughly classified into two categories: those which are applications of convolution, and those which are applications of correlation. Mathematically, correlation is closely related to convolution; however their interpretations are very different. The former category includes applications such as beam shaping, image enhancement, and Wiener filtering for image restoration and noise removal. The latter category includes matched filtering and more advanced approaches in pattern recognition.

The single-lens imaging system is a very important system in optics which consists of an "object" situated a distance $d_{\rm o}$ to the left of a lens with focal length $f$, with the "image" observed a distance $d_{\rm i}$ to the right, such that

$$\frac{1}{f}=\frac{1}{d_{\rm o}}+\frac{1}{d_{\rm i}}, \tag{3.158}$$

an equation known as the imaging condition. If the object (input) complex amplitude distribution is $\check{f}(x,y)$, the image (output) complex amplitude distribution is

$$\propto e^{-i\pi(x^2+y^2)/\lambda M f}\,\check{f}(x/M,y/M), \tag{3.159}$$

where $M\equiv -d_{\rm i}/d_{\rm o}$ is the *magnification* of the imaging system. This result is valid for an infinite lens whose transmittance function is given by equation 3.80. In reality, the lens

will have a finite aperture. If the aperture is characterized by an aperture function $\check{p}(x,y)$ which is 1 where the aperture is transparent and 0 where it is opaque, such a lens can be modeled by the transmittance function $\check{p}(x,y)\exp[-i\pi(x^2+y^2)/\lambda f]$. Using this new transmittance it is possible to re-analyze the single-lens coherent imaging system (as well as the $2f$ or $4f$ systems discussed above) to find the correct input-output relation. It is commonly stated in many sources that the overall effect of a finite lens aperture is to result in a space-invariant system whose transfer function is $\propto \check{p}(-\lambda d_{\rm i}\sigma_x, -\lambda d_{\rm i}\sigma_y)$. However, this widely stated result is not correct and the overall coherent imaging system is not space-invariant (Rhodes 1998). The lens aperture cannot be a Fourier-domain filter since the Fourier transform is in general not observed at the lens but elsewhere in the system. Greater insight into this system can be gained by using fractional Fourier transforms. We also note that when incoherent light is used, the same system does become space-invariant and can be modeled by a point-spread function or a Fourier-domain transfer function.

Real imaging systems may consist of several lenses with different aperture sizes. An approximate but useful way of analyzing such systems is to determine the limiting aperture of the system (known as the aperture stop) and find its image with respect to the output plane (known as the exit pupil), using geometrical optics techniques (see subsection 4.4.3). Then, for purposes of determining the resolution of the system, we can instead analyze the problem of a converging spherical wave incident on the exit pupil. If the exit pupil was infinite in extent, this converging spherical wave would be focused to a point at the image plane, corresponding to perfect imaging with no blur. The finiteness of the exit pupil leads to a finite spot size instead, determining the resolution of the system. If the exit pupil is a distance $d$ to the left of the image plane and is represented by the function $\check{p}(x,y)$, then at the image plane the distribution of light is proportional to

$$e^{i\pi(x^2+y^2)/\lambda d}\,\check{P}\left(\frac{x}{\lambda d},\frac{y}{\lambda d}\right). \tag{3.160}$$

Thus instead of a point, we observe a distribution of light given by the Fourier transform of the exit pupil function. If one considers a simple circular or rectangular aperture of diameter $D$, then it is easy to show that this distribution has a width of about $\lambda d/D$. This quantity is a measure of the resolution of the imaging system. Image features closer than $\lambda d/D$ cannot be distinctly resolved from each other. In certain cases, the distance $d$ corresponds to the focal length $f$ of the overall imaging system or compound lens, so that the size of the smallest resolvable feature is written as $f_{\#}\lambda$, where $f_{\#} \equiv f/D$ is referred to as the "$f$-number." Smaller $f$-numbers mean better resolution. The fact that the size of the smallest resolvable feature is inversely proportional to the size of the aperture is, of course, a consequence of the uncertainty relation discussed in section 1.7.

## 3.9 Further reading

Our coverage of elementary optics in this chapter has been brief and skewed towards the needs of later chapters. Classic introductory books on general optics include Saleh and Teich 1991, Möller 1988, Klein and Furtak 1986, Hecht, Zajac, and Guardino 1997, and Jenkins and White 1976. More advanced books include Born and Wolf 1980, Stavroudis 1972, Solimeno, Crosignani, and Di Porto 1986, and Mandel and Wolf 1995.

Classic introductory texts on Fourier optics and optical information processing include Goodman 1996, Lohmann 1986, Papoulis 1968, Iizuka 1987, Reynolds and others 1989, Yu 1983, Gaskill 1978, and Cathey 1974. More advanced treatments include VanderLugt 1992, Yu and Jutamulia 1992, Javidi and Horner 1994, Boone 1997, and Yu and Jutamulia 1998. For a discussion of sampling theory in the context of optics, see Gori 1993.

Although this book does not deal with quantum optics, the fact that the mathematical techniques discussed have found many applications in these areas warrants inclusion of a few references: Walls and Milburn 1994, Mandel and Wolf 1995, and Yamamoto and İmamoğlu 1999.

# Chapter 4

# Phase-Space Optics

## 4.1 Wave-optical and geometrical-optical phase spaces

In chapter 1 we discussed the basic concepts used in the study of signals and systems, and in chapter 3 we discussed optical signals and systems. In chapter 2 we discussed time- or space-frequency representations and linear canonical transforms, and in this chapter we discuss the use of these concepts in optics. Our main interest will be in the spatial distribution of light so that we will usually deal with space-frequency representations rather than time-frequency representations. Despite the fact that optical signals are most commonly two dimensional, leading to four-dimensional space-frequency representations, in this chapter we will mostly discuss one-dimensional optical signals leading to two-dimensional space-frequency representations, since the latter are much easier to visualize. *Phase space* is the space in which space-frequency representations exist, and is also referred to as the space-frequency plane for one-dimensional signals and systems. One of the two dimensions of phase space is usually a spatial coordinate, whereas the other dimension may be either spatial frequency, the angle or sine of the angle or slope of a ray, or a quantity corresponding to momentum.

We will discuss both wave optical and geometrical optical phase-space representations. Let us consider an optical signal corresponding to the amplitude distribution of light at a given plane. Although other alternatives are also possible, we will employ the Wigner distribution $\check{W}_{\check{f}}(x,\sigma_x)$ of the optical signal $\check{f}(x)$ as its phase-space representation (equation 2.16):

$$\check{W}_{\check{f}}(x,\sigma_x) = \int \check{f}(x+x'/2)\check{f}^*(x-x'/2)e^{-i2\pi\sigma_x x'}\,dx'. \tag{4.1}$$

The properties given in equations 2.17, 2.18, and 2.19 remain valid, with the appropriate replacement of dimensional variables. When we integrate the Wigner distribution over all spatial frequencies, we obtain the intensity distribution

$$\check{I}_{\check{f}}(x) = |\check{f}(x)|^2 = \int \check{W}_{\check{f}}(x,\sigma_x)\,d\sigma_x. \tag{4.2}$$

When we integrate the Wigner distribution over space, we obtain the spectral distribution of power

$$|\check{F}(\sigma_x)|^2 = \int \check{W}_{\check{f}}(x, \sigma_x)\, dx. \tag{4.3}$$

When we integrate over both $x$ and $\sigma_x$, we obtain the total power of the optical signal:

$$\iint \check{W}_{\check{f}}(x, \sigma_x)\, dx\, d\sigma_x = \text{signal power}. \tag{4.4}$$

Notice that the optical signal power (which might be expressed in Watts) is the quantity corresponding to the mathematical concept of "energy" defined in chapter 1. This is because our main interest is in the distribution of this total power in space and spatial frequency. Temporally, the light is assumed to be quasi-monochromatic, so that it is appropriate to consider the energy in a unit time interval (which is the power), rather than the total energy (the energy from $t = -\infty$ to $+\infty$ would be infinite). If we had been dealing with signals of finite duration rather than quasi-monochromatic signals, then one could consider the spatio-temporal Wigner distribution whose integral over all variables could be interpreted as the actual energy in Joules (Mendlovic and Zalevsky 1997).

Roughly speaking, the Wigner distribution of an optical signal gives us the distribution of optical power over space and spatial frequency. In other words, it provides us information about the local frequency content of the optical wave at a certain location.

The Wigner distributions of several elementary signals were given in table 2.1. We see that the Wigner distribution of a plane wave $\exp(i2\pi\sigma_{x0}x)$ (which makes an angle $\arcsin(\lambda\sigma_{x0})$ with the $z$ axis), is given by $\delta(\sigma_x - \sigma_{x0})$. A point source $\delta(x - x_0)$ has a Wigner distribution given by $\delta(x - x_0)$. Likewise, the remaining entries of the table can be respectively interpreted as a parabolic wave, a Gaussian beam, and light coming through a simple aperture.

We will not further repeat the content of chapter 2, but the reader should bear in mind that the results and discussions presented there are directly applicable to optical signals and systems. We will also not discuss space-frequency representations other than the Wigner distribution in an optical context. Use of the ambiguity function in optics is discussed in Papoulis 1974. A discussion of wavelets from an optics perspective is given in Li and Sheng 1998. We also recall that the Fresnel integral can be cast in the form of a wavelet transform (see page 64 and Onural 1993).

Whereas the coordinates of wave-optical phase space are space and spatial frequency, the coordinates of geometrical-optical phase space will be chosen as the position and angle of the rays. In some contexts the sine of the angle or the slope of the ray is considered instead; however, in the paraxial approximation where all angles are assumed to be small, all of these are equivalent. The reader should just bear in mind that the angle can also be interpreted as the slope of the ray or the derivative of the function $x(z)$ describing the trajectory of the ray.

As usual, the position will be denoted by $x$. As for the angle, we will choose to work with the normalized angle $\sigma_x \equiv \sin\theta_x/\lambda \approx \theta_x/\lambda$ introduced on page 140. Although we

will interpret $\sigma_x$ primarily as an angle in geometrical-optical contexts, we are using the same notation we use for spatial frequency, since $\sigma_x$ directly corresponds to the spatial frequency of the spatial harmonic at $z = 0$ associated with a plane wave whose wave vector is parallel to the ray with angle $\theta_x$. This association allows us to look at wave-optical and geometrical-optical phase spaces in a unified manner. It is also possible to interpret $\theta_x/\lambda$ as a quantity corresponding to momentum, which allows an analogy with the phase space of mechanics whose coordinates consist of spatial variables and momenta associated with these spatial variables.

We recall from chapter 3 that in geometrical optics a distribution of light is represented by a bundle of rays. In general, there will be a continuum of rays crossing a given plane at different points $x$ making different normalized angles $\sigma_x = \theta_x/\lambda$. Each such ray will be represented in phase space by the point $(x, \sigma_x)$. It is also possible to assign weights to each ray and thus to each point in phase space. These weights might be represented as a function of $(x, \sigma_x)$ defined over the whole $x$-$\sigma_x$ plane. If there is no ray at position $x$ making angle $\sigma_x$, the value of this function is zero at that point $(x, \sigma_x)$. We will further discuss the relationship between the two phase spaces throughout this chapter.

Figure 4.1 shows how several different ray bundles look in phase space, where we have assumed all rays to have the same weight. More general ray bundles will cover more general regions. The usual case is for these regions to consist of one or more well-connected regions, rather than a scattering of isolated points. As these bundles of rays pass through optical systems, the region and its boundary will be distorted such that any ray on the boundary will continue to remain on the boundary. The closest wave-optical analog of such a boundary is a contour of the Wigner distribution which contains "most" of the signal energy (say 99%).

Here we also briefly point out the relationship of the phase-space quantities discussed above to *radiometric* or *photometric* quantities (Born and Wolf 1980). Consider an extended light source illuminating a certain region of space. The power (energy per unit time) coming from an element of area $\delta A$ centered at a point $P$ on this source, and falling within an element of solid angle $\delta\Omega$ centered around the direction of $\Theta$, is given by $B(P, \Theta) \cos\theta \, \delta A \, \delta\Omega$. Here $\theta$ is the angle between the surface normal at $P$ and the direction of $\Theta$, and $B(P, \Theta)$ is referred to as the *photometric brightness*. Here the term *photometric* implies that we are referring to physical quantities measured in Watts and so on, rather than measures of visual sensation. The $\cos\theta$ term is an obliquity factor corresponding to the fact that the projection of the area element in the direction $\Theta$ is given by $\cos\theta \, \delta A$. We observe that as a function of position and angle, the brightness is also a kind of phase-space density, analogous to the Wigner distribution. When we integrate the brightness over the area of the source as $\int B(P, \Theta) \cos\theta \, dA$, we obtain the *photometric intensity* as a function of the direction of $\Theta$. Integration of this quantity over all relevant angles yields the total power. When we integrate the brightness throughout the solid angle as $\int B(P, \Theta) \cos\theta \, d\Omega$, we obtain the *photometric illumination* as a function of the point $P$. Integration of this quantity over the source area yields the total power.

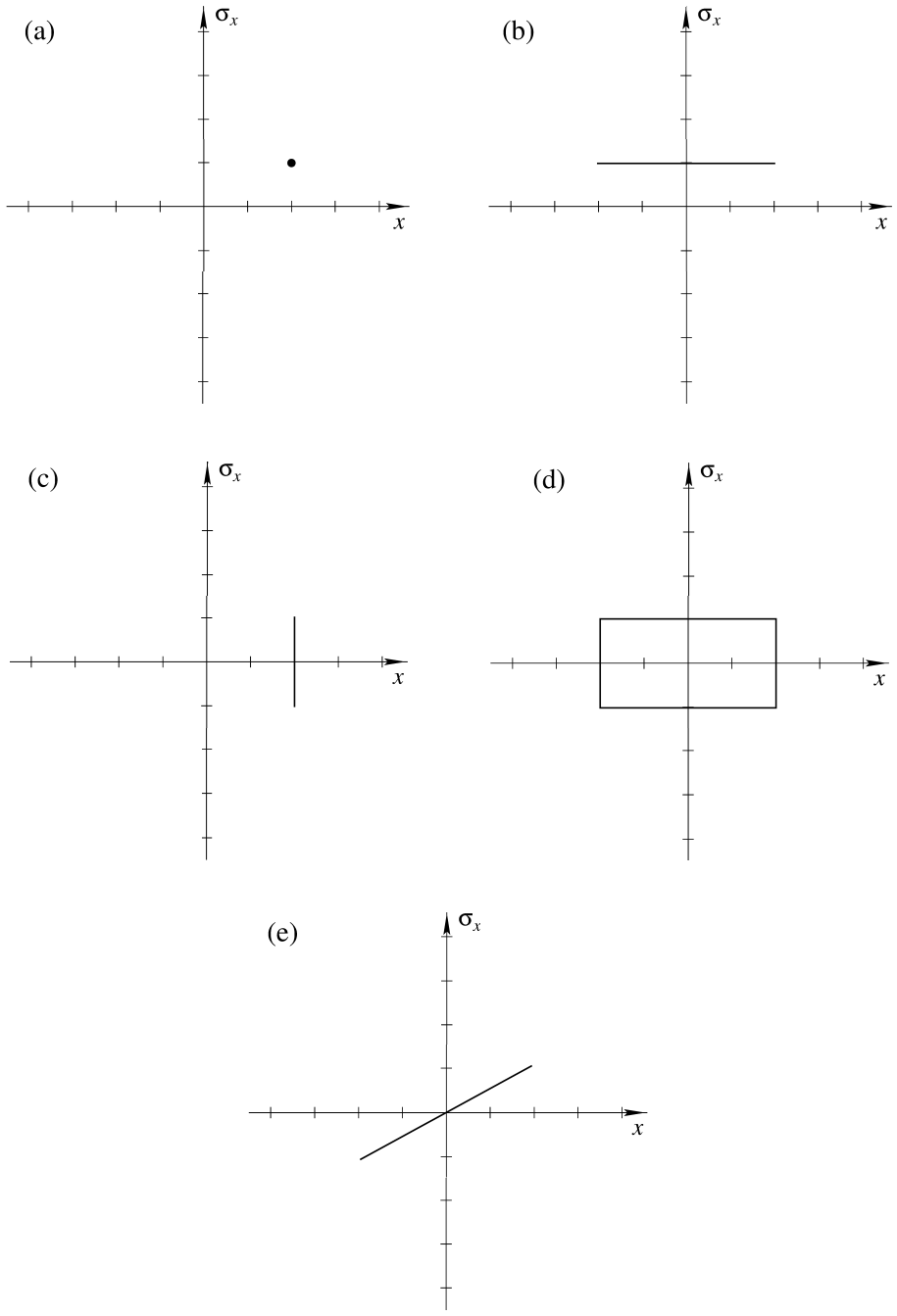


Figure 4.1: Representation of various ray bundles in phase space. (a) A single ray, (b) a bundle of parallel rays, (c) a bundle of rays emanating from the same point, (d) a bundle of rays uniformly distributed over different positions and angles, (e) a bundle of rays corresponding to a spherical wave emanating from a point to the left.

The integrals of $B(P,\Theta)$ over one of the variables are analogous to the projections of the Wigner distribution. Integration over both variables gives us the total power. Further discussion of these quantities may be found in Born and Wolf 1980, pages 181–183. It must be noted that the field of radiometry and photometry displays a wealth of terminology which might be potentially confusing. The terminology used above (Born and Wolf 1980) is not consistent with that used in some other sources and elsewhere in this book. When reference is made to visual sensation, as opposed to true physical magnitude, photometric power, brightness, intensity, and illumination are replaced by luminous energy, luminance, luminous intensity (candle power), and illumination, respectively. Yet another set of terms—in the same respective order—is power, radiance, intensity, and irradiance. A brief introduction to the relationship of radiometric quantities to the Wigner distribution may be found in Bastiaans 1997. What is important for our purpose is to realize that conventional radiometric and photometric quantities are essentially analogous to the Wigner distribution and its integrals (projections or marginals).

Two-dimensional phase space is defined by the variables $(x, y, \sigma_x, \sigma_y)$. Since working with these four variables makes the notation more cumbersome and requires dealing with four-dimensional functions, we will mostly restrict ourselves to the one-dimensional case.

## 4.2 Quadratic-phase systems and linear canonical transforms

Sections of free space in the Fresnel approximation, thin lenses, and sections of quadratic-graded index media, as well as arbitrary combinations of these belong to the class of *quadratic-phase systems* (Bastiaans 1979a), which are mathematically the same as the class of linear canonical transforms we discussed in chapter 2. The output of a one-dimensional quadratic-phase system is related to its input through

$$\check{g}(x) = \int \check{h}(x,x')\check{f}(x')\,dx', \tag{4.5}$$

$$\check{h}(x,x') = \sqrt{\check{\beta}}\,e^{-i\pi/4}\,\exp\left[i\pi(\check{\alpha}x^2 - 2\check{\beta}xx' + \check{\gamma}x'^2)\right],$$

where $\check{\alpha}$, $\check{\beta}$, $\check{\gamma}$ are the three independent parameters of the system (equation 2.67). We have chosen the normalization such that the system is unitary. In two dimensions

$$\check{g}(x,y) = \iint \check{h}(x,y;x',y')\check{f}(x',y')\,dx'\,dy', \tag{4.6}$$

$$\check{h}(x,y;x',y') = -i\check{\beta}\,\exp\left[i\pi\left(\check{\alpha}(x^2+y^2) - 2\check{\beta}(xx'+yy') + \check{\gamma}(x'^2+y'^2)\right)\right].$$

It is also possible to consider the case where the parameters are different for the two dimensions (Bastiaans 1979a, Sahin, Ozaktas, and Mendlovic 1998).

Again, $s$ denotes the implicit scale parameter relating the dimensionless variables $u, v$ to the variables $x, y$ with dimensions of length: $x \equiv su$ and $y \equiv sv$. As in chapter 3, $s\check{f}(x,y) \equiv f(x/s, y/s) \equiv f(u,v)$ and $s^{-1}\check{F}(\sigma_x,\sigma_y) \equiv F(s\sigma_x, s\sigma_y) \equiv F(\mu,\nu)$. We are

also introducing the definitions $\check{\alpha} \equiv \alpha/s^2$, $\check{\beta} \equiv \beta/s^2$, $\check{\gamma} \equiv \gamma/s^2$ and $\check{A} \equiv A$, $\check{B} \equiv s^2 B$, $\check{C} \equiv C/s^2$, $\check{D} \equiv D$. These relate the dimensionless parameters of chapters 1 and 2 to the dimensional parameters employed in this chapter and chapter 3. These sets of parameters are again related to each other through the relation (equation 2.75)

$$\begin{bmatrix} \check{A} & \check{B} \\ \check{C} & \check{D} \end{bmatrix} = \begin{bmatrix} \check{\gamma}/\check{\beta} & 1/\check{\beta} \\ -\check{\beta} + \check{\alpha}\check{\gamma}/\check{\beta} & \check{\alpha}/\check{\beta} \end{bmatrix} = \begin{bmatrix} \check{\alpha}/\check{\beta} & -1/\check{\beta} \\ \check{\beta} - \check{\alpha}\check{\gamma}/\check{\beta} & \check{\gamma}/\check{\beta} \end{bmatrix}^{-1}, \tag{4.7}$$

with $\check{A}\check{D} - \check{B}\check{C} = 1$. Once again, it is the case that the matrix associated with the system obtained by concatenating two systems, is given by the product of the matrices of the two systems. The kernel in equation 4.5 can be rewritten in terms of $\check{A}$, $\check{B}$, $\check{C}$, $\check{D}$ as

$$\check{h}(x, x') = \sqrt{\frac{1}{\check{B}}}\, e^{-i\pi/4} \exp\left[\frac{i\pi}{\check{B}}\left(\check{D}x^2 - 2xx' + \check{A}x'^2\right)\right]. \tag{4.8}$$

This formula is found in the literature in a variety of contexts and has been referred to by various names (for instance, "quadratic-phase systems" in Bastiaans 1979a, "linear canonical transforms" in Wolf 1979, "generalized Huygens integral" in Siegman 1986, "generalized Fresnel transform" in James and Agarwal 1996, "special affine Fourier transforms" in Abe and Sheridan 1994a, b).

The effect of a quadratic-phase system on the Wigner distribution of an optical signal follows from the discussion of subsection 2.4.2. The Wigner distribution of the output signal $\check{W}_{\check{g}}$ is related to that of the input signal by

$$\check{W}_{\check{g}}(x, \sigma_x) = \check{W}_{\check{f}}(\check{D}x - \check{B}\sigma_x, -\check{C}x + \check{A}\sigma_x). \tag{4.9}$$

The kernel relating these Wigner distributions is given by

$$\begin{aligned} \check{W}_{\check{g}}(x, \sigma_x) &= \iint \check{K}_{\check{h}}(x, \sigma_x; x', \sigma'_x)\, \check{W}_{\check{f}}(x', \sigma'_x)\, dx'\, d\sigma'_x, \\ \check{K}_{\check{h}}(x, \sigma_x; x', \sigma'_x) &= \delta(x' - \check{D}x + \check{B}\sigma_x)\, \delta(\sigma'_x + \check{C}x - \check{A}\sigma_x). \end{aligned} \tag{4.10}$$

The kernel $\check{K}_{\check{h}}(x, \sigma_x; x', \sigma'_x)$ associated with optical components which are not linear canonical transforms (such as a spatial filter) can be determined from equation 2.44.

As in chapter 2, we are restricting ourselves to systems with real $\check{A}$, $\check{B}$, $\check{C}$, $\check{D}$. Allowing complex parameters makes it possible to deal with attenuating apertures of Gaussian profile, and propagation in certain media exhibiting attenuation (or gain). Most of the results presented here are also valid for such systems with complex parameters (Siegman 1986).

When dealing with two-dimensional systems, the above matrices become four-dimensional matrices. The most compact approach is usually to define a two-dimensional space vector and a two-dimensional spatial frequency vector. In terms of these, the four-dimensional matrices become two-by-two block matrices of two-by-two matrices, and most results bear a direct formal similarity to the one-dimensional results. Two-dimensional linear canonical

transforms and their matrices were briefly discussed in section 2.5. The reader is referred to Bastiaans 1979a for a discussion of two-dimensional systems in an optical context. We also note that when dealing with axially symmetric systems, one-dimensional analysis is sufficient for most purposes.

## 4.3 Optical components

We now individually discuss the optical components introduced in section 3.4. Our discussion will be uniform and comprehensive, running in parallel to the wave-optical and geometrical-optical discussion of the same components in chapter 3. The results presented are one-dimensional versions of the results presented in that chapter, so that they are not rederived.

For each optical component, we will first state which mathematical operation the optical component corresponds to. These operations were already studied in detail in chapters 1 and 2 (for instance, see tables 1.2 and 2.7). We will then state the forward and inverse kernels $\check{h}(x,x')$ and $\check{h}^{-1}(x,x')$ of the component, and show how the output amplitude distribution $\check{g}(x)$ is related to the input distribution $\check{f}(x)$. We recall that these quantities are related by

$$\check{g}(x) = \int \check{h}(x,x')\check{f}(x')\,dx'. \tag{4.11}$$

We will also state how the Fourier transform of the output $\check{G}(\sigma_x)$ is related to the Fourier transform $\check{F}(\sigma_x)$ of the input. Comparing the kernels with the general form of the kernel for quadratic-phase systems

$$\check{h}(x,x') = e^{-i\pi/4}\sqrt{\frac{1}{\check{B}}}\,\exp\left[\frac{i\pi}{\check{B}}\left(\check{D}x^2 - 2xx' + \check{A}x'^2\right)\right], \tag{4.12}$$

will reveal that most of the components under consideration are quadratic-phase systems with different matrix parameters, and enable us to identify their $\check{A}$, $\check{B}$, $\check{C}$, $\check{D}$ parameters.

We will then state the kernel $\check{K}_{\check{h}}(x,\sigma_x;x',\sigma'_x)$ relating the Wigner distribution of the output $\check{W}_{\check{g}}(x,\sigma_x)$ to the Wigner distribution of the input $\check{W}_{\check{f}}(x,\sigma_x)$:

$$\check{W}_{\check{g}}(x,\sigma_x) = \iint \check{K}_{\check{h}}(x,\sigma_x;x',\sigma'_x)\,\check{W}_{\check{f}}(x',\sigma'_x)\,dx'\,d\sigma'_x, \tag{4.13}$$

as well as an expression directly relating $\check{W}_{\check{g}}(x,\sigma_x)$ to $\check{W}_{\check{f}}(x,\sigma_x)$. When the system in question is a quadratic-phase system, these relations take the form

$$\check{K}_{\check{h}}(x,\sigma_x;x',\sigma'_x) = \delta(x' - \check{D}x + \check{B}\sigma_x)\,\delta(\sigma'_x + \check{C}x - \check{A}\sigma_x), \tag{4.14}$$

$$\check{W}_{\check{g}}(x,\sigma_x) = \check{W}_{\check{f}}(\check{D}x - \check{B}\sigma_x, -\check{C}x + \check{A}\sigma_x). \tag{4.15}$$

The parameters $\check{A}$, $\check{B}$, $\check{C}$, $\check{D}$ of a component deduced by comparison with equation 4.12, are also always consistent with the above two equations.

In sections 3.5 and 3.6, we had characterized rays by noting their position $x$ and angle $\theta_x$ (or normalized angle $\sigma_x = \theta_x/\lambda$) at a given plane. We had derived linear relations between the output position and normalized angle $(x_2, \sigma_{x2})$ and the input position and normalized angle $(x_1, \sigma_{x1})$ for each component. We now observe that all of these linear relations can be written as a simple matrix equation of the form

$$\begin{bmatrix} x_2 \\ \sigma_{x2} \end{bmatrix} = \begin{bmatrix} \check{A} & \check{B} \\ \check{C} & \check{D} \end{bmatrix} \begin{bmatrix} x_1 \\ \sigma_{x1} \end{bmatrix}, \tag{4.16}$$

with the same parameters $\check{A}$, $\check{B}$, $\check{C}$, $\check{D}$ deduced from wave optical considerations.

We emphasize that the $\check{A}$, $\check{B}$, $\check{C}$, $\check{D}$ parameters of quadratic-phase components can be determined in two ways. We can obtain the kernel $\check{h}(x, x')$ characterizing the component based on wave optics, and compare it with equation 4.12 to identify $\check{A}$, $\check{B}$, $\check{C}$, $\check{D}$. Alternatively, we can write the relation between the input and output ray vectors in matrix form based on geometrical optics, and identify these four parameters from this relation. The fact that both approaches yield the same result reflects an underlying correspondence between the approximations employed in the wave-optical and geometrical-optical approaches. Wave optics and geometrical optics become operationally equivalent under these conditions, despite the fact that they are quite different descriptions of the behavior of light.

We will also provide a figure for each component, illustrating how a rectangle representing either the support of the Wigner distribution or a bundle of rays, is changed by this component. These figures illustrate the effect of the component in phase space. (The reader should compare these with figures 2.3 and 2.6.)

We will then note the eigenfunctions of the component, and also present the associated hyperdifferential form, as well as a differential-equation description of the component. However, we will not discuss these in great detail in this chapter, referring the reader to subsection 2.4.9 for further elaboration of the mathematical relationships between these concepts.

Bastiaans has been a major contributor to the use of Wigner distributions and quadratic-phase systems in optics (1978, 1979a, b, c, d, 1989, 1997).

### 4.3.1 Sections of free space

The physical process of propagation through a section of free space of length $d$ in the Fresnel approximation, mathematically corresponds to **chirp convolution**. Positive values of $d$ correspond to forward propagation and negative values of $d$ correspond to backward propagation. The associated kernel and its inverse are given by

$$\check{h}(x, x') = e^{-i\pi/4} \sqrt{\frac{1}{\lambda d}}\, e^{i\pi(x-x')^2/\lambda d}, \tag{4.17}$$

$$\check{h}^{-1}(x, x') = e^{i\pi/4} \frac{1}{\sqrt{\lambda d}}\, e^{-i\pi(x-x')^2/\lambda d}, \tag{4.18}$$

where the trivial phase factors $\exp(\pm i2\pi\sigma d)$ have been dropped. The output can be expressed in terms of the input as

$$\check{g}(x) = e^{-i\pi/4}\sqrt{\frac{1}{\lambda d}}\int e^{i\pi(x-x')^2/\lambda d}\check{f}(x')\,dx' = \left[e^{-i\pi/4}\sqrt{\frac{1}{\lambda d}}\,e^{i\pi x^2/\lambda d}\right] * f(x). \tag{4.19}$$

In the Fourier domain,

$$\check{G}(\sigma_x) = e^{-i\pi\lambda d\sigma_x^2}\,\check{F}(\sigma_x). \tag{4.20}$$

The kernel relating the Wigner distribution of the output to that of the input, and the Wigner distribution of the output are given by

$$\check{K}_{\check{h}}(x,\sigma_x;x',\sigma_x') = \delta(x-\lambda d\sigma_x - x')\,\delta(\sigma_x-\sigma_x'), \tag{4.21}$$

$$\check{W}_{\check{g}}(x,\sigma_x) = \check{W}_{\check{f}}(x-\lambda d\sigma_x,\sigma_x). \tag{4.22}$$

Upon comparison with equations 4.12 and 4.15, equations 4.17 and 4.22 imply

$$\begin{bmatrix} \check{A} & \check{B} \\ \check{C} & \check{D} \end{bmatrix} = \begin{bmatrix} 1 & \lambda d \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -\lambda d \\ 0 & 1 \end{bmatrix}^{-1}. \tag{4.23}$$

The system inverse of a section of free space of length $d$ is a section of free space of length $-d$, which can be interpreted as backward propagation. The equations relating the position and normalized angle of the output ray to those of the input ray can be written in the form

$$\begin{bmatrix} x_2 \\ \sigma_{x2} \end{bmatrix} = \begin{bmatrix} 1 & \lambda d \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ \sigma_{x1} \end{bmatrix} = \begin{bmatrix} \check{A} & \check{B} \\ \check{C} & \check{D} \end{bmatrix}\begin{bmatrix} x_1 \\ \sigma_{x1} \end{bmatrix}. \tag{4.24}$$

Figure 4.2 shows how a rectangle representing either the support of the Wigner distribution or a bundle of rays, is transformed through this operation.

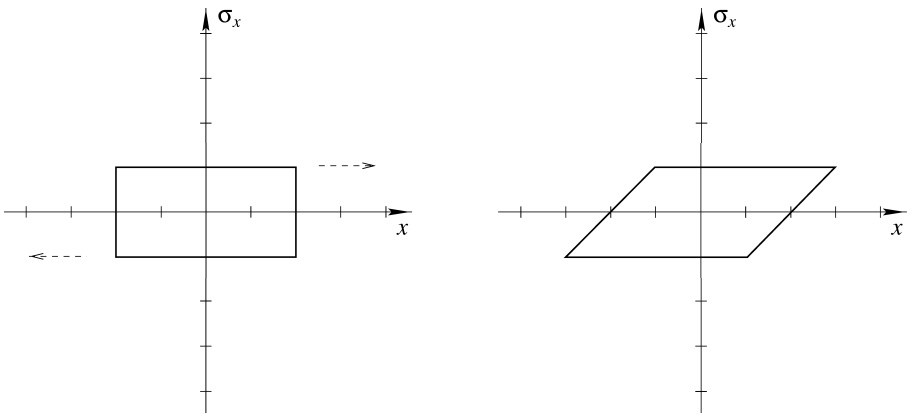


Figure 4.2: The effect of free-space propagation in phase space is horizontal shearing (in the space dimension).

The eigenfunctions of propagation through free space are of the form $\check{f}(x) = \exp(i2\pi\sigma_{x0}x)$, where $\sigma_{x0}$ is any real number. Physically, this corresponds to a spatial harmonic in the transverse plane, and a plane wave throughout space. The effect of chirp convolution on such functions is simply to multiply them with a unit-magnitude complex constant (a phase factor).

The hyperdifferential form of the free-space propagation operator is given by

$$\exp\left(i\frac{\lambda d}{4\pi}\frac{d^2}{dx^2}\right). \tag{4.25}$$

Since the operator corresponding to free-space propagation is a function of $d/dx$ only, all operators of $d/dx$ will remain invariant under free-space propagation. In other words, they commute with free-space propagation. For instance, if we propagate the derivative of an input field $\check{f}(x)$, we will obtain the derivative of the original output $\check{g}(x)$.

Let us now denote the amplitude distribution of light as a function of both $x$ and $z$ by $\check{f}(x,z)$. Then, if we solve the differential equation

$$-\frac{1}{4\pi}\frac{\partial^2\check{f}(x,z)}{\partial x^2} = i\frac{\partial\check{f}(x,z)}{\partial(\lambda z)} \tag{4.26}$$

subject to the boundary condition $\check{f}(x,0) = \check{f}(x)$, the solution $\check{f}(x,z)$ will be related to $\check{f}(x)$ through the Fresnel integral, with $z$ replacing the distance of propagation $d$.

### 4.3.2 Thin lenses

The physical process of passing through a thin lens of focal length $f$ in the paraxial approximation, mathematically corresponds to **chirp multiplication**. Positive values of $f$ correspond to positive lenses and negative values of $f$ correspond to negative lenses. The associated kernel and its inverse are given by

$$\check{h}(x,x') = e^{-i\pi x^2/\lambda f}\,\delta(x-x'), \tag{4.27}$$

$$\check{h}^{-1}(x,x') = e^{i\pi x^2/\lambda f}\,\delta(x-x'). \tag{4.28}$$

The output can be expressed in terms of the input as

$$\check{g}(x) = e^{-i\pi x^2/\lambda f}\,\check{f}(x). \tag{4.29}$$

In the Fourier domain,

$$\check{G}(\sigma_x) = e^{-i\pi/4}\sqrt{\lambda f}\int e^{i\pi\lambda f(\sigma_x-\sigma_x')^2}\check{F}(\sigma_x')\,d\sigma_x' = \left[e^{-i\pi/4}\sqrt{\lambda f}\,e^{i\pi\lambda f\sigma_x^2}\right] * \check{F}(\sigma_x). \tag{4.30}$$

The kernel relating the Wigner distribution of the output to that of the input, and the Wigner distribution of the output are given by

$$\check{K}_{\check{h}}(x,\sigma_x;x',\sigma_x') = \delta(\sigma_x + x/\lambda f - \sigma_x')\delta(x-x'), \tag{4.31}$$

$$\check{W}_{\check{g}}(x,\sigma_x) = \check{W}_{\check{f}}(x,\sigma_x + x/\lambda f). \tag{4.32}$$

Upon comparison with equations 4.12 and 4.15, equations 4.27 and 4.32 imply

$$\begin{bmatrix} \check{A} & \check{B} \\ \check{C} & \check{D} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1/\lambda f & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1/\lambda f & 1 \end{bmatrix}^{-1}. \tag{4.33}$$

The system inverse of a thin lens of focal length $f$ is a thin lens of focal length $-f$: the inverse of a positive lens is a negative lens and vice versa. The equations relating the position and normalized angle of the output ray to those of the input ray can be written in the form

$$\begin{bmatrix} x_2 \\ \sigma_{x2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1/\lambda f & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ \sigma_{x1} \end{bmatrix} = \begin{bmatrix} \check{A} & \check{B} \\ \check{C} & \check{D} \end{bmatrix} \begin{bmatrix} x_1 \\ \sigma_{x1} \end{bmatrix}. \tag{4.34}$$

Figure 4.3 shows how a rectangle representing either the support of the Wigner distribution or a bundles of rays, is transformed through this operation.

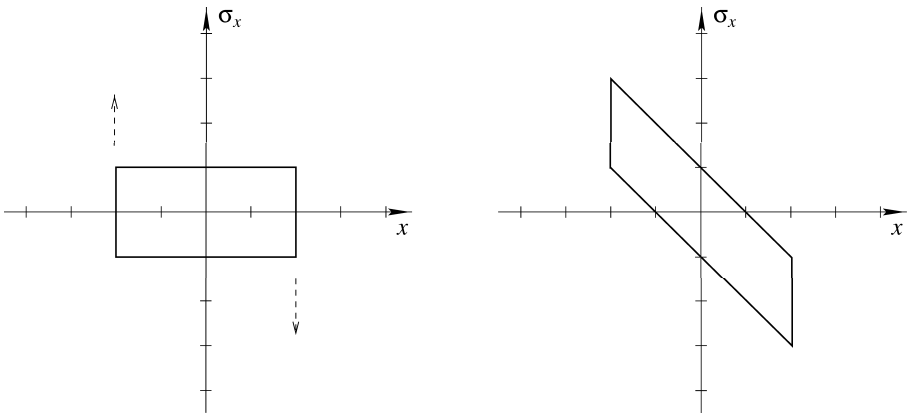


Figure 4.3: The effect of a thin lens in phase space is vertical shearing (in the frequency dimension).

The eigenfunctions of passing through a thin lens are of the form $\check{f}(x) = \delta(x - x_0)$, where $x_0$ is any real number. Physically, this corresponds to a point source. The effect of chirp multiplication on such functions is simply to multiply them with a unit-magnitude complex constant (a phase factor).

The hyperdifferential form of the thin lens operator is given by

$$\exp\left(-i\frac{\pi}{\lambda f}x^2\right). \tag{4.35}$$

Since the operator corresponding to a thin lens is a function of $x$ only, all operators of $x$ will remain invariant under passage through a thin lens. In other words, they commute with the thin lens. For instance, if $x\check{f}(x)$ is incident on the lens, at the output we will obtain $x\check{g}(x)$.

The associated differential equation does not have a meaningful physical interpretation and is thus not presented here.

### 4.3.3 Quadratic graded-index media

The physical process of propagation through a section of quadratic graded-index medium with parameter $\chi$ and of length $d$ mathematically corresponds to **fractional Fourier transformation**. Positive values of $d$ correspond to forward propagation and negative values of $d$ correspond to backward propagation. The associated kernel and its inverse are given by

$$\check{h}(x,x') = \begin{cases} \frac{e^{-id/2\chi}}{\sqrt{\lambda\chi}} A_\alpha \, e^{\frac{i\pi}{\lambda\chi}\left(\cot\alpha\, x^2 - 2\csc\alpha\, xx' + \cot\alpha\, x'^2\right)} & d \neq j\pi\chi \\ e^{-id/2\chi}\,\delta(x-x') & d = 2j\pi\chi \\ e^{-id/2\chi}\,\delta(x+x') & d = (2j\pm 1)\pi\chi \end{cases} \tag{4.36}$$

$$\check{h}^{-1}(x,x') = \begin{cases} \frac{e^{id/2\chi}}{\sqrt{\lambda\chi}} A_\alpha^* \, e^{-\frac{i\pi}{\lambda\chi}\left(\cot\alpha\, x^2 - 2\csc\alpha\, xx' + \cot\alpha\, x'^2\right)} & d \neq j\pi\chi \\ e^{id/2\chi}\,\delta(x-x') & d = 2j\pi\chi \\ e^{id/2\chi}\,\delta(x+x') & d = (2j\pm 1)\pi\chi \end{cases} \tag{4.37}$$

where $j$ is an integer, $\alpha = d/\chi$, and $A_\alpha = \sqrt{1 - i\cot\alpha}$. (The fractional Fourier order $a$ is related to $\alpha$ by $\alpha = a\pi/2$.) In the above equations, the trivial phase factors $\exp(\pm i2\pi\sigma d)$ have been dropped. This kernel can be derived from equation 3.103 by employing the last property in table 1.8.

The output can be expressed in terms of the input as

$$\check{g}(x) = e^{-id/2\chi}(\lambda\chi)^{-1/4} f_a(x/\sqrt{\lambda\chi}\,) = e^{-id/2\chi}\check{f}_a(x) \tag{4.38}$$

for $d \neq j\pi\chi$. Here $f_a(u)$ denotes the $a$th order fractional Fourier transform of $f(u)$, where $(\lambda\chi)^{-1/4} f(x/\sqrt{\lambda\chi}\,) = \check{f}(x)$. In the Fourier domain,

$$\check{G}(\sigma_x) = e^{-id/2\chi}(\lambda\chi)^{1/4} F_a(\sqrt{\lambda\chi}\,\sigma_x) = e^{-id/2\chi}\check{F}_a(\sigma_x). \tag{4.39}$$

Here $F_a(u)$ denotes the $a$th order fractional Fourier transform of $F(u)$, where $(\lambda\chi)^{1/4} F(\sqrt{\lambda\chi}\,\sigma_x) = \check{F}(\sigma_x)$. The kernel relating the Wigner distribution of the output to that of the input, and the Wigner distribution of the output are given by

$$\check{K}_{\check{h}}(x,\sigma_x;x',\sigma_x') = \delta(x\cos\alpha - \sigma_x\lambda\chi\sin\alpha - x')\,\delta(x\sin\alpha/\lambda\chi + \sigma_x\cos\alpha - \sigma_x'), \tag{4.40}$$

$$\check{W}_{\check{g}}(x,\sigma_x) = \check{W}_{\check{f}}(x\cos\alpha - \sigma_x\lambda\chi\sin\alpha, x\sin\alpha/\lambda\chi + \sigma_x\cos\alpha). \tag{4.41}$$

Upon comparison with equations 4.12 and 4.15, equations 4.36 and 4.41 imply

$$\begin{bmatrix} \check{A} & \check{B} \\ \check{C} & \check{D} \end{bmatrix} = \begin{bmatrix} \cos\alpha & \lambda\chi\sin\alpha \\ -\sin\alpha/\lambda\chi & \cos\alpha \end{bmatrix} = \begin{bmatrix} \cos\alpha & -\lambda\chi\sin\alpha \\ \sin\alpha/\lambda\chi & \cos\alpha \end{bmatrix}^{-1}. \tag{4.42}$$

The equations relating the position and normalized angle of the output ray to those of the input ray can be written in the form

$$\begin{bmatrix} x_2 \\ \sigma_{x2} \end{bmatrix} = \begin{bmatrix} \cos\alpha & \lambda\chi\sin\alpha \\ -\sin\alpha/\lambda\chi & \cos\alpha \end{bmatrix} \begin{bmatrix} x_1 \\ \sigma_{x1} \end{bmatrix} = \begin{bmatrix} \check{A} & \check{B} \\ \check{C} & \check{D} \end{bmatrix} \begin{bmatrix} x_1 \\ \sigma_{x1} \end{bmatrix}. \tag{4.43}$$

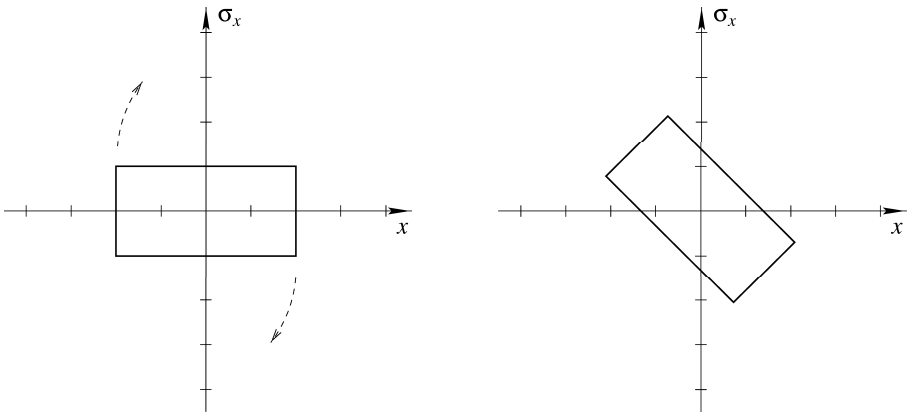


Figure 4.4: The effect of propagation through graded-index media in phase space is rotation.

Figure 4.4 shows how a rectangle representing either the support of the Wigner distribution or a bundles of rays, is transformed through this operation.

The eigenfunctions of propagation through graded-index media are of the form $\check{f}(x) = (\lambda\chi)^{-1/4}\psi_l(x/\sqrt{\lambda\chi})$, where $\psi_l(\cdot)$ is the $l$th order Hermite-Gaussian function. The effect of fractional Fourier transformation on such functions is simply to multiply them with a unit-magnitude complex constant (a phase factor).

The hyperdifferential form of the quadratic graded-index media propagation operator is given by

$$\exp\left[-i\alpha\left(-\frac{\lambda\chi}{4\pi}\frac{d^2}{dx^2}+\pi\frac{x^2}{\lambda\chi}\right)\right]. \tag{4.44}$$

Since the operator corresponding to propagation in this medium is a function of $\lambda\chi(i2\pi)^{-2}d^2/dx^2 + x^2/\lambda\chi$ only, all operators of $\lambda\chi(i2\pi)^{-2}d^2/dx^2 + x^2/\lambda\chi$ will remain invariant under propagation in this medium. In other words, they will commute with propagation in this medium.

Let us now denote the amplitude distribution of light as a function of both $x$ and $z$ by $\check{f}(x,z)$. Then, if we solve the differential equation

$$-\frac{\lambda\chi}{4\pi}\frac{\partial^2\check{f}(x,z)}{\partial x^2}+\pi\frac{x^2}{\lambda\chi}\check{f}(x,z)=i\chi\frac{\partial\check{f}(x,z)}{\partial z} \tag{4.45}$$

subject to the boundary condition $\check{f}(x,0) = \check{f}(x)$, the solution $\check{f}(x,z)$ will be related to $\check{f}(x)$ through a fractional Fourier transformation, with $z$ replacing the distance of propagation $d$.

Quadratic graded-index media also exhibiting a quadratically varying attenuation coefficient are discussed in Siegman 1986. We only note that propagation in such media can be related to complex-ordered fractional Fourier transforms.

### 4.3.4 Extensions

Although the components discussed in subsection 3.4.4 will not be of direct interest to us in this book, we briefly mention that they also have simple effects in phase space. As mentioned in subsection 3.4.4, homogeneous regions with refractive index $n \neq 1$ are transparently handled by normalizing spatial frequencies and angles by $n$, and spherical refractive surfaces are handled in the same manner as lenses. Prisms as well as tilts and displacements of the optical axis simply correspond to shifts in phase space. A prism or a tilt of the optical axis corresponds to a shift in the $\sigma_x$ dimension of phase space, and a displacement of the optical axis corresponds to a shift in the $x$ dimension.

### 4.3.5 Spatial filters

The physical process of spatial filtering with a spatial filter $\check{h}(x)$ mathematically corresponds to **multiplicative filtering**. The associated kernel and its inverse are given by

$$\check{h}(x, x') = \check{h}(x)\delta(x - x'), \tag{4.46}$$

$$\check{h}^{-1}(x, x') = [1/\check{h}(x)]\delta(x - x'). \tag{4.47}$$

The output can be expressed in terms of the input as

$$\check{g}(x) = \check{h}(x)\check{f}(x). \tag{4.48}$$

In the Fourier domain,

$$\check{G}(\sigma_x) = \int \check{H}(\sigma_x - \sigma'_x)\check{F}(\sigma'_x)\, d\sigma'_x = \check{H}(\sigma_x) * \check{F}(\sigma_x). \tag{4.49}$$

The kernel relating the Wigner distribution of the output to that of the input, and the Wigner distribution of the output are given by

$$\check{K}_{\check{h}}(x, \sigma_x; x', \sigma'_x) = \check{W}_{\check{h}}(x, \sigma_x - \sigma'_x)\, \delta(x - x'), \tag{4.50}$$

$$\check{W}_{\check{g}}(x, \sigma_x) = \int \check{W}_{\check{h}}(x, \sigma_x - \sigma'_x)\, \check{W}_{\check{f}}(x, \sigma'_x)\, d\sigma'_x. \tag{4.51}$$

The effect of a spatial filter in phase space is to convolve the Wigner distribution along the frequency dimension with the Wigner distribution of the filter function.

The eigenfunctions of spatial filters are of the form $\check{f}(x) = \delta(x - x_0)$, where $x_0$ is any real number. Physically, this corresponds to a point source. The effect of spatial filters on such functions is simply to multiply them with a complex constant.

### 4.3.6 Fourier-domain spatial filters

The physical process of Fourier-domain spatial filtering mathematically corresponds to **convolutive filtering**. The associated kernel and its inverse are given by

$$\check{h}(x, x') = \check{h}(x - x'), \tag{4.52}$$

$$\check{h}^{-1}(x, x') = \check{h}^{-1}(x - x'). \tag{4.53}$$

The output can be expressed in terms of the input as

$$\check{g}(x) = \int \check{h}(x - x')\check{f}(x')dx' = \check{h}(x) * \check{f}(x). \tag{4.54}$$

In the Fourier domain,

$$\check{G}(\sigma_x) = \check{H}(\sigma_x)\check{F}(\sigma_x). \tag{4.55}$$

The kernel relating the Wigner distribution of the output to that of the input, and the Wigner distribution of the output are given by

$$\check{K}_{\check{h}}(x, \sigma_x; x', \sigma_x') = \check{W}_{\check{h}}(x - x', \sigma_x)\, \delta(\sigma_x - \sigma_x'), \tag{4.56}$$

$$\check{W}_{\check{g}}(x, \sigma_x) = \int \check{W}_{\check{h}}(x - x', \sigma_x)\, \check{W}_{\check{f}}(x', \sigma_x)\, dx'. \tag{4.57}$$

The effect of a Fourier-domain spatial filter in phase space is to convolve the Wigner distribution along the space dimension with the Wigner distribution of the filter function.

The eigenfunctions of Fourier-domain spatial filters are of the form $\check{f}(x) = \exp(i2\pi\sigma_{x0}x)$, where $\sigma_{x0}$ is any real number. Physically, this corresponds to a spatial harmonic in the transverse plane, and a plane wave throughout space. The effect of a Fourier-domain spatial filter on such functions is simply to multiply them with a complex constant.

*Fourier transform stages:* We also discuss the effect in phase space of the Fourier transform stages presented on page 137. The following applies to all three systems, provided the appropriate value for the parameter $s$ is inserted.

The associated kernel and its inverse are given by

$$\check{h}(x, x') \propto e^{-i2\pi xx'/s^2}, \tag{4.58}$$

$$\check{h}^{-1}(x, x') \propto e^{i2\pi xx'/s^2}. \tag{4.59}$$

The output can be expressed in terms of the input as

$$\check{g}(x) \propto \int e^{-i2\pi xx'/s^2} \check{f}(x')\, dx' = \check{F}(x/s^2). \tag{4.60}$$

In the Fourier domain,

$$\check{G}(\sigma_x) \propto \int e^{-i2\pi s^2\sigma_x\sigma_x'} \check{F}(\sigma_x')\, d\sigma_x' = \check{f}(-s^2\sigma_x). \tag{4.61}$$

The kernel relating the Wigner distribution of the output to that of the input, and the Wigner distribution of the output are given by

$$\check{K}_{\check{h}}(x, \sigma_x; x', \sigma_x') = \delta(s^2\sigma_x + x')\, \delta(x/s^2 - \sigma_x'), \tag{4.62}$$

$$\check{W}_{\check{g}}(x, \sigma_x) = \check{W}_{\check{f}}(-s^2\sigma_x, x/s^2). \tag{4.63}$$

Upon comparison with equations 4.12 and 4.15, equations 4.58 and 4.63 imply

$$\begin{bmatrix} \check{A} & \check{B} \\ \check{C} & \check{D} \end{bmatrix} = \begin{bmatrix} 0 & s^2 \\ -1/s^2 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -s^2 \\ 1/s^2 & 0 \end{bmatrix}^{-1}. \tag{4.64}$$

The equations relating the position and normalized angle of the output ray to those of the input ray can be written in the form

$$\begin{bmatrix} x_2 \\ \sigma_{x2} \end{bmatrix} = \begin{bmatrix} 0 & s^2 \\ -1/s^2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ \sigma_{x1} \end{bmatrix} = \begin{bmatrix} \check{A} & \check{B} \\ \check{C} & \check{D} \end{bmatrix} \begin{bmatrix} x_1 \\ \sigma_{x1} \end{bmatrix}. \tag{4.65}$$

Figure 4.5 shows how a rectangle representing either the support of the Wigner distribution or a bundles of rays, is transformed through this operation.

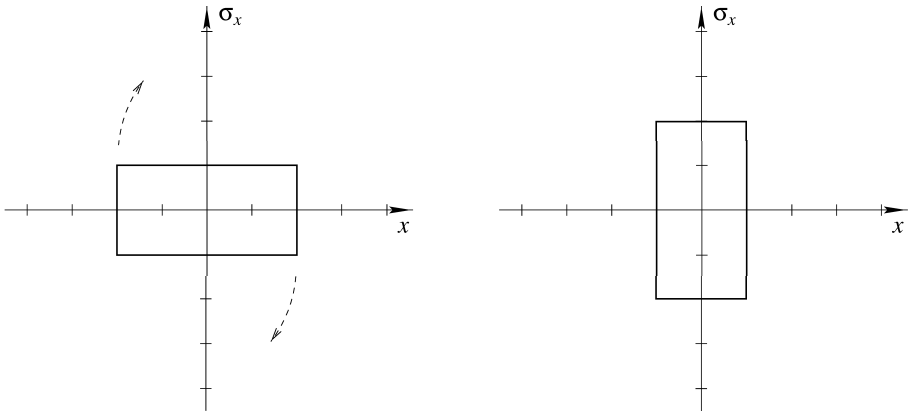


Figure 4.5: The effect of Fourier transform stages in phase space.

The eigenfunctions of propagation through these Fourier transform stages are of the form $\check{f}(x) = s^{-1/2}\psi_l(x/s)$, where $\psi_l(\cdot)$ is the $l$th order Hermite-Gaussian function. The effect of Fourier transformation on such functions is simply to multiply them with a unit-magnitude complex constant (a phase factor).

The hyperdifferential form of the Fourier transform operator is given by

$$\propto \exp\left[-i(\pi/2)\left(-\frac{s^2}{4\pi}\frac{d^2}{dx^2} + \pi\frac{x^2}{s^2}\right)\right]. \tag{4.66}$$

Since the operator corresponding to Fourier transformation is a function of $s^2(i2\pi)^{-2}d^2/dx^2 + x^2/s^2$ only, all operators of $s^2(i2\pi)^{-2}d^2/dx^2 + x^2/s^2$ will remain invariant under Fourier transformation. In other words, they will commute with the Fourier transform.

### 4.3.7 General linear systems

The effect of an arbitrary linear system

$$\check{g}(x) = \int \check{h}(x, x')\check{f}(x')\,dx' \tag{4.67}$$

can be described in phase space by the kernel $\check{K}_{\check{h}}(x, \sigma_x; x', \sigma'_x)$ as follows:

$$\check{W}_{\check{g}}(x, \sigma_x) = \iint \check{K}_{\check{h}}(x, \sigma_x; x', \sigma'_x)\,\check{W}_{\check{f}}(x', \sigma'_x)\,dx'\,d\sigma'_x, \tag{4.68}$$

$$\check{K}_{\check{h}}(x,\sigma_x;x',\sigma'_x) = \iint \check{h}(x+x''/2,x'+x'''/2)\,\check{h}^*(x-x''/2,x'-x'''/2) \times e^{-i2\pi x''\sigma_x + i2\pi x'''\sigma'_x}\,dx''\,dx''',$$

which is simply the dimensional form of equation 2.44.

### 4.3.8 Spherical reference surfaces

Since we have seen in our discussion of spherical reference surfaces in subsection 3.4.8 that the effect of employing such reference surfaces is mathematically equivalent to passage through a lens, their effect in phase space also corresponds to vertical shearing (along the frequency dimension).

### 4.3.9 Discussion

At this point we have completed a rather comprehensive and unified description of so-called *first-order optical systems*. This is simply another name for *quadratic-phase systems*, which are mathematically equivalent to the linear canonical transforms discussed in chapter 2. Free-space propagation in the Fresnel approximation, transmission through thin lenses, and propagation through quadratic graded-index media, and their arbitrary combinations fall into this class. The wealth of properties and results derived in chapter 2 for linear canonical transforms can be directly applied to these optical components and systems composed of them. For instance, we can readily translate the uncertainty relation given in equation 2.164 to an optical context:

$$\Delta x' \Delta x \geq |\check{B}|/4\pi, \tag{4.69}$$

where $x'$ denotes the coordinate of the input plane and $x$ denotes the coordinate of the output plane. For instance, specializing to free-space propagation over a distance $d$, we have $\check{B} = \lambda d$ so that the product of the spreads of the input and output light distributions cannot be less than $\lambda d/4\pi$.

When we combine first-order optical components and systems with arbitrary spatial filters (which are not first-order components), we obtain a class of systems which we will refer to as *Fourier-optical systems*.

We have discussed common optical components that will be of interest to us from both wave optics and geometrical optics perspectives, and also discussed the effects of these components in phase space, be it the wave-optical phase space of Wigner distributions, or the geometrical-optical phase space of ray bundles. We saw that the relationship of the wave and geometrical optical perspectives is particularly transparent in phase space: the support of the Wigner distribution and the region representing the ray bundle are transformed in the same manner.

At this point we have a rather broad array of tools for characterizing and analyzing first-order optical systems: their forward and inverse kernels, their input-output relations

in the space and Fourier domains, their effect on the Wigner distribution, the ray matrices (which are nothing but the matrices characterizing these systems as linear canonical transforms), the parallelogram-type geometric distortions which they effect in phase space, and their eigenfunctions, Hamiltonians $\mathcal{H}$, hyperdifferential forms, and associated differential equations. Most of these were already discussed in a purely mathematical and signal theoretical context in chapter 2. Here we have reexpressed them in an optical context and showed their relation to wave and geometrical optical concepts. The reader who has grasped these different forms in which common optical components and systems can be characterized, how to translate between them, and how to concatenate different components, can be said to have acquired a rather broad and comprehensive understanding of the theory of first-order and Fourier-optical systems. The importance of studying the parallels between the results presented in chapters 2 and 3 and the present chapter cannot be overstressed.

The $2 \times 2$ matrices characterizing first-order optical components and systems constitute one of the most efficient ways of manipulating and concatenating them. As already discussed, these matrices can be interpreted in many different ways. The following three $2 \times 1$ vectors (which are equivalent in the paraxial approximation) transform with the $\check{A}\check{B}\check{C}\check{D}$ matrix:

$$\begin{bmatrix} x \\ \sigma_x \end{bmatrix} = \begin{bmatrix} x \\ \theta_x/\lambda \end{bmatrix} = \begin{bmatrix} x \\ dx(z)/d(\lambda z) \end{bmatrix}. \tag{4.70}$$

Here $\sigma_x$ can be interpreted either as spatial frequency or normalized angle $\theta_x/\lambda$. Of course, under the paraxial approximation, $\theta_x$ is equal to $\sin\theta_x$ and $\tan\theta_x$ and thus also the slope and derivative of the ray. To this list we could also add the $2 \times 1$ vector formed by replacing the second component by *momentum*, which is the second phase-space variable most commonly encountered in classical and quantum mechanics. We also recall from chapter 2 that the coordinate multiplication and differentiation operators $[\mathcal{U}\ \mathcal{D}]^{\mathrm{T}}$ also transform according to the (dimensionless) $ABCD$ matrix. (We refrain from defining dimensional versions of these operators to avoid further proliferation of notation.) The intimate relation between the differentiation operator and spatial frequency is immediate if we recall that the effect of $\mathcal{D}$ in the frequency domain is merely to multiply by $\mu$. Alternatively, if we consider a signal of the form $\exp[i2\pi\check{\phi}(x)]$, we see that its derivative at a certain point is simply given by the signal multiplied by the instantaneous frequency $d\check{\phi}(x)/dx$ at that point.

The unity and underlying equivalence of these $2 \times 1$ vectors should already be evident from the discussions of chapters 1, 2, 3, and the present one. Any of these pairs of quantities constituting the two variables of phase space (for one-dimensional systems) are said to be *conjugate variables*. Although the first member of each pair, corresponding to time or space or the generic coordinate variable is often the "original" variable in which physical relations are first expressed, this priority disappears in phase space, and there is complete symmetry between the two conjugate variables. We also find it useful to refer to these variables as being *orthogonal* to each other. We further know that there is

also complete symmetry between variables associated with all fractional Fourier domains (variables which are not orthogonal to each other).

We will have more to say on the transformation of $2 \times 1$ vectors in section 4.7.

## 4.4 Imaging and Fourier transformation

In this section we will discuss imaging and Fourier transforming systems in the context of first-order optics (Ozaktas and Erden 1997), and also present a number of general theorems for determining image and Fourier transform planes.

### 4.4.1 Imaging systems

We begin with the historically prior case of imaging. The most general transform kernel, allowing for the possibility of a residual quadratic-phase term, may be expressed as

$$\check{h}(x,x') = Ke^{i\pi x^2/\lambda R}\sqrt{M}\,\delta(x - Mx'). \tag{4.71}$$

The system is energy preserving when the magnitude of the constant $K$ is unity. This kernel maps a function $\check{f}(x)$ to $K\exp(i\pi x^2/\lambda R)\,|M|^{-1/2}\check{f}(x/M)$. $M$ is referred to as the magnification and $R$ is the radius of the spherical surface on which the perfect image is observed. When $R = \infty$, the quadratic-phase term disappears and the perfect image is observed on a planar surface. The above kernel is a special case of equation 4.5 with $\check{\alpha}, \check{\beta}, \check{\gamma} \to \infty$ such that $\check{\gamma}/\check{\beta} = M$, $\check{\alpha} - \check{\beta}^2/\check{\gamma} = 1/\lambda R$. The matrix associated with this kernel is given by

$$\begin{bmatrix} M & 0 \\ M/\lambda R & 1/M \end{bmatrix}, \tag{4.72}$$

which the reader should compare with equation 2.116, from which it is seen that this matrix corresponds to perfect imaging (scaling) followed by chirp multiplication. Now, interpreting the above as a ray matrix, the following well recognized conditions may be easily deduced:

1. If and only if a given quadratic-phase system is an imaging system, a ray emanating from the axial point at the input plane ($x_1 = 0$) is mapped onto the axial point at the output plane ($x_2 = 0$) (figure 4.6a). Furthermore, all rays emanating from an off-axis point at the input plane will be mapped to a common point at the output plane.

2. If a ray parallel to the optical axis at the input plane ($\theta_{x1} = 0$) is also parallel to the optical axis at the output plane ($\theta_{x2} = 0$), the imaging is perfect with no residual phase curvature ($R = \infty$), and the magnification $M$ is given by the ratio of the distance of the output ray to the optical axis, to the distance of the input ray to the optical axis: $M = x_2/x_1$ (figure 4.6a).

3. If a ray parallel to the optical axis at the input plane is not parallel to the optical axis at the output plane, the magnification is found in the same way as in perfect imaging, and the radius of curvature of the residual quadratic-phase factor is simply that defined by the slope of the ray ($R = x_2/\theta_{x2}$) (figure 4.6b).

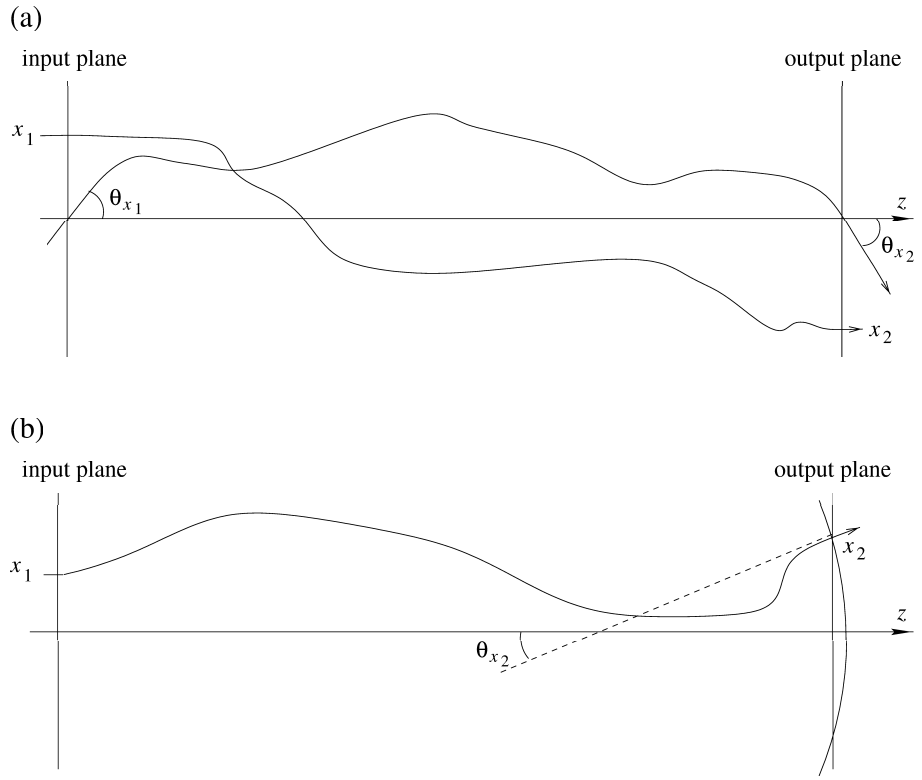


Figure 4.6: Rays in an imaging system (Ozaktas and Erden 1997).

As an example, we consider the classical single-lens imaging system which consists of a section of free space of length $d_1$ followed by a lens with focal length $f$ followed by another section of free space of length $d_2$, such that $1/f = 1/d_1 + 1/d_2$. The matrix associated with the overall system can be found easily by multiplying the matrices of its three components:

$$\begin{bmatrix} -d_2/d_1 & 0 \\ -1/\lambda f & -d_1/d_2 \end{bmatrix}, \tag{4.73}$$

from which we see that $M = -d_2/d_1$ and $R = f d_2/d_1$. We see that the single-lens imaging system does not provide a perfect image on a planar surface but results in an additional phase factor. This also means that parallel rays entering the system do not emerge parallel to each other.

Introduction of spherical reference surfaces at the input and/or output does not alter the basic imaging condition, but simply changes the overall phase factor appearing in

equation 4.71. (This is because the product of lower triangular matrices are always also lower triangular.) If we choose these input and output spherical reference surfaces in a particular manner, it is possible to eliminate the phase factor in the kernel relating the input and output with respect to these reference surfaces. To determine the radii $R_1$ and $R_2$ of the input and output spherical reference surfaces which eliminate the phase factor, we consider the following identity:

$$\begin{bmatrix} M & 0 \\ M/\lambda R & 1/M \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1/\lambda R_2 & 1 \end{bmatrix} \begin{bmatrix} M & 0 \\ 0 & 1/M \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1/\lambda R_1 & 1 \end{bmatrix} \tag{4.74}$$

where $R_1$ and $R_2$ satisfy $M/R = M/R_2 - 1/MR_1$. Here the rightmost matrix takes us from a planar reference surface to a spherical reference surface with radius $R_1$ and the matrix immediately to the right of the equal sign takes us from a spherical reference surface with radius $R_2$ to a planar reference surface. A particular choice of $R_1$ and $R_2$ which satisfies this relation is $R_1 = (R/M)(1 - 1/M)$ and $R_2 = R(1 - 1/M)$. Figure 4.7 shows these reference surfaces, between which perfect imaging with no additional phase factor is obtained.

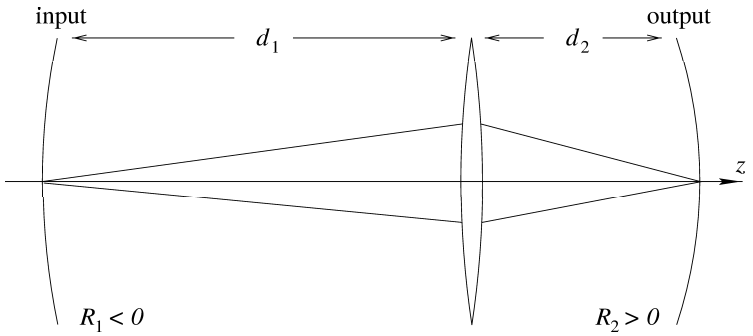


Figure 4.7: Single-lens imaging system with spherical reference surfaces. The figure is drawn for a positive lens ($f > 0$) and $d_1, d_2 > 0$ for which $R_1 < 0$ and $R_2 > 0$.

The reader may also wish to similarly analyze the telescopic imaging system which consists of a section of free space of length $f_1$, followed by a lens of focal length $f_1$, a section of free space of length $f_1 + f_2$, a lens of focal length $f_2$, and finally a section of free space of length $f_2$. This system has magnification $M = -f_2/f_1$. The image is perfect without any residual phase factor.

### 4.4.2 Fourier transforming systems

We now turn our attention to Fourier transformation. The most general transform kernel allowing for the possibility of a residual quadratic-phase term may be expressed as

$$\check{h}(x, x') = K e^{i\pi x^2/\lambda R} \sqrt{\frac{1}{s^2 M}}\, e^{-i2\pi x x'/s^2 M}. \tag{4.75}$$

The system is energy preserving when the magnitude of the constant $K$ is unity. This kernel maps a function $\check{f}(x)$ into $K\exp(i\pi x^2/\lambda R)\sqrt{1/s^2M}\,\check{F}(x/s^2M)$, where $F(\mu)$ is the Fourier transform of $f(u)$. Here $s$ is the implicit scale factor discussed in section 3.2 relating functions which take dimensionless and dimensional arguments: $s^{1/2}\check{f}(x) \equiv f(x/s) \equiv f(u)$ and $s^{-1/2}\check{F}(\sigma_x) \equiv F(s\sigma_x) \equiv F(\mu)$. $M$ is the magnification associated with the Fourier transformation and $R$ is the radius of the spherical surface on which the perfect Fourier transform is observed. When $R = \infty$, the quadratic-phase term disappears and the perfect Fourier transform is observed on a planar surface. The above kernel is a special case of equation 4.5 with $\check{\gamma} = 0$, $\check{\alpha} = 1/\lambda R$, and $\check{\beta} = 1/s^2M$. The matrix associated with the Fourier transform kernel is given by

$$\begin{bmatrix} 0 & s^2M \\ -1/s^2M & s^2M/\lambda R \end{bmatrix}, \tag{4.76}$$

which the reader should compare with equation 2.121, from which it is seen that this matrix corresponds to perfect Fourier transformation (with a certain magnification) followed by chirp multiplication. Now, interpreting the above as a ray matrix, the following well recognized conditions may be easily deduced:

1. If and only if a given quadratic-phase system is a Fourier transforming system, a ray parallel to the optical axis at the input plane ($\theta_{x1} = 0$) will pass through the axial point at the output plane ($x_2 = 0$) (figure 4.8a). Furthermore, all parallel rays making the same angle with the optical axis at the input plane, will be mapped to a common point at the output plane.

2. If a ray emanating from the axial point at the input plane ($x_1 = 0$) emerges parallel to the optical axis at the output plane ($\theta_{x2} = 0$), the Fourier transformation is perfect with no residual phase curvature ($R = \infty$), and the magnification $M$ is given by the ratio of the distance of the output ray to the optical axis, to the angle of the input ray: $M = x_2\lambda/\theta_{x1}s^2$ (figure 4.8a).

3. If a ray emanating from the axial point at the input plane does not emerge parallel to the optical axis at the output plane, the scale factor is found in the same way as in perfect Fourier transformation, and the radius of curvature of the residual quadratic-phase factor is simply that defined by the slope of the ray ($R = x_2/\theta_{x2}$) (figure 4.8b).

Overall, we see that Fourier transforming systems map ray positions into ray angles, and ray angles into ray positions. This is to be expected due to the close association between ray angles and spatial frequencies.

We have already discussed three examples of Fourier transforming systems on pages 137 and 171. All of these systems provide perfect Fourier transforms without any residual phase. A system with residual phase can be corrected by either working with spherical reference surfaces or appending thin lenses at the input or output planes.

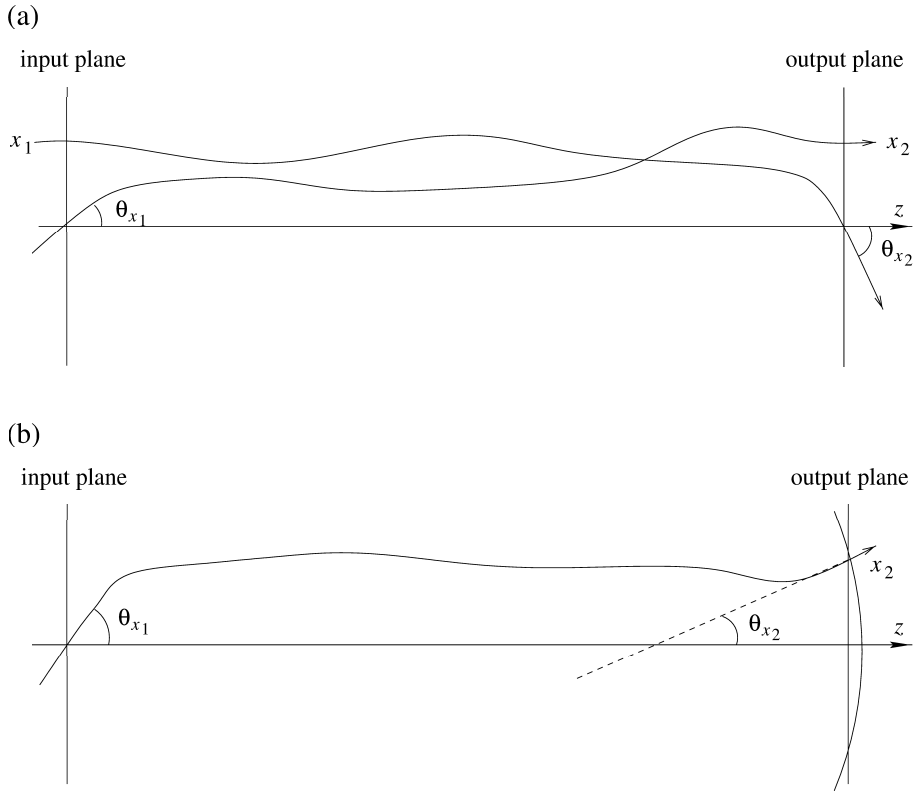


Figure 4.8: Rays in a Fourier transforming system (Ozaktas and Erden 1997).

We conclude with a result which states that any quadratic-phase system can be interpreted as a magnified Fourier transform by choosing appropriate spherical reference surfaces. This fact directly follows from the following matrix identity:

$$\begin{bmatrix} \check{A} & \check{B} \\ \check{C} & \check{D} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1/\lambda R_2 & 1 \end{bmatrix} \begin{bmatrix} 0 & M' \\ -1/M' & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1/\lambda R_1 & 1 \end{bmatrix}, \tag{4.77}$$

where $R_1 = -\check{B}/\check{A}\lambda$, $R_2 = \check{B}/\check{D}\lambda$, and $M' = \check{B}$. Alternatively, any quadratic-phase system can be realized by appending lenses with focal lengths $f_1 = R_1$ and $f_2 = -R_2$ to the input and output planes of a magnified Fourier transforming system.

### 4.4.3 General theorems for image and Fourier transform planes

We now consider a generic first-order optical system of the general form shown in figure 4.9, consisting of several thin lenses separated by arbitrary distances, an object $\check{f}(x)$, and a point light source. Although not shown in the figure, the system may also contain sections of quadratic graded-index media or more generally any system which can be characterized by a linear $\check{A}\check{B}\check{C}\check{D}$ relation. That part of the system between the point source and object

serves to illuminate the object, and that part to the right of the object forms a sequence of images and Fourier transforms. Typically, we expect to first observe a Fourier transform, followed by an inverted image, followed by an inverted Fourier transform, followed by an erect image, and so on.

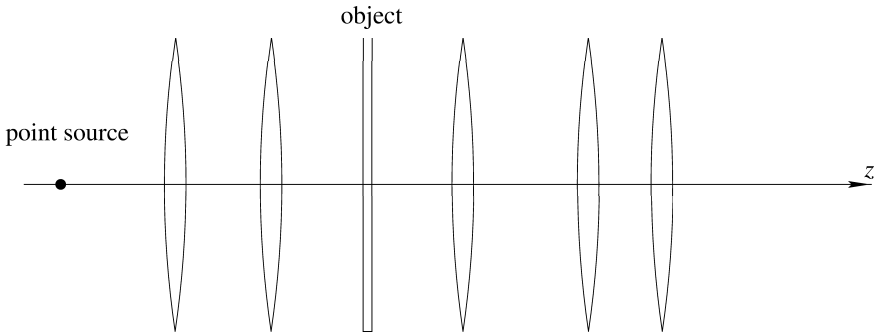


Figure 4.9: Generic first-order optical system.

The planes at which the images are observed are quite easily determined by finding those planes at which the system matrix from the object to that plane is of the form of equation 4.72. The magnifications and residual phase factors are also readily obtained from this procedure. (The image planes can also be less systematically found by consecutive application of the equation $1/f = 1/d_1 + 1/d_2$.)

Likewise, the planes at which the Fourier transforms are observed are quite easily determined by finding those planes at which the system matrix from the object to that plane is of the form of equation 4.76. The magnifications and residual phase factors are again readily obtained from this procedure. It turns out that *the planes at which the Fourier transforms are observed are precisely the planes at which the images of the point source would have been observed had there been no object.* Furthermore, *the phase factor preceding the Fourier transform is the same as that which would have been observed at the transform plane had the object been replaced with a point source.* Thus the location of the Fourier transform does not depend on the location of the object, although the residual phase factor does. This result encompasses and generalizes the many Fourier transforming configurations given in texts on Fourier optics. We now present a derivation of this very general result which is stated in several sources without proof (VanderLugt 1992, Goodman 1996).

This result is intuitively plausible in that having no object is the same as having an object whose transmittance is unity, so that the Fourier transform plane being the plane where the point source is imaged is consistent with the fact that the Fourier transform of unity is a delta function.

To derive this result, we refer to figure 4.10. Here we have lumped all components lying between the point source and object transparency into a single first-order system

characterized by $\check{A}_1$, $\check{B}_1$, $\check{C}_1$, $\check{D}_1$, and we have lumped all components lying between the object and a point to which the source is imaged into another first-order system characterized by $\check{A}_2$, $\check{B}_2$, $\check{C}_2$, $\check{D}_2$. If $P'$ is an image of $P$ when there is no object, the

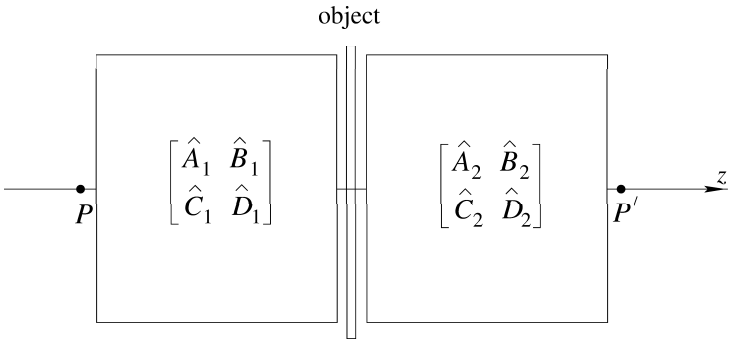


Figure 4.10: Derivation of general Fourier transforming condition. $P$ represents the point source and $P'$ one of its images had there been no object transparency.

following necessarily holds (equation 4.72)

$$\begin{bmatrix} M & 0 \\ M/\lambda R & 1/M \end{bmatrix} = \begin{bmatrix} \check{A}_2 & \check{B}_2 \\ \check{C}_2 & \check{D}_2 \end{bmatrix} \begin{bmatrix} \check{A}_1 & \check{B}_1 \\ \check{C}_1 & \check{D}_1 \end{bmatrix} \tag{4.78}$$

for some $M$ and $R$. Since the upper right element of an imaging matrix is zero, all rays emanating from point $P$, regardless of their angle, arrive at $P'$. The result we are seeking to prove states that the Fourier transform of the object is observed at $P'$ regardless of how the left hand side is decomposed into the two matrices appearing on the right hand side. The left part of the system characterized by $\check{A}_1, \check{B}_1, \check{C}_1, \check{D}_1$ merely illuminates the object transparency with a spherical wave. This is fully equivalent to putting a thin lens of appropriate focal length $f$ immediately before the object and illuminating it with a plane wave. Since the positions of the object transparency and thin lens can be interchanged without effect, this in turn is equivalent to plane wave illumination of the object transparency followed by the thin lens followed by the system characterized by $\check{A}_2, \check{B}_2, \check{C}_2, \check{D}_2$. Thus if we can show that the thin lens followed by the system characterized by $\check{A}_2, \check{B}_2, \check{C}_2, \check{D}_2$ is a Fourier transforming system, we will have demonstrated the claimed result. This amounts to showing that the matrix

$$\begin{bmatrix} \check{A}_2 & \check{B}_2 \\ \check{C}_2 & \check{D}_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1/\lambda f & 1 \end{bmatrix} \tag{4.79}$$

is in the form of equation 4.76. In order to show this, it remains to determine the focal length $f$ of the thin lens which illuminates the object with the same spherical wave which would have been generated by the point source followed by the system characterized by $\check{A}_1, \check{B}_1, \check{C}_1, \check{D}_1$. Thinking in terms of rays, the point source is represented by a bundle

of rays of the general form $[0\ \sigma_x]^{\mathrm{T}}$, and the plane wave illuminating the thin lens is represented by a bundle of rays of the general form $[x\ 0]^{\mathrm{T}}$. Thus, equating the bundle of rays resulting from a point source followed by the system $\check{A}_1, \check{B}_1, \check{C}_1, \check{D}_1$, to a plane wave followed by the lens of focal length $f$, we obtain

$$\begin{bmatrix} \check{A}_1 & \check{B}_1 \\ \check{C}_1 & \check{D}_1 \end{bmatrix} \begin{bmatrix} 0 \\ \sigma_x \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1/\lambda f & 1 \end{bmatrix} \begin{bmatrix} x \\ 0 \end{bmatrix}, \tag{4.80}$$

from which we get $f = -\check{B}_1/\lambda\check{D}_1$. This lens, when illuminated by a plane wave, produces a spherical wave whose radius of curvature is $\check{B}_1/\lambda\check{D}_1$. Now, we can evaluate the matrix product in equation 4.79 using this value of $f$:

$$\begin{bmatrix} \check{A}_2 + \check{B}_2\check{D}_1/\check{B}_1 & \check{B}_2 \\ \check{C}_2 + \check{D}_1\check{D}_2/\check{B}_1 & \check{D}_2 \end{bmatrix}. \tag{4.81}$$

Using the condition $\check{A}_2\check{B}_1 + \check{B}_2\check{D}_1 = 0$ which follows from equation 4.78, the above equation reduces to

$$\begin{bmatrix} 0 & \check{B}_2 \\ -1/\check{B}_2 & \check{D}_2 \end{bmatrix}, \tag{4.82}$$

which upon comparison with equation 4.76 is seen to be a Fourier transform matrix with magnification $M = \check{B}_2$ and residual phase radius of curvature $R = \check{B}_2/\lambda\check{D}_2$ (here we take $s = 1$ for simplicity). Thus we have shown that the Fourier transform is observed at locations where the point source is imaged and does not depend on the location of the object. The magnification and curvature, however, do depend on where the object is. We now also show that the residual phase factor observed is the same phase factor that a point source located in place of the object would result in. Again representing a point source by $[0\ \sigma_x]^{\mathrm{T}}$ and passing it through the system $\check{A}_2, \check{B}_2, \check{C}_2, \check{D}_2$, we obtain

$$\begin{bmatrix} \check{A}_2 & \check{B}_2 \\ \check{C}_2 & \check{D}_2 \end{bmatrix} \begin{bmatrix} 0 \\ \sigma_x \end{bmatrix} = \begin{bmatrix} \check{B}_2\sigma_x \\ \check{D}_2\sigma_x \end{bmatrix}. \tag{4.83}$$

The right hand side corresponds to a bundle of rays representing a spherical wave with radius of curvature $\check{B}_2/\lambda\check{D}_2$, which is the same as the radius of curvature of the residual phase factor accompanying the Fourier transform.

Naturally, it is also possible to think of the above derivation in terms of direct illumination of an object by a diverging or converging spherical wave, rather than a point source followed by illumination optics. (Such illumination can of course be obtained by a plane wave followed by a thin lens.) In this case, we would say that the Fourier transform is observed at the plane where the illuminating spherical wave would have come to focus, had the object transparency been removed. We finally note that the results also hold for virtual sources and when the point source is at $-\infty$ (corresponding to plane wave illumination).

We now turn our attention to another closely related result. But first, we must very briefly introduce the *aperture stop*, and the *entrance pupil* and *exit pupil* of an optical

system, referring the reader to textbooks such as Jenkins and White 1976 and Born and Wolf 1980 for a more complete explanation. These concepts are most easily understood in terms of a centered rotationally symmetric optical system consisting of many sections of free space, lenses, and apertures. The aperture stop is the limiting aperture in such an optical system: It is the aperture which limits the solid angle of the fan or cone of rays (the ray bundle) which can pass through the system. Increasing the diameter of any aperture other than the aperture stop will not increase the solid angle of the cone of rays which can pass through the system. Increasing the diameter of the aperture stop, on the other hand, will let a greater number of rays pass. (However, if the diameter is increased too much, some other aperture may start limiting the solid angle and become the aperture stop.) The exit pupil and the entrance pupil are the images of the aperture stop with respect to the image and object sides. That is, the image of the stop formed by everything to the right of the stop is the exit pupil, and the image of the stop formed by everything to the left of the stop is the entrance pupil. Alternatively, the exit pupil is where the stop seems to be when we look from the image side and the entrance pupil is where the stop seems to be when we look from the object side. Ray matrices can be used to determine their position and size. Since we cannot offer a more complete discussion, an example should help make these concepts more concrete. Figure 4.11 shows an imaging system consisting of two lenses and an aperture stop located between the lenses (the lenses themselves are assumed to have large apertures). The entrance and exit pupils are shown.

Now, consider an imaging system which under ideal conditions (no apertures) would map an on-axis delta function object to an on-axis delta function image. In reality there will always be a limiting aperture (the aperture stop) determining the spatial frequency response of the system. If the object is an on-axis delta function, the Fourier transform of the aperture stop will be observed at the output plane. To see this, we again refer to figure 4.10, this time interpreting $P$ as an on-axis point object, $P'$ as its ideal on-axis image, and the object in that figure as an aperture stop. It now follows from the analysis associated with figure 4.10 that we observe the Fourier transform of the aperture stop at the image plane.

We can now state the result we wish to demonstrate: *The amplitude distribution observed at the image plane will be the same as that which would have been observed had the exit pupil been illuminated with a spherical wave converging towards the ideal point image through free space. This observed amplitude distribution is a scaled version of the Fourier transform of the exit pupil.* The latter part of this statement is consistent with the result stated in the previous paragraph because the exit pupil is simply an image of the aperture stop so that their Fourier transforms are simply scaled versions of each other.

We now proceed to demonstrate the stated result. The left part of the system characterized by $\check{A}_1, \check{B}_1, \check{C}_1, \check{D}_1$ merely serves to illuminate the aperture stop with a spherical wave of radius $\check{B}_1/\lambda\check{D}_1$. Thus, the result stated in the preceding paragraph is equivalent to the following claim: The second part of the system consisting of the aperture stop illuminated by a spherical wave followed by the system characterized by $\check{A}_2, \check{B}_2, \check{C}_2, \check{D}_2$,

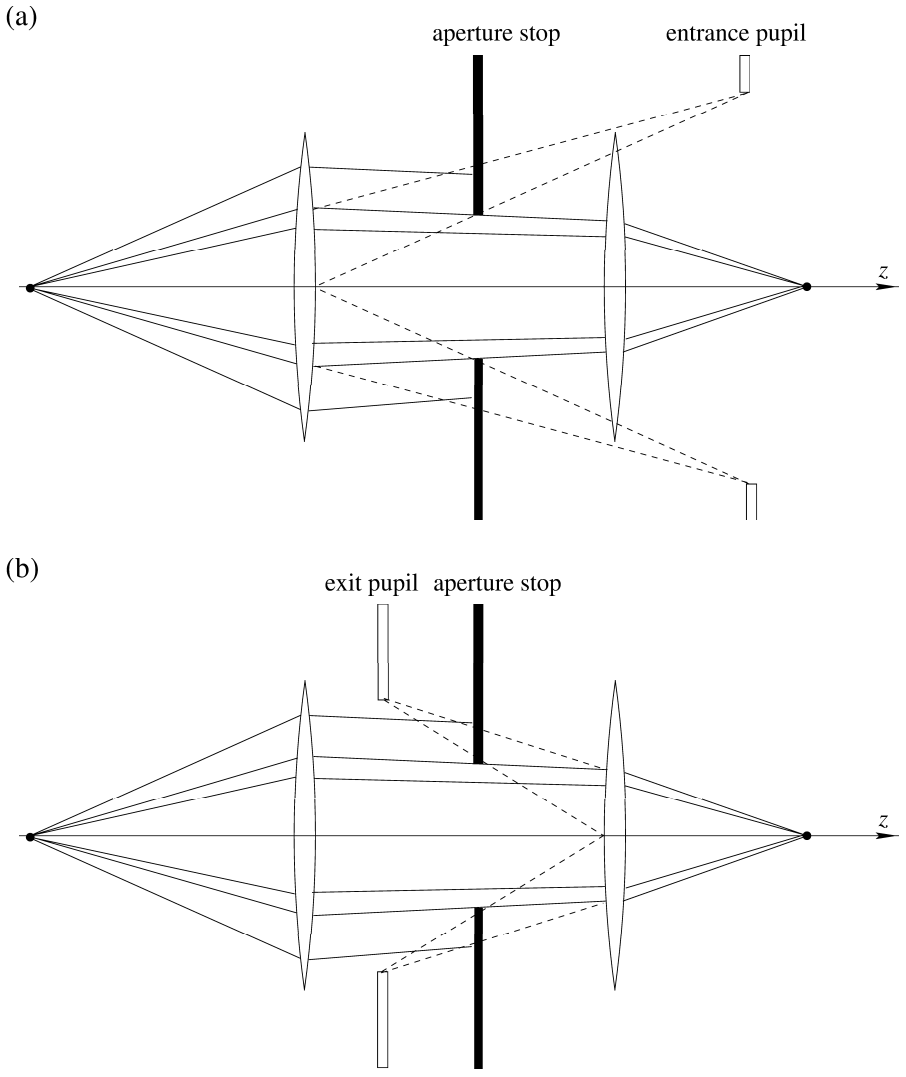


Figure 4.11: Entrance (a) and exit (b) pupils.

produces the same amplitude distribution as, the exit pupil illuminated by a spherical wave of radius $-d_{\text{exit}}$ followed by a section of free space of length $d_{\text{exit}}$ (figure 4.12). Here $d_{\text{exit}}$ is the distance of the exit pupil from the image plane. First, we write the equation

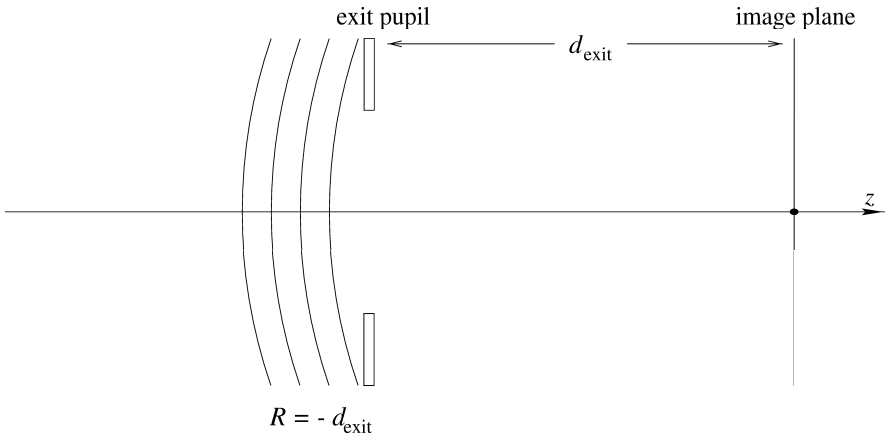


Figure 4.12: The exit pupil illuminated by a converging spherical wave.

which enables us to determine the size and location of the exit pupil:

$$\begin{bmatrix} 1 & -\lambda d_{\text{exit}} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \check{A}_2 & \check{B}_2 \\ \check{C}_2 & \check{D}_2 \end{bmatrix} = \begin{bmatrix} M_{\text{exit}} & 0 \\ M_{\text{exit}}/\lambda R_{\text{exit}} & 1/M_{\text{exit}} \end{bmatrix}. \tag{4.84}$$

The left hand side of this equation consists of the matrix associated with that part of the system falling to the right of the aperture stop, followed by free-space propagation over a distance $-d_{\text{exit}}$. (Here the minus sign reflects the fact that $d_{\text{exit}}$ is measured towards the left, so that if $d_{\text{exit}} > 0$ the location of the exit pupil is to the left of the image plane.) The right hand side of this equation is of the form of equation 4.72 and reflects the fact that the exit pupil is the image of the aperture stop as seen from the image side. By solving this equation we can determine the magnification $M_{\text{exit}}$ between the exit pupil and the aperture stop as $1/\check{D}_2$, and the distance of the exit pupil towards the left of the image plane as $d_{\text{exit}} = \check{B}_2/\lambda\check{D}_2$. Now, we can demonstrate the claimed equivalence result through the following identity which the reader should be able to verify:

$$\begin{bmatrix} \check{A}_2 & \check{B}_2 \\ \check{C}_2 & \check{D}_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \check{D}_1/\check{B}_1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & \lambda d_{\text{exit}} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1/\lambda d_{\text{exit}} & 1 \end{bmatrix} \begin{bmatrix} M_{\text{exit}} & 0 \\ 0 & 1/M_{\text{exit}} \end{bmatrix}. \tag{4.85}$$

The second term on the left hand side corresponds to illumination of the aperture stop by a spherical wave with radius $\check{B}_1/\lambda\check{D}_1$, which is equal to $-\check{B}_2/\lambda\check{A}_2$ by virtue of the imaging condition equation 4.78. The last term on the right hand side reflects the fact that the exit pupil is a magnified image of the aperture stop. The central term on the right hand side represents a spherical wave converging towards the on-axis image point, and the first

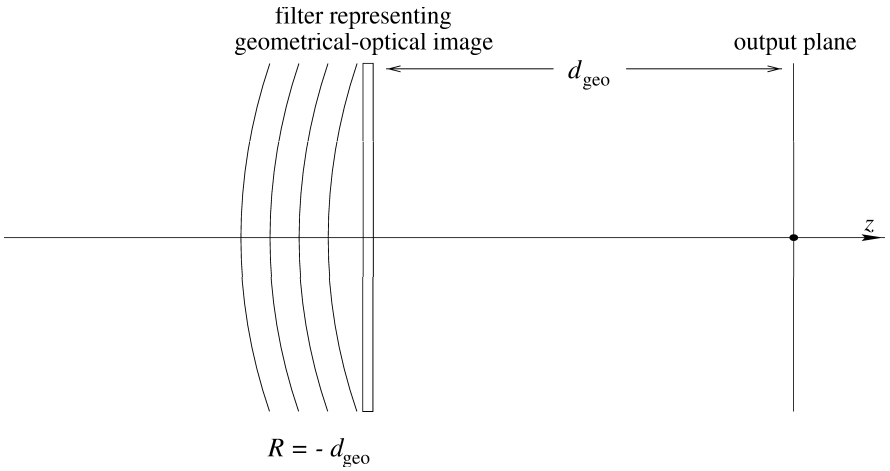


Figure 4.13: The input of the system is the geometrical-optical image of the original input object as seen from the output side, and is situated at the same location where this image is observed. It is illuminated by a spherical wave converging towards the axis at the output plane.

term corresponds to free-space propagation over a distance $d_{\text{exit}}$. Thus the two sides of the above identity represent the two systems whose equivalence we were seeking to prove.

The above discussion has been carried out for an on-axis object and image point. If it can be assumed that the images corresponding to off-axis object points are simply shifted versions of the Fourier transform of the exit pupil, we are then led to a simple space-invariant model for the whole system with the impulse response being given by the Fourier transform of the exit pupil. Unfortunately, the system is often not space-invariant so that this is not possible. (See the discussion following equation 3.159.)

We will conclude this section by considering the demonstrated result from a different perspective. Let us assume $\check{f}(x)$ (or $\check{f}(x,y)$) is the input of an optical system characterized by $\check{A}_2, \check{B}_2, \check{C}_2, \check{D}_2$. We assume $\check{f}(x)$ is illuminated with a spherical wave which, in the absence of $\check{f}(x)$, would have been focused to the axis at the output plane. The output of the system is then the same as that resulting from the system shown in figure 4.13. In the figure the input is the geometrical-optical image of the original input object $\check{f}(x)$ as seen from the output side, illuminated with a spherical wave converging towards the axis at the output plane. This result is simply a more general form of the previous result with an arbitrary function replacing the aperture stop, and thus the matrix algebra behind this result is the same. The significance of this result is that it allows a separation of geometrical-optical effects from diffraction effects. By replacing the original $\check{f}(x)$ by its geometrical-optical image, and the system characterized by $\check{A}_2, \check{B}_2, \check{C}_2, \check{D}_2$ by free space, we separate out the geometric effect, leaving us with the diffraction effect in the pure form represented by figure 4.13.

A complementary discussion of some of these issues may be found in Goodman 1996.

## 4.5 Decompositions and duality in optics

In subsection 2.4.4 we discussed how general linear canonical transforms (quadratic-phase systems) could be expressed in terms of each other or decomposed into simpler component systems. These decompositions can be interpreted in terms of matrices, abstract operators, operations such as chirp multiplications and convolutions, integral transforms, and geometric distortions. Now we are in a position to see that they can also be interpreted in terms of optical components such as sections of free space, thin lenses, quadratic graded-index media, and imaging and Fourier transforming systems. Such decompositions of optical systems have many applications. Two particularly important decompositions are those showing how any quadratic-phase system can be realized by using only two sections of free space and a thin lens (equation 2.122) or two thin lenses and a section of free space (equation 2.123).

As an example, consider the decompositions

$$\begin{aligned}\begin{bmatrix} 0 & \lambda f \\ -1/\lambda f & 0 \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ -1/\lambda f & 1 \end{bmatrix} \begin{bmatrix} 1 & \lambda d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1/\lambda f & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & \lambda d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1/\lambda f & 1 \end{bmatrix} \begin{bmatrix} 1 & \lambda d \\ 0 & 1 \end{bmatrix}, \end{aligned} \tag{4.86}$$

with $f = d$, which are special cases of equations 2.123 and 2.122. These decompositions show how an optical Fourier transform can be realized either as a lens followed by free space followed by another lens, or free space followed by a lens followed by free space. These are nothing but the Fourier transform stages (ii) and (i) we had introduced on page 137. These decompositions can also be cast in abstract operator notation if desired (equations 2.131 and 2.130). The optical fractional Fourier transforms can also be realized by similar decompositions but with $f \neq d$.

As another example, let us consider the simulation of a section of free space of negative length $-d$, where $d > 0$. Using

$$\begin{bmatrix} 1 & -\lambda d \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -Ms^2 \\ 1/Ms^2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1/\lambda f & 1 \end{bmatrix} \begin{bmatrix} 0 & Ms^2 \\ -1/Ms^2 & 0 \end{bmatrix}, \tag{4.87}$$

with the condition $\lambda^2 f d = -s^4 M^2$, we see that a negative section of free space can be simulated with two Fourier transform stages and a negative lens. This configuration can also be useful for simulating an inconveniently long section of free space. Alternatively, a lens of given focal length we happen not to have at hand, can be simulated by using a section of free space of appropriate length and two Fourier transform stages. A zoom lens can likewise be simulated by employing a variable section of free space.

In an optical context, equation 2.139 (or 2.140) means that any quadratic-phase system can be interpreted as a magnified fractional (or ordinary) Fourier transform by choosing appropriate spherical reference surfaces. An alternative interpretation is that any quadratic-phase system can be implemented by appending lenses at the input and output surfaces of a fractional (or ordinary) Fourier transformer.

More generally, equation 2.138 means that any quadratic-phase system can be interpreted as a magnified version of another simply by choosing appropriate spherical reference surfaces. A more general treatment of the synthesis of quadratic-phase systems, also covering anamorphic two-dimensional systems, is Sahin, Ozaktas, and Mendlovic 1998.

The decompositions under consideration can also be used to discuss a number of traditional concepts in geometrical optics. Considering an imaging system, the second focal plane is defined as the plane where a ray coming from the left parallel to the optical axis intersects the optical axis (figure 4.14). Likewise, the first focal plane is the plane where a ray coming from the right parallel to the optical axis intersects the optical axis. The

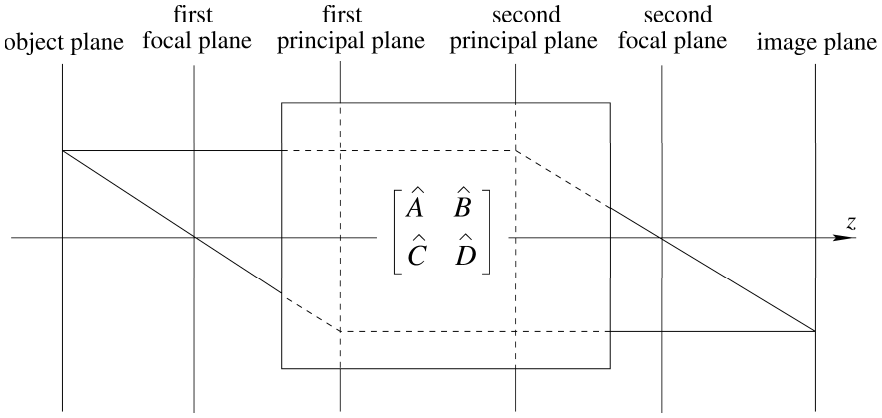


Figure 4.14: Principal and focal planes.

second principal plane is that plane where the ray coming from the left seems to bend towards the second focal point. In reality, when we have a complex optical system consisting of several components, the ray will actually undergo several bends before finally heading for the second focal point. Thus the figure does not show the actual trajectory of the ray, but a fictitious trajectory corresponding to the total refractive effect being concentrated at a single plane, which we refer to as the second principal plane. The first principal plane is likewise defined for a ray coming from the right. The first and second focal lengths $f_1$ and $f_2$ are measured from the principal planes to the focal planes and are equal in magnitude when the refractive indices of the object and image spaces are equal. A third pair of planes called the first and second nodal planes are also often introduced. However, we will not separately define them since they coincide with the principal planes when the refractive indices of the object and image spaces are equal. The rationale behind these definitions

is that they enable us to treat a complex optical system consisting of many components as if it were a thin lens. The first principal plane plays the role of the input plane of the thin lens and the second principal plane plays the role of the output plane of the thin lens. The light distribution or bundle of rays incident on the first principal plane emerges from the second principal plane just as if these planes were the input and output faces of a thin lens. In other words, the matrix relating the ray vector at the second principal plane to that at the first principal plane is of lower triangular form. With reference to the optical system characterized by $\check{A}$, $\check{B}$, $\check{C}$, $\check{D}$ shown in figure 4.15, let us consider the following

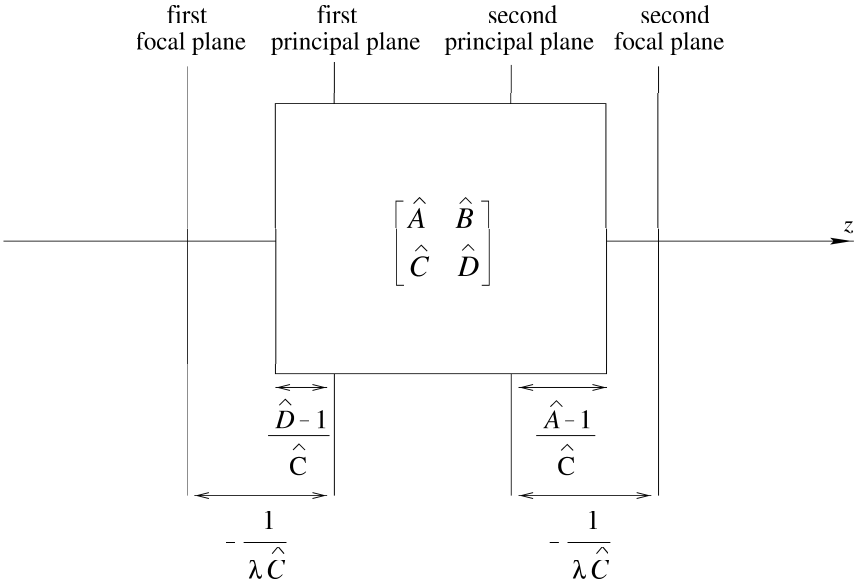


Figure 4.15: Determination of principal and focal planes. The figure is drawn for the case $\check{C} < 0$, $(1-\check{A})/\check{C} < 0$ and $(1-\check{D})/\check{C} < 0$.

decomposition:

$$\begin{bmatrix} 1 & 0 \\ \check{C} & 1 \end{bmatrix} = \begin{bmatrix} 1 & (1-\check{A})/\check{C} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \check{A} & \check{B} \\ \check{C} & \check{D} \end{bmatrix} \begin{bmatrix} 1 & (1-\check{D})/\check{C} \\ 0 & 1 \end{bmatrix}. \tag{4.88}$$

The left hand side of this equation is the matrix from the first principal plane to the second principal plane, and indeed corresponds to a thin lens with focal length $f = -1/\lambda\check{C}$. The two matrices surrounding the original system matrix on the right hand side, represent sections of free space and thus tell us the locations of the principal planes with respect to the actual physical input and output planes of the original system.

Here we also mention the concept of *dual* optical systems (Lohmann 1954, Papoulis 1968). The concept of duality was already introduced in equations 2.142 and 2.143. In an optical context, a thin lens of focal length $f$ is the dual of a section of free space of length $d$ provided $s^4 = \lambda^2 f d$, where $s$ is some implicit scale factor with the dimension of length.

This means that the action of one is the same as the action of the other under a Fourier transform with this implicit scale factor $s$:

$$\begin{bmatrix} 1 & 0 \\ -1/\lambda f & 1 \end{bmatrix} = \begin{bmatrix} 0 & -s^2 \\ 1/s^2 & 0 \end{bmatrix} \begin{bmatrix} 1 & \lambda d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & s^2 \\ -1/s^2 & 0 \end{bmatrix}. \tag{4.89}$$

The dual of a section of quadratic graded-index medium is again a section of quadratic graded-index medium. The dual of a system consisting of several components is easily found by replacing each component by its dual, with the value of $s$ fixed. The wavefields $\check{f}(x)$ are replaced by their Fourier transforms $\check{F}(x/s^2)$. Rays are replaced by dual rays: if a ray is represented by the vector $[x\ \sigma_x]^{\mathrm{T}}$, the dual of this ray is represented by $[s^2\sigma_x\ x/s^2]^{\mathrm{T}}$. If an optical system maps a given input $\check{f}(x)$ into an output $\check{g}(x)$, then the dual optical system will map the input $\check{F}(x/s^2)$ into $\check{G}(x/s^2)$. Likewise, if an optical system maps an input ray into an output ray, the dual optical system will map the dual of the original input ray into the dual of the original output ray.

In closing this section, we mention a closely related approach to the study of first-order optical systems known as *operator optics* or *operator algebra*. A brief exposition may be found in Goodman 1996, or the reader may wish to consult some of the original references, of which we provide a selection: Butterweck 1977, 1981; Shamir 1979; Nazarathy and Shamir 1980, 1981, 1982a, b; Stoler 1981; Nazarathy, Hardy, and Shamir 1982, 1986; Ojeda-Castañeda and Noyola-Isgleas 1988; Ruiz and Rabal 1997. This approach has also been generalized to deal with higher-order systems. A useful introductory exposition to the use of operators and the applications of linear canonical transforms and group theoretical methods in optics may be found in Seger and Lenz 1992.

The operators employed in such approaches are simply dimensional forms of the operators $\mathcal{Q}_q$, $\mathcal{R}_r$, $\mathcal{F}$, $\mathcal{M}$ and the like which appear in this book, and are used to represent optical components such as thin lenses and sections of free space. Once a list of commutation relations and formulas are tabulated, one can easily manipulate concatenations of operators representing components, or decompose complex systems into elementary components. In this book we preferred to use matrices to represent these operators since the ease with which these matrices are multiplied allows us to dispense with a pretabulated list of commutation relations and formulas. (Incidentally, this is also the reason why Baker-Campbell-Hausdorff type relations have not been employed in this book—these relations are not transparent and directly verifiable (see equation 1.245 and the accompanying discussion). Nevertheless, the operator method has its place as an elegant and useful formulation of Fourier optics and interested readers are especially recommended to study the works of Nazarathy and Shamir cited above.

## 4.6 Relations between wave and geometrical optics

In this section we will present several results which will serve to further highlight the parallels between wave-optics and geometrical-optics perspectives for first-order systems.

### 4.6.1 Phase of the system kernel and Hamilton's point characteristic

This subsection assumes slight familiarity with Hamilton's characteristic functions (Born and Wolf 1980).

First, we begin by showing how the phase of the kernel transforming the wavefields can be related to geometrical optical concepts. Let the kernel of a general linear system be expressed as

$$\check{h}(x, x') = |\check{h}(x, x')| e^{i2\pi \check{V}(x,x')}, \tag{4.90}$$

where the magnitude and phase are both assumed to be slowly varying. Approximating the phase as

$$\check{V}(x, x') \approx \check{V}(0, 0) + \frac{\partial \check{V}}{\partial x}\, x + \frac{\partial \check{V}}{\partial x'}\, x', \tag{4.91}$$

and taking the magnitudes out of the integral in equation 4.68, we can explicitly calculate the kernel transforming the Wigner distributions as (Bastiaans 1979b)

$$\check{K}_{\check{h}}(x, \sigma_x; x', \sigma'_x) = |\check{h}(x, x')|^2\, \delta[\sigma_x - \partial \check{V}(x, x')/\partial x]\, \delta[\sigma'_x + \partial \check{V}(x, x')/\partial x']. \tag{4.92}$$

These delta functions imply a pointwise mapping of the values of the Wigner distribution in phase space. Which phase-space point is mapped to which phase-space point is governed by the relations

$$\sigma_x = \frac{\partial \check{V}(x, x')}{\partial x}, \qquad \sigma'_x = -\frac{\partial \check{V}(x, x')}{\partial x'}. \tag{4.93}$$

Recalling that $\sigma_x$ can be interpreted as a normalized ray angle, the above equations for $\check{V}(x, x')$ can be recognized as those satisfied by *Hamilton's point characteristic* (Bastiaans 1979b). The point characteristic gives the optical path length from the input point $x'$ to the output point $x$, and is known to satisfy equations 4.93 (Born and Wolf 1980). (To be precise, we note that the quantity $\check{V}(x, x')$ appearing here is actually Hamilton's point characteristic normalized by the wavelength of light in free space.) Thus it may be concluded that the wave-optical phase $\check{V}(x, x')$ corresponds to the geometrical-optical entity known as Hamilton's point characteristic. The reader may also wish to relate this discussion to the eikonal (page 142) and Hamilton's equations (page 143).

Similar derivations are also possible for the other three representations of the system kernel: $\check{h}_{\text{spac}\to\text{freq}}(\sigma_x, x')$, $\check{h}_{\text{freq}\to\text{spac}}(x, \sigma'_x)$, and $\check{h}_{\text{freq}\to\text{freq}}(\sigma_x, \sigma'_x)$. (For instance, $\check{h}_{\text{spac}\to\text{freq}}(\sigma_x, x')$ relates the Fourier transform of the output to the input: $\check{G}(\sigma_x) = \int \check{h}_{\text{spac}\to\text{freq}}(\sigma_x, x') \check{f}(x')\, dx'$.) Bastiaans (1979b) shows that these likewise correspond to Hamilton's angle and mixed characteristics.

We will now derive Hamilton's point characteristic for a first-order optical system in terms of the $\check{A}\check{B}\check{C}\check{D}$ parameters. Since this discussion is more readily interpreted from a geometrical-optics perspective, we will use $[x_1\ \sigma_{x1}]^{\mathrm{T}}$ and $[x_2\ \sigma_{x2}]^{\mathrm{T}}$, rather than primed quantities, to denote the input and output ray vectors respectively. First-order systems can be characterized by a linear relation between the input and output ray vectors:

$$x_2 = \check{A} x_1 + \check{B} \sigma_{x1},$$

$$\sigma_{x2} = \check{C}x_1 + \check{D}\sigma_{x1}, \tag{4.94}$$

which can be solved to yield

$$\begin{aligned} \sigma_{x1} &= \frac{1}{\check{B}}\left(-\check{A}x_1 + x_2\right), \\ \sigma_{x2} &= \frac{1}{\check{B}}\left(-(\check{A}\check{D} - \check{B}\check{C})x_1 + \check{D}x_2\right). \end{aligned} \tag{4.95}$$

Now, using these together with equations 4.93 which are known to be satisfied by the point characteristic, we can write

$$\begin{aligned} \frac{\partial \check{V}(x_2, x_1)}{\partial x_1} &= \frac{1}{\check{B}}\left(\check{A}x_1 - x_2\right), \\ \frac{\partial \check{V}(x_2, x_1)}{\partial x_2} &= \frac{1}{\check{B}}\left(-(\check{A}\check{D} - \check{B}\check{C})x_1 + \check{D}x_2\right). \end{aligned} \tag{4.96}$$

Integrating these with respect to $x_1$ and $x_2$ respectively, we obtain

$$\begin{aligned} \check{V}(x_2, x_1) &= \frac{1}{\check{B}}\left(\check{A}x_1^2/2 - x_2x_1\right) + \text{term independent of } x_1, \\ \check{V}(x_2, x_1) &= \frac{1}{\check{B}}\left(\check{D}x_2^2/2 - (\check{A}\check{D} - \check{B}\check{C})x_2x_1\right) + \text{term independent of } x_2. \end{aligned} \tag{4.97}$$

Consistency of these equations requires $\check{A}\check{D} - \check{B}\check{C} = 1$ and

$$\check{V}(x_2, x_1) = \frac{1}{2\check{B}}\left(\check{D}x_2^2 - 2x_2x_1 + \check{A}x_1^2\right) + \text{term independent of } x_1 \text{ and } x_2. \tag{4.98}$$

Thus we have shown that the point characteristic of an optical system characterized by a linear $\check{A}\check{B}\check{C}\check{D}$ relation is of the above quadratic form. Referring back to equation 4.8, we see that $2\pi\check{V}(x_2, x_1)$ is identical to the quadratic form appearing as the phase of that equation. This does not surprise us since the derivation starting with equation 4.90 showed that the phase of the transform kernel corresponds to Hamilton's point characteristic. Also note that the quadratic form of $\check{V}$ is consistent with the approximation made in equation 4.91: neglecting second- and higher-order derivatives of $\check{V}$ amounts to assuming it can be expressed as a quadratic function.

To consolidate, first-order systems are characterized by linear relations between input and output ray vectors, or equivalently, linear distortions in phase space: The output ray intercepts and angles are related linearly to the input ray intercepts and angles, and the effect in phase space is a parallelogram-type distortion. These linear relations are a consequence of the paraxial and other simplifying approximations. In terms of wave kernels, first-order systems are characterized by quadratic terms in the phase. It is assumed that higher-order terms can be neglected.

We conclude this subsection by pointing out the work of Alonso and Forbes (1995) which discusses a generalization of Hamilton's formalism for geometrical optics.

### 4.6.2 Transport equations for the Wigner distribution

We now turn our attention to *transport equations* for the Wigner distribution (Bastiaans 1979c, d, 1997), which are differential equations describing how the Wigner distribution $\check{W}_{\check{f}_z(x)}(x, \sigma_x)$ of the transverse field $\check{f}_z(x)$ evolves with $z$. We first repeat equation 3.52 for a single transverse dimension:

$$\frac{\partial \check{f}_z(x)}{\partial z} = i\sqrt{4\pi^2\sigma^2 + \frac{\partial^2}{\partial x^2}}\ \check{f}_z(x), \tag{4.99}$$

which can be rearranged to read

$$\frac{\partial \check{f}_z(x)}{\partial z} = i2\pi\sqrt{\sigma^2 - \left(\frac{1}{i2\pi}\frac{\partial}{\partial x}\right)^2}\ \check{f}_z(x), \tag{4.100}$$

which in turn can be expressed in terms of the Hamiltonian given in equation 3.125 with $\sigma = 1/\lambda = n(x, z)f_{\mathrm{oc}}/c$ as

$$\frac{-1}{i2\pi}\frac{\partial \check{f}_z(x)}{\partial z} = \check{H}(x, (i2\pi)^{-1}\partial/\partial x; z)\ \check{f}_z(x), \tag{4.101}$$

where $\check{H}(x, (i2\pi)^{-1}\partial/\partial x; z)$ is a differential operator. We are not surprised to see $(i2\pi)^{-1}\partial/\partial x$ in place of $\sigma_x$, since we know that the operator version of $\sigma_x$, corresponding to multiplication by $\sigma_x$ in the frequency domain, corresponds to a derivative with respect to $x$ in the space domain. Now, if $\check{f}_z(x)$ evolves according to an equation of this general form, it is possible to show that $\check{W}_{\check{f}_z(x)}(x, \sigma_x)$ evolves according to the equation (Bastiaans 1997)

$$\frac{\partial \check{W}_{\check{f}_z(x)}(x, \sigma_x)}{\partial z} = \frac{\partial \check{H}}{\partial x}\frac{\partial \check{W}_{\check{f}_z(x)}}{\partial \sigma_x} - \frac{\partial \check{H}}{\partial \sigma_x}\frac{\partial \check{W}_{\check{f}_z(x)}}{\partial x}. \tag{4.102}$$

(This equation can also be put in the form $\partial \check{W}_{\check{f}_z(x)}/\partial z = [\cdots]\check{W}_{\check{f}_z(x)}$ if desired.) Here $\check{f}_z(x)$ is the amplitude distribution of light at a plane intersecting the optical axis at $z$, and $\check{W}_{\check{f}_z(x)}$ is the Wigner distribution of $\check{f}_z(x)$. The solution of this equation tells us how the Wigner distribution of the transverse light distribution propagates along the optical axis $z$. (We already knew how the Wigner distribution propagates as a function of $z$ in a quadratic-phase system, such as free space or quadratic graded-index media. It simply undergoes a shearing, rotation, or more general linear distortion. The present formulation is more general and is valid for arbitrary weakly inhomogeneous refractive index distributions.)

Now, let us consider the total derivative of $\check{W}_{\check{f}_z(x)}(x, \sigma_x)$ with respect to $z$:

$$\frac{d\check{W}_{\check{f}_z(x)}}{dz} = \frac{\partial \check{W}_{\check{f}_z(x)}}{\partial x}\frac{dx}{dz} + \frac{\partial \check{W}_{\check{f}_z(x)}}{\partial \sigma_x}\frac{d\sigma_x}{dz} + \frac{\partial \check{W}_{\check{f}_z(x)}}{\partial z}. \tag{4.103}$$

The partial derivative of $\check{W}_{\check{f}_z(x)}$ with respect to $z$ tells us how the value of $\check{W}_{\check{f}_z(x)}$ changes with increasing $z$ for a given $(x, \sigma_x)$. The total derivative, on the other hand, tells us the

total change in $\check{W}_{\check{f}_z(x)}$ when $x(z)$ and $\sigma_x(z)$ are also varying with $z$. If $x(z)$ and $\sigma_x(z)$ are governed by Hamilton's equations (which are equivalent to the ray equation), the total derivative will tell us the change in $\check{W}_{\check{f}_z(x)}$ along an optical ray. Now, using Hamilton's equations $dx/dz = \partial\check{H}/\partial\sigma_x$ and $d\sigma_x/dz = -\partial\check{H}/\partial x$ and equation 4.102 we obtain

$$\frac{d\check{W}_{\check{f}_z(x)}}{dz} = 0. \tag{4.104}$$

*The value of the Wigner distribution remains constant along a ray.* (The same result can also be written in the form $d\check{W}_{\check{f}_z(x)}/ds = 0$, where $ds$ is a differential element along a ray which satisfies the ray equation.) Referring the reader to Bastiaans 1979c, d for details and a precise statement of the conditions under which this result holds, and Bastiaans 1997 for a complementary discussion, here we focus on the interpretation of this result. For instance, let us consider a particular phase-space point $(x_1, \sigma_{x1})$ at the input plane, which corresponds to a ray emanating from the point $x_1$ with normalized angle $\sigma_{x1}$. The value of the Wigner distribution along all points along the ray will be the same. Alternatively, if we want to determine the value of the Wigner distribution at a certain phase-space point $(x_2, \sigma_{x2})$ at the output plane corresponding to a ray incident at the point $x_2$ with normalized angle $\sigma_{x2}$, all we need to do is to trace back this ray to the input plane and determine the value of the Wigner distribution at the phase-space point the ray originated from (Bastiaans 1979c). In other words, solving equation 4.102 is equivalent to solving the ray equation, since if we know how the rays propagate, we can also determine the Wigner distribution at any plane.

The kernel $\check{K}_{\check{h}}(x, \sigma_x; x', \sigma'_x)$ can be interpreted as the response of the system to a distribution of light whose Wigner distribution is $\delta(x-x')\,\delta(\sigma_x-\sigma'_x)$, which can be thought of as representing a single ray (Bastiaans 1978). There does not exist a distribution of light with such a Wigner distribution. However, such a fictitious Wigner distribution represents the same idealization and approximation as the concept of a ray of light. For reasons that should by now be obvious, Bastiaans refers to $\check{K}_{\check{h}}$ as the *ray spread function*, as opposed to the *point spread function* $\check{h}(x, x')$.

As another viewpoint, we may consider the value of the Wigner distribution at the phase-space point $(x, \sigma_x)$ as representing the intensity or weight of the ray passing through the point $x$ with normalized angle $\sigma_x$. In a lossless system, each ray belonging to a bundle preserves its intensity or weight as it propagates through an inhomogeneous medium. Then, the Wigner distribution function represents the intensities or weights associated with the whole bundle of rays. As light propagates through such a medium, the values of the Wigner distribution at given phase-space points are mapped in one-to-one fashion to other phase-space points, precisely according to the mapping of the ray vectors. In yet other words, the effect of propagation through this medium on the Wigner distribution is a distortion which perfectly parallels the distortion of the bundle of rays representing the same light distribution. We already knew this to be the case for quadratic-phase systems (recall that the ray matrix was found identical to the matrix governing the distortion of

the Wigner distribution as it passes through a quadratic-phase system). We now see that this result is more generally true, and have further confirmation for the analogy between the support of the Wigner distribution and the region representing the bundle of rays.

It is also possible to relate our discussion to the local spatial spectrum of a light distribution $\check{f}(x)$. We had already discussed the global spatial spectrum of $\check{f}(x)$ in chapter 3, seeing that an arbitrary wavefield can be expressed as the superposition of plane waves with different amplitudes and directions. In the local spatial spectrum, we concentrate on a certain small region and examine the spatial spectrum within this region. Mathematically, this is precisely what space-frequency representations such as the windowed Fourier transform or the Wigner distribution provide us. The value of the Wigner distribution at a certain phase-space point $(x, \sigma_x)$ tells us the strength of the localized plane wave component at point $x$ traveling along the direction defined by the normalized angle $\sigma_x$. It is such a localized plane wave component passing through a given point at a certain angle with which we can associate a ray (perpendicular to the wavefronts of the plane wave component).

### 4.6.3 Discussion

When dealing with first-order systems, equivalent results are obtained regardless of whether one uses wave optics or geometrical optics, so that a unified mathematical treatment is possible. The $\check{A}\check{B}\check{C}\check{D}$ matrix may be interpreted either as the ray matrix, or the matrix characterizing the linear canonical transform describing how the wavefields are transformed. Speaking of the equivalence of geometrical optics and wave optics requires care, since geometrical optics is normally considered to be an approximation to wave optics. There are two points to be made in this regard: (i) We are concerned with a certain restricted class of systems, namely first-order systems characterized by only three parameters; and (ii) even for these systems, the equivalence in question does not imply that the geometrical-optics viewpoint will properly predict the wavefields everywhere (for instance, near a focal point). However, given the limited number of parameters of the systems under consideration, the ray matrix properly summarizes the operational characteristics of the kernel governing the propagation of the wavefields, so that this kernel and the output wavefield can be easily obtained when desired.

## 4.7 Quadratic-exponential signals

In this section we will discuss two special families of signals whose dependence on the spatial coordinate $x$ is quadratic in the exponent. Apart from a constant multiplicative factor, quadratic-phase *systems* are characterized by 3 parameters (one of the four matrix parameters is redundant because of the unit-determinant condition). A single ray is characterized by 2 real parameters, the position $x$ and normalized angle $\sigma_x$, which we write as a $2 \times 1$ vector $[x \ \sigma_x]^{\mathrm{T}}$. The quadratic-exponential signals we will discuss are likewise

characterized by only two parameters.

This section may be skipped without loss of continuity.

### 4.7.1 Ray-like signals

In the previous section, we noted that a distribution of light whose Wigner distribution is $\delta(x - x_0)\,\delta(\sigma_x - \sigma_{x0})$ could be associated with a ray whose position is $x_0$ and whose normalized angle is $\sigma_{x0}$. We also commented that no actual distribution of light has such a Wigner distribution. A signal whose Wigner distribution comes as close as physically possible to $\delta(x - x_0)\,\delta(\sigma_x - \sigma_{x0})$ is

$$\check{w}_{x_0,\sigma_{x0}}(x) = 2^{1/4}{\Delta_x}^{-1/2}\, e^{-\pi(x-x_0)^2/{\Delta_x}^2}\, e^{i2\pi\sigma_{x0}x}. \tag{4.105}$$

The Wigner distribution of this signal can be expressed as (Bastiaans 1978, 1989)

$$2\exp\left[-2\pi\left(\frac{(x-x_0)^2}{\Delta_x^2} + \frac{(\sigma_x - \sigma_{x0})^2}{\Delta_{\sigma_x}^2}\right)\right], \tag{4.106}$$

where $\Delta_{\sigma_x} = 1/\Delta_x$. This is a distribution of light which is spatially concentrated within a region of extent $\Delta_x$ around the point $x_0$, and spectrally concentrated within a region of extent $\Delta_{\sigma_x}$ around the frequency $\sigma_{x0}$. This distribution of light is as close as we can physically get to a ray passing through $x_0$ with normalized angle $\sigma_{x0}$.

Such a signal $\check{w}_{x_0,\sigma_{x0}}(x)$ is characterized by two parameters corresponding to position and frequency, just as a ray is characterized by two parameters corresponding to position and normalized angle, and of course we already know that normalized angle corresponds to frequency. Thus both a ray and $\check{w}_{x_0,\sigma_{x0}}(x)$ can be characterized by a $2 \times 1$ vector $[x_0\ \sigma_{x0}]^{\mathrm{T}}$ which is transformed by the $\check{A}\check{B}\check{C}\check{D}$ matrix upon passage through a quadratic-phase system. When $\check{w}_{x_0,\sigma_{x0}}(x)$ is input to a quadratic-phase system, the output is another similar signal characterized by the vector $[(\check{A}x_0 + \check{B}\sigma_{x0}), (\check{C}x_0 + \check{D}\sigma_{x0})]^{\mathrm{T}}$, which tells us where the output signal is centered in the space-frequency plane. (We are not dealing with the shape parameters of $\check{w}_{x_0,\sigma_{x0}}(x)$ which are also transformed in passing through the system, concentrating only on its center $(x_0, \sigma_{x0})$.)

A ray and the signal $\check{w}_{x_0,\sigma_{x0}}(x)$ can both be seen as elementary signals characterized by two real parameters. Just as more general distributions of light can be represented geometrical-optically as a bundle of rays, they can also be expressed wave-optically as a linear superposition of signals of the form of $\check{w}_{x_0,\sigma_{x0}}(x)$. If we let $x_0 = l\,\delta x$ and $\sigma_{x0} = m\,\delta\sigma_x$, we immediately recognize these signals as Gaussian basis functions appearing in a Gabor expansion (equation 2.7), so that the linear superposition in question is nothing but a Gabor expansion. Thinking of a distribution of light as a bundle of rays corresponds to thinking of it in terms of its Gabor expansion. The Gabor coefficient may be associated with the weight of either the ray in question or a "plane wave" localized in both space and spatial frequency, with its wavefronts perpendicular to the ray.

In subsection 4.6.2 we discussed how to associate a bundle of weighted rays with an arbitrary wavefield $\check{f}(x)$. We simply find the Wigner distribution of the wavefield and assign $\check{W}_{\check{f}}(x, \sigma_x)$ as a weight to the ray $[x\ \sigma_x]^{\mathrm{T}}$. We now consider an alternative approach. Let us write $\check{f}(x) = |\check{f}(x)| \exp[i2\pi\check{\phi}(x)]$ and concentrate on a localized part of this signal by multiplying it with a window function $\propto \exp[-\pi(x - x_0)^2/{\Delta_x}^2]$ centered around $x_0$. Assuming the magnitude and phase are slowly varying we obtain approximately

$$\propto e^{-\pi(x-x_0)^2/{\Delta_x}^2}\, |\check{f}(x_0)|\, e^{i2\pi(d\check{\phi}(x_0)/dx)x}. \tag{4.107}$$

We immediately recognize this to be of the same form as the signal $\check{w}_{x_0,\sigma_{x0}}(x)$ in equation 4.105, with $\sigma_{x0} = d\check{\phi}(x_0)/dx$. What we have done is to associate a localized part of $\check{f}(x)$ with $\check{w}_{x_0,\sigma_{x0}}(x)$, which in turn we had associated with the ray $[x_0\ \sigma_{x0}]^{\mathrm{T}}$. What we have found is that the normalized angle $\sigma_{x0}$ corresponds to the instantaneous spatial frequency $d\check{\phi}(x_0)/dx$. This derivative can also be recognized as the transverse component of the gradient of the phase function. (The gradient itself is perpendicular to the wavefronts and points along the direction of the ray.) Finally, let us recall from table 2.2 the following theorem:

$$\int \sigma_x \check{W}_{\check{f}}(x, \sigma_x)\, d\sigma_x \ \propto\ \frac{d\check{\phi}(x)}{dx}, \tag{4.108}$$

from which we conclude that the derivative of the phase for a certain value of $x$ also corresponds to the average frequency of the signal at that value of $x$, where the average is taken in the space-frequency plane by weighting with the Wigner distribution. This result is more general than interpreting the derivative of the phase as an instantaneous spatial frequency, which has a meaningful physical interpretation only when a narrow band of frequencies dominates at each spatial position.

### 4.7.2 Complex Gaussian signals

We now turn our attention to another family of quadratic-exponential functions which may be referred to as complex Gaussian signals. They are also characterized by 2 parameters (equation 2.149):

$$\check{f}(x) = (2/W^2)^{1/4}\, e^{i\pi x^2/\lambda\check{q}}, \tag{4.109}$$

where $\check{q}$ is the *complex radius of curvature* corresponding to 2 real parameters, and $W^2 > 0$. These signals were already discussed in a purely mathematical context in subsection 2.4.6. As in subsection 3.3.4, we will denote the real and imaginary parts of $\check{q}$ in the form $1/\check{q} = 1/R + i\lambda/W^2$. (Signals with a term linear in $x$ in the exponent are also complex Gaussian signals from a physical viewpoint, but are not considered here for simplicity.) In passing, we note that such signals can also be represented by what is known as the *complex ray representation* (Arnaud 1973).

Now, it follows from the discussion of subsection 2.4.6 that when such a signal passes through a quadratic-phase system, it remains a complex Gaussian signal, and the complex

radius of curvature $\check{q}'$ at the output is related to $\check{q}$ through

$$(\lambda\check{q}') = \frac{\check{A}(\lambda\check{q}) + \check{B}}{\check{C}(\lambda\check{q}) + \check{D}}. \tag{4.110}$$

The Wigner distribution of a complex Gaussian signal can be written as (Bastiaans 1979a, 1989)

$$\begin{aligned}\check{W}_{\check{f}}(x,\sigma_x) &= 2\exp\left\{-2\pi\left[\frac{x^2}{W^2} + W^2\left(\sigma_x - \frac{x}{\lambda R}\right)^2\right]\right\} \\ &= 2\exp\left\{-2\pi\left[\left(\frac{1}{W^2} + \frac{W^2}{\lambda^2 R^2}\right)x^2 - 2\left(\frac{W^2}{\lambda R}\right)x\sigma_x + \left(W^2\right)\sigma_x^2\right]\right\},\end{aligned} \tag{4.111}$$

and retains this form when the signal passes through a quadratic-phase system. The values of $R'$ and $W'$ at the output can be easily found by using equation 4.110. When $1/W^2 = 0$, we expect to recover the simple chirp function $\propto \exp(i\pi x^2/\lambda R)$. However, since the chirp function does not have unit energy, this does not follow as a special case of equation 4.109 by letting $1/W^2 \to 0$. Nevertheless, the result $(\lambda R') = [\check{A}(\lambda R) + \check{B}]/[\check{C}(\lambda R) + \check{D}]$ still holds. We know that the Wigner distribution of the chirp function is given by $\check{W}_{\check{f}}(x,\sigma_x) = \delta(\sigma_x - x/\lambda R)$, which is a line delta concentrated along $\sigma_x = x/\lambda R$. When such a function passes through a quadratic-phase system, its Wigner distribution becomes $\delta(\sigma_x - x/\lambda R')$, where $R'$ is given by the linear fractional transformation formula above.

We now consider the moments of the Wigner distribution of complex Gaussian signals, which were mathematically defined in subsection 2.4.5. Since constant or linear terms were excluded from the exponent of our signal, the first-order moments $\overline{x}$ and $\overline{\sigma_x}$ will be zero. The second-order moments are given by (Bastiaans 1989)

$$\overline{(x-\overline{x})^2} = \overline{x^2} = \iint x^2\check{W}_{\check{f}}(x,\sigma_x)\,dx\,d\sigma_x = W^2/4\pi, \tag{4.112}$$

$$\overline{(\sigma_x-\overline{\sigma_x})^2} = \overline{\sigma_x^2} = \iint \sigma_x^2\check{W}_{\check{f}}(x,\sigma_x)\,dx\,d\sigma_x = (1/W^2 + W^2/\lambda^2R^2)/4\pi, \tag{4.113}$$

$$\overline{(x-\overline{x})(\sigma_x-\overline{\sigma_x})} = \overline{x\sigma_x} = \iint x\sigma_x\check{W}_{\check{f}}(x,\sigma_x)\,dx\,d\sigma_x = W^2/4\pi\lambda R. \tag{4.114}$$

Upon passage through a quadratic-phase system, these moments transform according to equation 2.148, which can be cast in the form of expressions relating $R', W'$ to $R, W$ (which the reader may show are consistent with similar expressions obtained directly from equation 4.110). The reader may also wish to relate the present considerations to the discussion on page 105 by noting that $\overline{x\sigma_x}/\overline{x^2} + i\sqrt{\overline{x^2}\,\overline{\sigma_x^2} - \overline{x\sigma_x}^2}\,/\overline{x^2}$ evaluates to $1/\lambda R + i/W^2 = 1/\lambda\check{q}$. We also note that had we considered the more general case where the first-order moments are not zero, they would have been transformed by equation 2.146.

More generally, equations 2.146 and 2.148 (or the equivalent linear fractional transformation given in equation 2.156) hold for the first- and second-order moments of any signal. Thus, in situations where it is not necessary to fully characterize a signal by its

functional representation $\check{f}(x)$, and it is sufficient to characterize it by its moments, the moments of the output can be easily related to the moments of the input through simple matrix relations. This amounts to approximating the signal by the closest-fitting complex Gaussian signal. Of course, complex Gaussian signals are fully characterized by their first and second moments, so that if we have a complex Gaussian signal itself, we will have complete knowledge of the output once we find its moments.

The second-order moments of the Wigner distribution have been discussed in an optical context in Bastiaans 1979a, 1989, 1991b. Higher-order moments have been discussed in Dragoman 1994.

## 4.8 Optical invariants

Some of the more fundamental statements in physics take the form of conservation laws. Certain quantities or properties have the quality of remaining invariant as a function of time, after propagation through space, or upon passage through certain classes of systems. One such quantity is of course energy, which within our framework, is given by the square of the norm of a signal. We know that the energy of a signal remains unchanged upon passage through a system if the system is unitary, which physically means that it does not contain attenuating or amplifying components. Quadratic-phase systems are unitary and thus conserve energy. On the other hand, a multiplicative filter $\check{h}(x)$ is usually not, unless $|\check{h}(x)| = 1$ for all $x$.

In this section, we will discuss another invariant which is of central importance in optics, often called the *optical invariant*. It takes many forms, has many different interpretations and is referred to by many different names. We have already seen one manifestation of it: the area of support of the Wigner distribution or the ray bundle corresponding to a distribution of light.

As preparation, we recall that for an inhomogeneous medium of one transverse dimension with refractive index distribution $n(x, z)$, the Hamiltonian $\check{H}$ and Hamilton's equations are given by (page 143)

$$\check{H}(x, \sigma_x; z) = -\sqrt{n^2(x, z) f_{\text{oc}}^2/c^2 - \sigma_x^2}, \tag{4.115}$$

$$\frac{dx}{dz} = \frac{\partial \check{H}}{\partial \sigma_x}, \qquad \frac{d\sigma_x}{dz} = -\frac{\partial \check{H}}{\partial x}. \tag{4.116}$$

Hamilton's equations can be written compactly as a single vector equation of the form (Goldstein 1980)

$$\begin{bmatrix} dx/dz \\ d\sigma_x/dz \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} \partial \check{H}/\partial x \\ \partial \check{H}/\partial \sigma_x \end{bmatrix}. \tag{4.117}$$

The $2 \times 2$ matrix appearing in this expression is conventionally denoted by the symbol $\mathbf{J}$. Under the paraxial approximation the Hamiltonian takes the form (equation 3.131)

$$\check{H}(x, \sigma_x; z) = \frac{\sigma_x^2}{2n(x, z) f_{\text{oc}}/c} - n(x, z) f_{\text{oc}}/c. \tag{4.118}$$

Hamilton's equations are equivalent to the ray equation and thus embody the laws of geometrical optics. We also recall the discussion starting on page 145 which showed that (at least in the paraxial case) Hamilton's equations corresponded to Snell's law and the continuity of light rays. In this section we will derive many consequences of Hamilton's equations.

Our discussion of optical invariants is intended to be instructive, rather than complete and general. Our derivations are not always rigorous or complete, nor are they set up for the most general case possible.

### 4.8.1 Invariance of density and area in phase space

One form the optical invariant takes is the density or area in phase space. The invariance of phase-space density and area in optics is an instance of a more general theorem appearing in many branches of physics called *Liouville's theorem* (Arnaud 1976, Marcuse 1982).

Let us consider a bundle of rays occupying an arbitrary region in phase space. We will let $\rho$ denote the density of the rays in phase space. As $z$ increases, each ray will move in phase space according to the functions $x(z)$, $\sigma_x(z)$. Thus, the "velocity" of the phase-space point corresponding to a ray is given by $\mathbf{v} = [dx(z)/dz \; d\sigma_x(z)/dz]^{\mathrm{T}}$. We now write the continuity equation

$$\frac{\partial \rho}{\partial z} + \nabla \cdot (\rho \mathbf{v}) = 0, \tag{4.119}$$

which is essentially a statement of the fact that rays do not disappear and are not created. (That this equation holds can be seen by considering the equation $(\partial/\partial z) \iint \rho \, dx \, d\sigma_x = -\int \rho \mathbf{v} \cdot \mathbf{dl}$. The double integral on the left hand side is taken over some region and simply gives the total number of rays in that region. $\mathbf{dl}$ is a vector line element normal to the boundary of the region. The integral on the right hand side is taken over the boundary and tells us the number of rays exiting the region. The divergence theorem states that the right hand side is equal to $-\iint \nabla \cdot (\rho \mathbf{v}) \, dx \, d\sigma_x$, from which the continuity equation follows upon noting that these equations hold for any arbitrary region.) The continuity equation can be explicitly written as

$$\frac{\partial \rho}{\partial z} + \frac{\partial}{\partial x}\left(\rho \frac{dx(z)}{dz}\right) + \frac{\partial}{\partial \sigma_x}\left(\rho \frac{d\sigma_x(z)}{dz}\right) = 0. \tag{4.120}$$

Now, using Hamilton's equations, expanding the product rule for differentiation, and again using Hamilton's equations, we obtain

$$\frac{\partial \rho}{\partial z} + \frac{\partial \rho}{\partial x}\frac{\partial x}{\partial z} + \frac{\partial \rho}{\partial \sigma_x}\frac{\partial \sigma_x}{\partial z} = 0, \tag{4.121}$$

whose left hand side we recognize as the total derivative of $\rho$ and thus conclude that

$$\frac{d\rho}{dz} = 0, \tag{4.122}$$

which means that the density of rays (or phase-space points) remains invariant upon propagation through a medium with arbitrary refractive index. As we move along the rays with increasing $z$, the boundary of the region enclosing the ray bundle changes shape. However, since the density of rays remains constant, the fact that the total number of rays is constant also implies that the area of this region remains invariant. Mathematically,

$$\text{number of rays} = \text{density of rays} \times \text{phase-space area}, \tag{4.123}$$

so that invariance of the density implies invariance of the phase-space area. The area occupied by a bundle of rays in phase space is one of the most common forms of the optical invariant. (The reader may wish to imagine a fluid spread on a table. If the surface density of the fluid remains constant, the area occupied by it will also remain constant no matter how the region changes.) The above discussion is based on Marcuse 1982, where it is also shown that the invariance of density and area also holds when the refractive index distribution is discontinuous.

Now, let us write the phase-space coordinates of the rays $[x\ \sigma_x]^{\mathrm{T}}$ at an arbitrary plane $z$ as a function of their respective phase-space coordinates $[x_0\ \sigma_{x0}]^{\mathrm{T}}$ at some other plane $z_0$ as follows:

$$\begin{aligned} x &= x(x_0, \sigma_{x0}), \\ \sigma_x &= \sigma_x(x_0, \sigma_{x0}). \end{aligned} \tag{4.124}$$

These equations tell us the phase-space location of a ray at the plane $z$ in terms of its phase-space location at the plane $z_0$. Now, consider the area integrals $\iint dx_0\, d\sigma_{x0}$ and $\iint dx\, d\sigma_x$ taken over the respective regions enclosing the ray bundle at the planes $z_0$ and $z$. If we make a substitution of variables using equations 4.124 in the second of these integrals, this integral will be identical to the first integral except that it will be multiplied by the Jacobian of the variable transformation in question. Since we have seen that the area is invariant, the two integrals must be equal, and it follows that the Jacobian associated with the transformation must be unity.

(We recall that the area elements in such a variable transformation are related by $dx\, d\sigma_x = |\mathbf{M}|\, dx_0\, d\sigma_{x0}$, where the Jacobian matrix $\mathbf{M}$ is given by

$$\mathbf{M} = \begin{bmatrix} \partial x/\partial x_0 & \partial x/\partial \sigma_{x0} \\ \partial \sigma_x/\partial x_0 & \partial \sigma_x/\partial \sigma_{x0} \end{bmatrix}, \tag{4.125}$$

and $|\mathbf{M}|$ is the determinant of the matrix, referred to as the Jacobian of the transformation.)

The results just derived are generalizations of results we already knew to hold for quadratic-phase systems: The area of the region enclosing a ray bundle is invariant and the determinant $\check{A}\check{D} - \check{B}\check{C}$ of the linear transformation matrix (the $\check{A}\check{B}\check{C}\check{D}$ matrix) is equal to unity.

The reader can also show (in a manner similar to that leading us to equation 4.122) that $\nabla \cdot \mathbf{v} = 0$: the divergence of the velocity vector is zero. The total flux of the phase-space "fluid" out of the boundary enclosing the ray bundle is zero. In other words, the phase-space "fluid" behaves like an incompressible fluid (Lanczos 1970).

Before we close this subsection, we make note of the physical interpretations of the total derivative and the partial derivative of an entity $\check{K}(x, \sigma_x; z)$ with respect to $z$. The total derivative $d\check{K}/dz$ is the rate of change of $\check{K}$ as we follow a ray $[x(z)\ \sigma_x(z)]^{\mathrm{T}}$, whereas the partial derivative $\partial\check{K}/\partial z$ is the explicit rate of change with respect to $z$ for fixed $x$ and $\sigma_x$. The two derivatives are of course related by

$$\frac{d\check{K}}{dz} = \frac{\partial\check{K}}{\partial x}\frac{dx}{dz} + \frac{\partial\check{K}}{\partial \sigma_x}\frac{d\sigma_x}{dz} + \frac{\partial\check{K}}{\partial z}. \tag{4.126}$$

If $d\check{K}/dz = 0$, the entity $\check{K}$ remains invariant along a ray. We have already seen that the phase-space density $\rho$ is such an invariant entity (equation 4.122). We had also seen that the value of the Wigner distribution remains constant along a ray (equation 4.104), consistent with our interpretation of the Wigner distribution as the phase-space density in wave optics.

### 4.8.2 The symplectic condition and canonical transformations

If a ray has the phase-space coordinates $[x_0\ \sigma_{x0}]^{\mathrm{T}}$ at $z_0$, the phase-space coordinates $[x\ \sigma_x]^{\mathrm{T}}$ at an arbitrary value of $z$ will be given by transformation expressions of the form of equation 4.124 (Goldstein 1980):

$$\begin{aligned} x &= x(x_0, \sigma_{x0}; z), \\ \sigma_x &= \sigma_x(x_0, \sigma_{x0}; z), \end{aligned} \tag{4.127}$$

which are solutions $x(z)$, $\sigma_x(z)$ of equation 4.117 (or the ray equation) corresponding to the "initial conditions" $x(z_0) = x_0$ and $\sigma_x(z_0) = \sigma_{x0}$. The Jacobian matrix $\mathbf{M}(z)$ of the transformation given in equation 4.127 is

$$\mathbf{M}(z) = \begin{bmatrix} \partial x/\partial x_0 & \partial x/\partial \sigma_{x0} \\ \partial \sigma_x/\partial x_0 & \partial \sigma_x/\partial \sigma_{x0} \end{bmatrix}. \tag{4.128}$$

We will now show that this Jacobian matrix satisfies the following relation for all $z$:

$$\mathbf{J} = \mathbf{M}^{\mathrm{T}}\mathbf{J}\mathbf{M}. \tag{4.129}$$

This relation, known as the *symplectic condition*, already appeared as equation 2.180, where we had mentioned that matrices $\mathbf{M}$ satisfying such a relation are referred to as *symplectic matrices* (Goldstein 1980). The *symplectic form* which they preserve is nothing but phase-space area or density, as will be further discussed in subsection 4.8.3.

To demonstrate equation 4.129, we first note that it is a matter of simple matrix algebra to show that this equation is equivalent to the unit-determinant condition $|\mathbf{M}| = 1$:

$$\frac{\partial x}{\partial x_0}\frac{\partial \sigma_x}{\partial \sigma_{x0}} - \frac{\partial x}{\partial \sigma_{x0}}\frac{\partial \sigma_x}{\partial x_0} = 1. \tag{4.130}$$

But we had already seen in the previous subsection that this determinant, the Jacobian of the transformation, is equal to unity. Thus equation 4.129 also holds. We conclude that for systems with only one transverse dimension, *the unit-Jacobian condition and phase-space area conservation, are fully equivalent to the symplectic condition.*

The unit-Jacobian condition and thus equation 4.129 can also be derived directly from Hamilton's equations. First, note from equation 4.127 that $|\mathbf{M}|$ is a function of $(x_0, \sigma_{x0}; z)$. Second, note that when $z = z_0$ the Jacobian matrix is the identity matrix and its determinant is unity. Thus if we show that $d|\mathbf{M}(z)|/dz = 0$ for all $(x_0, \sigma_{x0}; z)$, we will have shown that $|\mathbf{M}(z)| = 1$ along all rays for all $z$. Concentrating on a particular ray, $(x_0, \sigma_{x0})$ is merely a label for that ray, so that there is no distinction between the total and partial derivatives of $x$, $\sigma_x$, and $|\mathbf{M}|$ with respect to $z$. Now, we explicitly take the derivative of $|\mathbf{M}|$ with respect to $z$:

$$\frac{\partial |\mathbf{M}|}{\partial z} = \frac{\partial^2 x}{\partial z \partial x_0}\frac{\partial \sigma_x}{\partial \sigma_{x0}} + \frac{\partial x}{\partial x_0}\frac{\partial^2 \sigma_x}{\partial z \partial \sigma_{x0}} - \frac{\partial^2 x}{\partial z \partial \sigma_{x0}}\frac{\partial \sigma_x}{\partial x_0} - \frac{\partial x}{\partial \sigma_{x0}}\frac{\partial^2 \sigma_x}{\partial z \partial x_0}. \tag{4.131}$$

By exchanging the order of the mixed partial derivatives and using Hamilton's equations we obtain

$$\frac{\partial |\mathbf{M}|}{\partial z} = \frac{\partial^2 \check{H}}{\partial x_0 \partial \sigma_x}\frac{\partial \sigma_x}{\partial \sigma_{x0}} - \frac{\partial x}{\partial x_0}\frac{\partial^2 \check{H}}{\partial \sigma_{x0} \partial x} - \frac{\partial^2 \check{H}}{\partial \sigma_{x0} \partial \sigma_x}\frac{\partial \sigma_x}{\partial x_0} + \frac{\partial x}{\partial \sigma_{x0}}\frac{\partial^2 \check{H}}{\partial x_0 \partial x}. \tag{4.132}$$

The factor $\partial^2 \check{H}/\partial x_0 \partial \sigma_x$ appearing in the first term is equal to

$$\frac{\partial^2 \check{H}}{\partial x_0 \partial \sigma_x} = \frac{\partial}{\partial x_0}\frac{\partial \check{H}}{\partial \sigma_x} = \left(\frac{\partial}{\partial x}\frac{\partial \check{H}}{\partial \sigma_x}\right)\frac{\partial x}{\partial x_0} + \left(\frac{\partial}{\partial \sigma_x}\frac{\partial \check{H}}{\partial \sigma_x}\right)\frac{\partial \sigma_x}{\partial x_0}. \tag{4.133}$$

It is possible to write similar chain rules for the corresponding factors appearing in the three remaining terms. Upon substitution, all terms cancel and we are left with zero, completing the derivation. A similar derivation is given in Arnaud 1976.

The reader may rightly think that equation 4.129 is just a fancy way of writing the unit-Jacobian condition. One advantage of equation 4.129 is that it remains true in two and higher dimensions, with the zeros and ones in the matrix $\mathbf{J}$ being replaced by higher-dimensional zero and identity matrices. In this case, the symplectic condition is equivalent to a set of conditions involving the blocks of $\mathbf{M}$ which is more complicated than the unit-determinant condition (for instance, see Bastiaans 1979a and Folland 1989).

What we have shown is that, a transformation of the form of equation 4.127 which is a solution of Hamilton's equations or the ray equation, and thus which corresponds to the propagation of actual light rays (rather than being some arbitrary transformation), satisfies the symplectic condition.

Equation 4.127 was introduced as a relation which gives us the output phase-space coordinates of rays in terms of their input phase-space coordinates. Mathematically, this equation describes a change of coordinates in phase space from $(x_0, \sigma_{x0})$ to $(x, \sigma_x)$. Thus, the propagation of light rays from one plane to another can be viewed as a coordinate transformation in phase space. The coordinates of the ray at the output plane, is the location of the ray in the input plane expressed with respect to the new coordinate system.

Transformations of the form of equation 4.127 satisfying the symplectic condition are called *canonical transformations* (Goldstein 1980). Since transformations corresponding to the propagation of actual light rays satisfy the symplectic condition for all $z$, the equations describing the propagation of light rays take the form of continuously unfolding canonical transformations. In a broad class of optical systems, this canonical transformation can be interpreted as the fractional Fourier transform. Also recall that linear canonical transforms (quadratic-phase systems) corresponded to linear transformations of the form $x = \check{A}x_0 + \check{B}\sigma_{x0}$, $\sigma_x = \check{C}x_0 + \check{D}\sigma_{x0}$ in phase space, constituting a special case of equation 4.127.

We had seen in chapter 2 that linear canonical transforms (as well as the matrices characterizing them) constituted a group. More generally, the canonical transformations discussed here (as well as their Jacobian matrices satisfying the symplectic condition) also constitute a group (Goldstein 1980).

### 4.8.3 The Lagrange invariant

We had motivated the concept of symplectic forms in subsection 2.4.11. Most optical invariants, in their various forms, are essentially symplectic forms, and the optical systems we are considering are symplectic systems preserving these forms. Certain optical invariants which are explicit symplectic forms are collectively referred to as the *Lagrange invariant*.

Again, we consider a ray whose phase-space coordinates at $z$ are given by $[x(x_0, \sigma_{x0}; z)\ \sigma_x(x_0, \sigma_{x0}; z)]^{\mathrm{T}}$ where $[x_0\ \sigma_{x0}]^{\mathrm{T}}$ are the coordinates at $z_0$. Now, consider the following symplectic form:

$$j_{\mathrm{L}} \equiv [\partial x/\partial x_0\ \partial x/\partial \sigma_{x0}] \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} \partial \sigma_x/\partial x_0 \\ \partial \sigma_x/\partial \sigma_{x0} \end{bmatrix}, \tag{4.134}$$

which is one form of the Lagrange invariant. If the symplectic condition (equation 4.129) holds, then the symplectic form $j_{\mathrm{L}}$ will have the same value for all $z$. In subsection 4.8.7, we will see that the symplectic condition ensures the invariance of a more general class of entities known as *Poisson brackets*. (The form given in equation 4.134 will then be seen to be the Poisson bracket $[x, \sigma_x]_{x_0, \sigma_{x0}}$.) In the particular case of equation 4.134, however, invariance is trivial since the right hand side evaluates to

$$j_{\mathrm{L}} = \frac{\partial x}{\partial x_0}\frac{\partial \sigma_x}{\partial \sigma_{x0}} - \frac{\partial x}{\partial \sigma_{x0}}\frac{\partial \sigma_x}{\partial x_0}, \tag{4.135}$$

which is equal to the Jacobian determinant $|\mathbf{M}|$. The fact that the symplectic condition ensures the invariance of $j_{\mathrm{L}}$ becomes a trivial statement since the symplectic condition is

equivalent to the unit-determinant condition to begin with. (Nevertheless, thus closing the circle serves as a consistency check.)

It remains to give a physical interpretation of the entity $j_{\mathrm{L}}$ whose invariance we have just demonstrated. In general, rays emanating from lines of constant $x_0$ and $\sigma_{x0}$ in the phase plane at $z_0$ will appear as curved lines in the phase plane at $z$, and the differential elements $dx_0$ and $d\sigma_{x0}$ will be mapped to differentials which are not orthogonal to each other. $dx$ and $d\sigma_x$ are related to $dx_0$ and $d\sigma_{x0}$ by

$$\begin{bmatrix} dx \\ d\sigma_x \end{bmatrix} = \begin{bmatrix} \partial x/\partial x_0 & \partial x/\partial \sigma_{x0} \\ \partial \sigma_x/\partial x_0 & \partial \sigma_x/\partial \sigma_{x0} \end{bmatrix} \begin{bmatrix} dx_0 \\ d\sigma_{x0} \end{bmatrix} = \mathbf{M} \begin{bmatrix} dx_0 \\ d\sigma_{x0} \end{bmatrix}, \tag{4.136}$$

where $\mathbf{M}$ is the Jacobian matrix. The area of the parallelogram area element defined by the images of $dx_0$ and $d\sigma_{x0}$ can be calculated by taking the "cross product" of $[\partial x/\partial x_0 \ \partial \sigma_x/\partial x_0]^{\mathrm{T}} dx_0$ and $[\partial x/\partial \sigma_{x0} \ \partial \sigma_x/\partial \sigma_{x0}]^{\mathrm{T}} d\sigma_{x0}$ (corresponding to $d\sigma_{x0} = 0$ and $dx_0 = 0$ respectively), which evaluates to $j_{\mathrm{L}} dx_0 d\sigma_{x0} = |\mathbf{M}| dx_0 d\sigma_{x0}$, a result familiar from calculus. Since $j_{\mathrm{L}} = |\mathbf{M}| = 1$, the area of the differential phase-area element corresponding to a differential bundle of rays remains constant for all values of $z$.

Now, we consider another form of the Lagrange invariant (Bastiaans 1979a). Considering two distinct rays $[x_1 \ \sigma_{x1}]^{\mathrm{T}}$ and $[x_2 \ \sigma_{x2}]^{\mathrm{T}}$ at a given plane, we define their symplectic form as

$$[x_1 \ \sigma_{x1}] \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} x_2 \\ \sigma_{x2} \end{bmatrix} = x_1 \sigma_{x2} - x_2 \sigma_{x1}. \tag{4.137}$$

This symplectic form corresponds to the (signed) area of the parallelogram defined by the two ray vectors (see subsection 2.4.11 and Guillemin and Sternberg 1984). We had already discussed the invariance of this form in subsection 2.4.11. Note that here $[x_1 \ \sigma_{x1}]^{\mathrm{T}}$ and $[x_2 \ \sigma_{x2}]^{\mathrm{T}}$ are *not* the positions and angles of the same ray at the input and output, but rather the positions and angles of two distinct rays at the same plane, consistent with subsection 2.4.11.

### 4.8.4 The Smith-Helmholtz invariant and Abbe's sine condition

An important special case of the Lagrange invariant will here be referred to as the Smith-Helmholtz invariant. The invariants we discussed until now remained unchanged throughout the optical system (for all $z$). The Smith-Helmholtz invariant, on the other hand, has the same value at the object plane and any number of image planes. These planes are connected to each other through a matrix of the form (equation 4.72)

$$\begin{bmatrix} M & 0 \\ M/\lambda R & 1/M \end{bmatrix}. \tag{4.138}$$

Again considering the two rays $[x_1 \ \sigma_{x1}]^{\mathrm{T}}$ and $[x_2 \ \sigma_{x2}]^{\mathrm{T}}$ of the previous subsection (Bastiaans 1979a), let us take the second ray to be a ray passing through the origin at the object plane ($x_2 = 0$). If the system is an imaging system, this ray will also pass through the

origin at the image plane. Thus the second term of the Lagrange invariant $x_1\sigma_{x2} - x_2\sigma_{x1}$ will be zero at both the object and image planes, implying the invariance of the first term $x_1\sigma_{x2}$. This result can also be deduced directly from the above matrix. We simply consider two rays, the first crossing the object plane at $x_1$ and the other crossing the object plane at the origin making angle $\sigma_{x2}$. It immediately follows from the matrix that the first ray will cross the image plane at $Mx_1$ and the second ray will cross it at the origin making angle $\sigma_{x2}/M$. Thus the product $(Mx_1)(\sigma_{x2}/M) = x_1\sigma_{x2}$ is the same at both object and image planes. This is directly related to the fact that if the ray bundle (or Wigner distribution) occupies a rectangular region in phase space, the effect of imaging is to simply change the aspect ratio of this rectangle without changing its area.

Here we see the optical invariant in its barest and most transparent form. Nothing more is involved than the fact that spatial and angular magnifications are inverses of each other (as implied by the unit-determinant condition), so that the product of spatial and angular features remains invariant. Remembering that angles correspond to spatial frequencies, we can also discuss the same in the language of wave optics. A perfect imaging system maps an object $\check{f}(x)$ into the image $\propto \check{f}(x/M)$. If the spatial extent of the object is $\Delta x$, that of the image is $M\Delta x$. We will let $\check{F}(\sigma_x)$ denote the Fourier transform of $\check{f}(x)$ with bandwidth $\Delta\sigma_x$. The Fourier transform of the image is $\propto \check{F}(M\sigma_x)$ and is of extent $\Delta\sigma_x/M$. We see that as spatial extent is magnified, frequency extent (bandwidth) is demagnified by the same amount. The phase-space area $(M\Delta x)(\Delta\sigma_x/M)$ of the image is equal to that of the object $\Delta x\Delta\sigma_x$. Now, let us assume that both the function and its Fourier transform are centered around the origin, so that the largest non-negligible frequency component is $\sigma_{x\max} = \Delta\sigma_x/2$ and the largest value of $x$ for which the function is not negligible is $x_{\max} = \Delta x/2$. By recalling the association between ray angles $\theta_x$ and spatial frequencies $\sigma_x$ given by $\sin\theta_x = \lambda\sigma_x$, we find that the largest ray angle $\theta_{x\max}$ satisfies $\sin\theta_{x\max} = \lambda\Delta\sigma_x/2$. From this we conclude that $x_{\max}\sin\theta_{x\max}$ is an invariant between object and image planes. In the paraxial approximation, this reduces to the invariance of $x_{\max}\theta_{x\max}$ discussed in the previous paragraph.

The invariance of the product of a transverse extent and the sine of an angle between object and image planes is known as *Abbe's sine condition*. Thus the Smith-Helmholtz invariant is the paraxial form of Abbe's sine condition. Another way of thinking about this invariant is as follows (Born and Wolf 1980). We consider a first-order expansion of Hamilton's point characteristic $\check{V}(x, x')$ around the origins of the object (input) and image (output) planes:

$$\begin{aligned}\check{V}(x,x') \approx \check{V}(0,0) + \frac{\partial\check{V}}{\partial x}x + \frac{\partial\check{V}}{\partial x'}x' &= \check{V}(0,0) + \sigma_x x - \sigma'_x x', \\ \check{V}(x,x') - \check{V}(0,0) &\approx \sigma_x x - \sigma'_x x',\end{aligned} \tag{4.139}$$

where we have used equation 4.93. Here $x'$ and $x$ are points close to the axis, and since the derivatives are evaluated at the origin, $\sigma'_x$ and $\sigma_x$ are normalized angles of rays passing

through the origin at the object and image planes. If $\sigma'_x = 0$, we have a ray coinciding with the optical axis so that also $\sigma_x = 0$, implying that the right hand side and thus the left hand side of the last displayed equation are equal to zero. But since the left hand side of the displayed equation is independent of the normalized angles $\sigma'_x$ and $\sigma_x$, it follows that it is always equal to zero and therefore $\sigma_x x = \sigma'_x x'$.

We also mention a much less known special case of the Lagrange invariant which holds not between object and image planes, but between planes connected by a perfect Fourier transform matrix (equation 4.76 with $R \to \infty$):

$$\begin{bmatrix} 0 & s^2 M \\ -1/s^2 M & 0 \end{bmatrix}. \tag{4.140}$$

It is easy to see that $|x\sigma_x|$ is an invariant since the same product formed at the Fourier plane is $|s^2 M \sigma_x(-x/s^2 M)| = |x\sigma_x|$.

### 4.8.5 The constant brightness theorem

We saw on page 159 that the *brightness* represented the power per area per solid angle. Thus, expressed in suitable units, the brightness can be thought to represent power per unit phase-space area (the power density in phase space). Assuming the brightness is uniform over the region of interest in phase space, we can therefore write

$$\text{total power} = \text{brightness} \times \text{phase-space area}. \tag{4.141}$$

Now, since the phase-space area is invariant, the brightness will also be invariant in a lossless and gainless system. This is the constant brightness theorem. (A more traditional approach may be found in Born and Wolf 1980, pages 188–189.) The above relation is closely related to equation 4.123, with brightness corresponding to ray density and total power corresponding to the total number of rays.

More generally, when the brightness is not uniform, $B(P, \Theta)$ essentially corresponds to the Wigner distribution $\check{W}_{\check{f}}(x, y, \sigma_x, \sigma_y)$. In one-dimensional notation, the generalization of equation 4.141 may be written as the familiar phase-space integral

$$\text{total power} = \iint \check{W}_{\check{f}}(x, \sigma_x)\, dx\, d\sigma_x. \tag{4.142}$$

The many results derived throughout our discussion of optical invariants are essentially consequences of Hamilton's equations and therefore the basic laws governing the propagation of light. The law of constant brightness is also known to be a consequence of the second law of thermodynamics (if it could be violated, one could heat up a hotter body with a colder one, Boyd 1983), a fact which provides further insight into the physics underlying various forms of the optical invariant.

### 4.8.6 The unit-determinant condition for inhomogeneous media

In this subsection we present an alternative derivation of the unit-determinant condition for inhomogeneous media, which may be more appealing to those who prefer to think in physical rather than abstract terms. We consider a weakly inhomogeneous medium with refractive index distribution $n(x,z)$. With the paraxial Hamiltonian given in equation 3.131, Hamilton's equations take the form

$$\frac{dx}{dz} = \frac{\sigma_x}{nf_{\rm oc}/c}, \qquad \frac{d\sigma_x}{dz} = (f_{\rm oc}/c)\frac{\partial n}{\partial x}, \tag{4.143}$$

leading to

$$\begin{aligned} x(z+dz) &= x(z) + \frac{1}{nf_{\rm oc}/c}\, dz\, \sigma_x, \\ \sigma_x(z+dz) &= \sigma_x(z) + (f_{\rm oc}/c)\, dz\, \frac{\partial n}{\partial x}. \end{aligned} \tag{4.144}$$

A linear matrix relation will result only if $n(x,z)$ can be approximated by a quadratic function of the form $K_2(z)x^2/2 + K_0(z)$ in which case

$$\begin{bmatrix} x(z+dz) \\ \sigma_x(z+dz) \end{bmatrix} = \begin{bmatrix} 1 & dz(nf_{\rm oc}/c)^{-1} \\ dz(f_{\rm oc}/c)K_2(z) & 1 \end{bmatrix} \begin{bmatrix} x(z) \\ \sigma_x(z) \end{bmatrix}. \tag{4.145}$$

More generally, the Jacobian matrix will replace the above matrix, corresponding to the lower left term being replaced by $dz(f_{\rm oc}/c)\partial^2 n/\partial x^2$. In either case, the determinant of the matrix is of the form $1-[\cdots]dz^2$. Since the deviation of this determinant from unity is only of second order, the determinant of the system between any two planes will be unity. For instance, if the distance between two planes is given by $L$, then the overall determinant will be the product of $L/dz$ terms of the form $1-[\cdots]dz^2$, where the bracketed factor varies from term to term. Although we do not provide the details, readers may convince themselves that this product is unity by considering the case where the bracketed factor is some constant, say $\iota$. Then, by noting the standard limit $\lim_{w\to\infty}(1-c/w)^w = \exp(-c)$, we find that $\lim_{dz\to 0}(1-\iota dz^2)^{L/dz} = \exp(-dz\,\iota L) = 1$.

### 4.8.7 Poisson brackets

The *Poisson bracket* $[\check{K},\check{L}]_{x,\sigma_x}$ of $\check{K}(x,\sigma_x)$ and $\check{L}(x,\sigma_x)$ with respect to $(x,\sigma_x)$ is defined as (Goldstein 1980)

$$[\check{K},\check{L}]_{x,\sigma_x} \equiv \frac{\partial\check{K}}{\partial x}\frac{\partial\check{L}}{\partial\sigma_x} - \frac{\partial\check{K}}{\partial\sigma_x}\frac{\partial\check{L}}{\partial x} = [\partial\check{K}/\partial x \;\; \partial\check{K}/\partial\sigma_x]\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} \partial\check{L}/\partial x \\ \partial\check{L}/\partial\sigma_x \end{bmatrix} \tag{4.146}$$

The matrix Poisson bracket of two vectors $[\check{K}_1\;\check{K}_2]^{\rm T}$ and $[\check{L}_1\;\check{L}_2]^{\rm T}$ is defined as

$$\Big[[\check{K}_1\;\check{K}_2]^{\rm T}, [\check{L}_1\;\check{L}_2]^{\rm T}\Big]_{x,\sigma_x} \equiv \begin{bmatrix} \big[\check{K}_1,\check{L}_1\big]_{x,\sigma_x} & \big[\check{K}_1,\check{L}_2\big]_{x,\sigma_x} \\ \big[\check{K}_2,\check{L}_1\big]_{x,\sigma_x} & \big[\check{K}_2,\check{L}_2\big]_{x,\sigma_x} \end{bmatrix}. \tag{4.147}$$

Using this definition, a number of special cases and properties follow immediately:

$$\Big[[x_0\ \sigma_{x0}]^{\mathrm{T}}, [x_0\ \sigma_{x0}]^{\mathrm{T}}\Big]_{x_0,\sigma_{x0}} = \mathbf{J}, \tag{4.148}$$

$$\Big[[x\ \sigma_x]^{\mathrm{T}}, [x\ \sigma_x]^{\mathrm{T}}\Big]_{x,\sigma_x} = \mathbf{J}, \tag{4.149}$$

$$\Big[[x\ \sigma_x]^{\mathrm{T}}, [x\ \sigma_x]^{\mathrm{T}}\Big]_{x_0,\sigma_{x0}} = \mathbf{M}\mathbf{J}\mathbf{M}^{\mathrm{T}}, \tag{4.150}$$

where $\mathbf{J}$ is the $2\times2$ matrix appearing in equation 4.146 and $\mathbf{M}$ is the Jacobian matrix given in equation 4.125. The first two equations are merely two instances of the same equation. Now, if $[x\ \sigma_x]^{\mathrm{T}}$ is related to $[x_0\ \sigma_{x0}]^{\mathrm{T}}$ through an actual physical ray transformation of the form of equation 4.127, then we know that $\mathbf{M}$ is symplectic and the right hand side of the third equation becomes equal to $\mathbf{J}$. (This can be demonstrated in a manner similar to equation 4.129. In one dimension, the conditions $\mathbf{M}\mathbf{J}\mathbf{M}^{\mathrm{T}} = \mathbf{J}$ and $\mathbf{M}^{\mathrm{T}}\mathbf{J}\mathbf{M} = \mathbf{J}$ are equivalent to each other and the unit-determinant condition.) Thus the right hand sides of the second and third equations become identical, and we conclude that the Poisson bracket on the left hand side is invariant under canonical transformations of the form of equation 4.127 (Goldstein 1980). In other words, it has the same value regardless of which variables it is evaluated with respect to (provided these are related through canonical transformations). The reader may have already noted that Poisson brackets are *symplectic forms* (subsection 2.4.11).

We now show the invariance of all Poisson brackets (Goldstein 1980). First, we write the following chain rule:

$$\begin{bmatrix} \partial\check{L}/\partial x_0 \\ \partial\check{L}/\partial\sigma_{x0} \end{bmatrix} = \mathbf{M}^{\mathrm{T}} \begin{bmatrix} \partial\check{L}/\partial x \\ \partial\check{L}/\partial\sigma_x \end{bmatrix},$$

$$\Big[\partial\check{L}/\partial x_0\ \partial\check{L}/\partial\sigma_{x0}\Big] = \Big[\partial\check{L}/\partial x\ \partial\check{L}/\partial\sigma_x\Big]\,\mathbf{M}, \tag{4.151}$$

which of course also hold for $\check{K}$. Now, we write the Poisson bracket $[\check{K},\check{L}]_{x_0,\sigma_{x0}}$ as

$$\begin{aligned}[\partial\check{K}/\partial x_0\ \partial\check{K}/\partial\sigma_{x0}] \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} \partial\check{L}/\partial x_0 \\ \partial\check{L}/\partial\sigma_{x0} \end{bmatrix} \\ = [\partial\check{K}/\partial x\ \partial\check{K}/\partial\sigma_x]\mathbf{M} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \mathbf{M}^{\mathrm{T}} \begin{bmatrix} \partial\check{L}/\partial x \\ \partial\check{L}/\partial\sigma_x \end{bmatrix}.\end{aligned} \tag{4.152}$$

The condition $\mathbf{M}\mathbf{J}\mathbf{M}^{\mathrm{T}} = \mathbf{J}$ leads us to conclude that $[\check{K},\check{L}]_{x_0,\sigma_{x0}} = [\check{K},\check{L}]_{x,\sigma_x}$. Poisson brackets are invariant under canonical transformations, which physically means they are invariants of ray propagation. Because of this invariance, the subscripts of the brackets are sometimes dropped. The reader may check that as binary operators, Poisson brackets are not associative. They obey a nonassociative algebra, known as a *Lie algebra*. The vector cross product and the commutator of two operators can also be used to construct similar Lie algebras. (Goldstein 1980)

We now consider the total derivative of some entity $\check{K}(x, \sigma_x; z)$ with respect to $z$:

$$\frac{d\check{K}}{dz} = \frac{\partial \check{K}}{\partial x}\frac{dx}{dz} + \frac{\partial \check{K}}{\partial \sigma_x}\frac{d\sigma_x}{dz} + \frac{\partial \check{K}}{\partial z} = \frac{\partial \check{K}}{\partial x}\frac{\partial \check{H}}{\partial \sigma_x} - \frac{\partial \check{K}}{\partial \sigma_x}\frac{\partial \check{H}}{\partial x} + \frac{\partial \check{K}}{\partial z} = [\check{K}, \check{H}] + \frac{\partial \check{K}}{\partial z}. \quad (4.153)$$

This is an equation for the evolution of the entity $\check{K}$ in terms of Poisson brackets (Goldstein 1980). If we take $\check{K} = x$ or $\check{K} = \sigma_x$ as special cases, we obtain $dx/dz = [x, \check{H}]$ or $d\sigma_x/dz = [\sigma_x, \check{H}]$, which the reader can show to be equivalent to Hamilton's equations. If we take $\check{K} = \check{H}$, we recover the familiar result $d\check{H}/dz = \partial \check{H}/\partial z$. If $\check{K}$ is an invariant ($d\check{K}/dz = 0$), then we must have $-[\check{K}, \check{H}] = [\check{H}, \check{K}] = \partial \check{K}/\partial z$. An example of the latter case has already been encountered in equation 4.102, whose right hand side we can now recognize to be $[\check{H}, \check{W}_{\check{f}_z(x)}]$. (Goldstein 1980)

When $\partial \check{K}/\partial z = 0$, it is possible to write a formal series solution for the equation $d\check{K}/dz = [\check{K}, \check{H}]$, and by recognizing the similarity of this expansion to that of an exponential function, express the series in "hyperbracket" form (Goldstein 1980). Although we do not pursue this any further, we mention that it closely corresponds to the hyperdifferential form for operators discussed in chapter 2.

### 4.8.8 The number of degrees of freedom

In chapter 2 we discussed at length the concept of the number of degrees of freedom of a set of signals, seeing that it is given by the area of the space-frequency support of the set of signals. In this section, we have seen that the area of the space-frequency support is an invariant upon passage through a broad class of optical systems. If we consider members of the set of signals in question to be inputs to such an optical system, the set of signals consisting of the outputs will also have the same number of degrees of freedom. Thus the number of degrees of freedom is another form of the optical invariant. Such optical systems preserve the information content of signals passing through them. This is also consistent with the reciprocity of wave and ray propagation.

In a typical optical system, the number of degrees of freedom of the set of signals which can pass through that system, can be determined by examining the *space-frequency aperture* of the system. Although this phase-space aperture may in general have different shapes (Lohmann and others 1996a), it is commonly assumed to be of rectangular form with the spatial extent determined by a spatial aperture in the object or image plane, and the frequency extent determined by an aperture in a Fourier plane. If these apertures are of length $\Delta x_{\text{sp}}$ and $\Delta x_{\text{fr}}$ respectively, and the scale factor relating these planes to each other is $s$, then the number of degrees of freedom is given by $\Delta x_{\text{sp}} \Delta x_{\text{fr}}/s^2$ (see page 83). The number of degrees of freedom and information carrying capacity of refractive and diffractive lenses, as well as how these scale with various parameters have been discussed in Ozaktas and Urey 1993 and Ozaktas, Urey, and Lohmann 1994.

The concept of the number of degrees of freedom or the space-bandwidth product has received considerable attention in the area of information optics in a way which is not

reflected by our limited discussion. For a sampling of historical works, see for instance, Toraldo di Francia 1955, 1969, Gabor 1961, Walther 1967, Winthrop 1971, Frieden 1971, Gori and Guattari 1971, 1973, Bendinelli and others 1974, Lohmann 1986, and the further references in these works. More recent works are diverse and too numerous to list here. We will note van Dekker and van den Bos 1997, Mendlovic and Lohmann 1997, Mendlovic, Lohmann, and Zalevsky 1997, and Miller 1998. The first is a review of past and present approaches to optical resolution. The second and third deal with the adaptation of the space-frequency aperture of an optical system so as to maximally benefit from the available space-frequency area, even when its shape does not match the space-frequency support of the signals to be processed. Miller examines the (functional) singular value decomposition for propagation between two volumes. This amounts to finding pairs of orthonormal functions $\check{\psi}_l(x')$ (at the input) and $\check{\phi}_l(x)$ (at the output) which couple into each other with strength $g_l$ (the singular value). (This means that $\check{\psi}_l$ at the input is mapped to $g_l\check{\phi}_l$ at the output.) These pairs of functions constitute independent spatial "channels" of information transfer, or "communication modes" between the input and the output. The "strength" of each channel, given by the singular value, tells us how reliable that channel is. Channels with very small strengths will couple the signal weakly and perhaps below the noise floor and are thus not reliable. Miller shows (at least when the medium of propagation is free space) that the sum of the squared connection strengths is a constant, so that even if there are an infinite number of channels, only a finite number of them can have substantial strength and thus be considered reliable: there is effectively a finite number $N$ of channels or spatial degrees of freedom. In our context, $N$ corresponds to the phase-space area or space-bandwidth product. In many typical physical systems, the eigenvalues (or singular values), when listed in decreasing order, are seen to have values comparable to each other (and often to unity) until we come close to the $N$th eigenvalue. Then the eigenvalues decrease sharply and become close to zero as we pass the $N$th eigenvalue.

These considerations are closely related to the problem of solving so-called *Fredholm integral equations of the first kind*:

$$\check{g}(x) = \int \check{h}(x, x')\check{f}(x')\, dx', \tag{4.154}$$

where $\check{g}(x)$ is measured with a finite degree of accuracy, $\check{h}(x, x')$ is assumed known, and we seek to find $\check{f}(x')$. Such problems are known as *inverse problems* and they are considered to be *ill-posed* in the sense that even if the error in measuring $\check{g}(x)$ is small, the error in the estimate of $\check{f}(x')$ may be large (Image Recovery 1987). This is a consequence of the fact that if we look at the discrete spectral expansion of some $\check{h}(x, x')$ corresponding to a real physical system, we see that the number of terms for which the eigenvalues are substantial are finite, and thus the system effectively filters out a portion of the information inherent in $\check{f}(x')$. The system does not pass all of the information inherent in the function $\check{f}(x')$, but rather only partial information represented by a finite number of degrees of freedom.

The remaining information is lost. (In most systems, the lost information corresponds to the high-frequency content of $\check{f}(x')$.)

## 4.9 Partially coherent light

The study of partially coherent light is an area in which the phase-space picture and the Wigner distribution have found great use. We have already discussed in chapter 3 that partially coherent light is characterized as a random process, and we have already defined the Wigner distribution of a random process in chapter 2. Thus, we already know how to obtain the Wigner distribution of partially coherent light and characterize it in phase space.

Unfortunately, we will not have space in this book to further develop this subject, and must satisfy ourselves by referring the reader to a short list of references: Bastiaans 1981b, 1986a, b, 1991a; Friberg 1981, 1986. Bastiaans 1997 is a particularly useful starting point.

## 4.10 Further reading

The subject matter of this chapter naturally leads the way to a deeper study of mathematical optics. Similar to the study of advanced classical and quantum mechanics, the study of advanced mathematical optics is heavily based on phase-space concepts. (In fact, there is a rather close mathematical analogy between mechanics and optics, which the reader may have already noted. For instance, see Lanczos 1970 or Goldstein 1980.) The calculus of variations plays an important part in the development of variational principles such as that of Fermat. Elementary expositions may be found in Hildebrand 1965, Born and Wolf 1980, and Goldstein 1980. Lanczos 1970 is an insightful text and Caratheodory 1965 is a translation of a classic work on the subject.

Relatively advanced texts containing discussions on general geometrical optics and mathematical optics including the Hamiltonian formulation include Synge 1937, 1954, Born and Wolf 1980, Stavroudis 1972, Arnaud 1976, Marcuse 1982, and Solimeno, Crosignani, and Di Porto 1986. Luneburg 1964, Kline and Kay 1965, and Deschamps 1972 are noted for the development of geometrical optics based on the foundation of Maxwell's equations. Luneburg's work also contains extensions of several of the concepts discussed in this chapter. Pegis 1961, Arnaud 1973, and Sekiguchi and Wolf 1987 specifically concentrate on Hamiltonian methods. Wolf and Krötzsch 1995 discusses discontinuities and refraction in the Hamiltonian framework.

A sampling of works dealing with Lie methods might include Dragt and Finn 1976, Dragt 1982, Wolf 1986a, Ferraro and Caelli 1988, and the edited volumes Lie Methods in Optics 1986, 1989. As for symplectic and group-theoretical techniques we might mention the mathematical text by Folland (1989) and the physics or optics oriented works Stavroudis 1972 (especially chapter 13: The fundamental optical invariant, chapter 14:

The lens equation, and chapter 15: The lens group), Guillemin and Sternberg 1981, 1984 (especially chapter 1: Introduction), Bacry and Cadilhac 1981, Navarro-Saad and Wolf 1986a, Perelomov 1986, Wolf 1986b, and Kauderer 1990.

An additional sampling of relatively recent works on phase-space optics more from a geometrical-optics perspective includes Tanaka 1986, Navarro-Saad and Wolf 1986b, Wright and Garrison 1987, Wolf 1991, 1993, Wolf and Kurmyshev 1993, and Campbell 1994.

Bastiaans has been a major contributor to the use of phase-space techniques from a Fourier optics perspective (1978, 1979a, b, c, d, 1980, 1981a, b, 1982a, b, 1986a, b, 1989, 1991a, b, 1994). This body of work deals with the use of the Wigner distribution, the Gabor expansion and other space-frequency representations as well as linear canonical transforms in optics. A very readable exposition is Bastiaans 1997, which is particularly recommended as a complement to this chapter. A number of other relatively early works which have been found instructive are Papoulis 1974, Bartelt, Brenner, and Lohmann 1980, Bamler and Glünder 1983, and Easton, Ticknor, and Barrett 1984. An earlier work bringing together partial coherence, radiometry, and the Wigner distribution is Walther 1968. The Wigner distribution of polychromatic wavefields is discussed in Wolf 1996.

A review of the Wigner distribution in optics is Dragoman 1997. A recent collection of works on the subject is Wigner Distributions and Phase Space in Optics 2000.

In this book we have totally excluded the discussion of quantum optics in phase space. For this the reader is referred to Gardiner 1991, Schleich, Mayr, and Krähmer 1999, and Carmichael 1999.

# Acknowledgments

We are grateful to several individuals for many fruitful discussions, reading and commenting on substantial parts of the text, for answering our many questions, or providing information: Laurence Barker, M. Fatih Erden, David Mendlovic, David Mustard, K. Bernardo Wolf, and Zeev Zalevsky. We would also like to extend our gratitude to others who commented on various parts of the text, answered our questions, made suggestions or pointed out errors in early drafts, or otherwise lent us their expertise: Yamaç Dikmelik, Lütfiye Durak, David A. B. Miller, and A. Pınar Tan.

# Bibliography


S. Abe and J. T. Sheridan. Generalization of the fractional Fourier transformation to an arbitrary linear lossless transformation: an operator approach. *J Phys A*, 27:4179–4187, 1994. Corrigenda in 7937–7938.

S. Abe and J. T. Sheridan. Optical operations on wave functions as the Abelian subgroups of the special affine Fourier transformation. *Opt Lett*, 19:1801–1803, 1994

S. Abe and J. T. Sheridan. Almost-Fourier and almost-Fresnel transformations. *Opt Commun*, 113:385–388, 1995.

S. Abe and J. T. Sheridan. Comment on 'The fractional Fourier transform in optical propagation problems.' *J Mod Opt*, 42:2373–2378, 1995.

S. Abe and J. T. Sheridan. The Wigner distribution function, and the special affine Fourier transform: signal processing and optical imaging. *Zoological Studies*, 34:121–122, 1995.

S. Abe and J. T. Sheridan. An optical implementation for the estimation of the fractional-Fourier order. *Opt Commun*, 137:214–218, 1997.

G. S. Agarwal and R. Simon. A simple realization of fractional Fourier transforms and relation to harmonic oscillator Green's function. *Opt Commun*, 110:23–26, 1994.

A. Akan and L. F. Chaparro. Discrete rotational Gabor transform. In *Proc 1996 IEEE-SP Int Symp Time-Frequency Time-Scale Analysis*, IEEE, Piscataway, New Jersey, 1996. Pages 169–172.

O. Akay and G. F. Boudreaux-Bartels. Linear fractionally invariant systems: fractional filtering and correlation via fractional operators. In *Proc 31st Asilomar Conf Signals, Systems, Computers*, IEEE, Piscataway, New Jersey, 1997. Pages 1494–1498.

O. Akay and G. F. Boudreaux-Bartels. Unitary and Hermitian fractional operators and their relation to the fractional Fourier transform. *IEEE Signal Processing Lett*, 5:312–314, 1998.

O. Akay and G. F. Boudreaux-Bartels. Fractional autocorrelation and its applications to detection and estimation of linear FM signals. In *Proc 1998 IEEE-SP Int Symp Time-Frequency Time-Scale Analysis*, IEEE, Piscataway, New Jersey, 1998. Pages 213–216.

O. Akay and G. F. Boudreaux-Bartels. Joint fractional representations. In *Proc 1998 IEEE-SP Int Symp Time-Frequency Time-Scale Analysis*, IEEE, Piscataway, New Jersey, 1998. Pages 417–420.

O. Akay and G. F. Boudreaux-Bartels. Fractional Mellin transform: an extension of frac-

tional frequency concept for scale. In *Proc 8th IEEE Int Digital Signal Processing Workshop*, IEEE, Piscataway, New Jersey, 1998. CD-ROM.

O. Akay and G. F. Boudreaux-Bartels. Broadband interference excision in spread spectrum communication systems via fractional Fourier transform. In *Proc 32nd Asilomar Conf Signals, Systems, Computers*, IEEE, Piscataway, New Jersey, 1998. Pages 832–837.

T. Alieva. On the self-fractional Fourier functions. *J Phys A*, 29:L377–L379, 1996.

T. Alieva. Fractional Fourier transform as a tool for investigation of fractal objects. *J Opt Soc Am A*, 13:1189–1192, 1996.

T. Alieva and F. Agulló-López. Reconstruction of the optical correlation function in a quadratic refractive index medium. *Opt Commun*, 114:161–169, 1995. Erratum in 118:657, 1995.

T. Alieva and F. Agullo-Lopez. Optical wave propagation of fractal fields. *Opt Commun*, 125:267–274, 1996.

T. Alieva and F. Agullo-Lopez. Imaging in first-order optical systems. *J Opt Soc Am A*, 13:2375–2380, 1996.

T. Alieva and F. Agullo-Lopez. Diffraction analysis of random fractal fields. *J Opt Soc Am A*, 15:669–674, 1998.

T. Alieva and A. M. Barbé. Self-fractional Fourier functions and selection of modes. *J Phys A*, 30:L211–L215, 1997.

T. Alieva and A. M. Barbé. Fractional Fourier and Radon-Wigner transforms of periodic signals. *Signal Processing*, 69:183–189, 1998.

T. Alieva and A. M. Barbé. About quantum state characterization. *J Phys A*, 31:L685–L688, 1998.

T. Alieva and A. M. Barbé. Self-fractional Fourier images. *J Mod Opt*, 46:83–99, 1999.

T. Alieva, V. Lopez, F. Agullo-Lopez, and L. B. Almeida. The fractional Fourier transform in optical propagation problems. *J Mod Opt*, 41:1037–1044, 1994.

T. Alieva, V. Lopez, F. Agullo-Lopez, and L. B. Almeida. Reply to the comment on the fractional Fourier transform in optical propagation problems. *J Mod Opt*, 42:2379–2383, 1995.

A. M. Almanasreh and M. A. G. Abushagur. Fractional correlations based on the modified fractional order Fourier transform. *Opt Eng*, 37:175–184, 1998.

L. B. Almeida. An introduction to the angular Fourier transform. In *Proc 1993 IEEE Int Conf Acoustics, Speech, Signal Processing*, IEEE, Piscataway, New Jersey, 1993. Pages III-257–III-260.

L. B. Almeida. The fractional Fourier transform and time-frequency representations. *IEEE Trans Signal Processing*, 42:3084–3091, 1994.

L. B. Almeida. Product and convolution theorems for the fractional Fourier transform. *IEEE Signal Processing Lett*, 4:15–17, 1997.

M. A. Alonso and G. W. Forbes. Fractional Legendre transformation. *J Phys A*, 28:5509–5527, 1995.

M. A. Alonso and G. W. Forbes. Uniform asymptotic expansions for wave propagators via

fractional transformations. *J Opt Soc Am A*, 14:1279–1292, 1997.

P. Andrés, W. D. Furlan, G. Saavedra, and A. W. Lohmann. Variable fractional Fourier processor: a simple implementation. *J Opt Soc Am A*, 14:853–858, 1997.

P. Antosik, J. Mikusiński, and R. Sikorski. *Theory of Distributions: The Sequential Approach.* Elsevier, Amsterdam, 1973.

O. Arıkan, M. A. Kutay, H. M. Özaktaş, and Ö. K. Akdemir. The discrete fractional Fourier transformation. In *Proc 1996 IEEE-SP Int Symp Time-Frequency Time-Scale Analysis*, IEEE, Piscataway, New Jersey, 1996. Pages 205–207.

V. Arrizón, J. G. Ibarra, and J. Ojeda-Castañeda. Matrix formulation of the Fresnel transform of complex transmittance gratings. *J Opt Soc Am A*, 13:2414–2422, 1996.

N. M. Atakishiyev, L. E. Vicent, and K. B. Wolf. Continuous vs. discrete fractional Fourier transforms. *J Computational Applied Mathematics*, 107:73–95, 1999.

N. M. Atakishiyev and K. B. Wolf. Fractional Fourier-Kravchuk transform. *J Opt Soc Am A*, 14:1467–1477, 1997.

N. M. Atakishiyev, E. I. Jafarov, S. M. Nagiyev, and K. B. Wolf. Meixner oscillators. *Revista Mexicana Fisica*, 44:235–244, 1998.

O. Aytür and H. M. Ozaktas. Non-orthogonal domains in phase space of quantum optics and their relation to fractional Fourier transforms. *Opt Commun*, 120:166-170, 1995.

D. H. Bailey and P. N. Swarztrauber. The fractional Fourier transform and applications. *SIAM Review*, 33:389–404, 1991.

D. H. Bailey and P. N. Swarztrauber. A fast method for the numerical evaluation of continuous Fourier and Laplace transforms. *SIAM J Sci Comput*, 15:1105–1110, 1994.

S. Barbarossa, Analysis of multicomponent LFM signals by a combined Wigner-Hough transform. *IEEE Trans Signal Processing*, 43:1511–1515, 1995.

V. Bargmann. On a Hilbert space of analytic functions and an associated integral transform, Part I. *Comm Pure and Applied Mathematics*, 14:187–214, 1961.

L. Barker. The discrete fractional Fourier transform and Harper's equation. Technical Report BU-CE-0007, Bilkent University, Department of Computer Engineering, Ankara, January 2000.

L. Barker, Ç. Candan, T. Hakioğlu, M. A. Kutay, and H. M. Ozaktas. The discrete harmonic oscillator, Harper's equation, and the discrete fractional Fourier transform. *J Phys A*, 33:2209–2222, 2000.

B. Barshan, M. A. Kutay, and H. M. Ozaktas. Optimal filtering with linear canonical transformations. *Opt Commun*, 135:32–36, 1997.

M. J. Bastiaans and A. J. van Leest. From the rectangular to the quincunx Gabor lattice via fractional Fourier transformation. *IEEE Signal Processing Lett*, 5:203–205, 1998.

M. Beck, M. G. Raymer, I. A. Walmsley, and V. Wong. Chronocyclic tomography for measuring the amplitude and phase structure of optical pulses. *Opt Lett*, 18:2041–2043, 1993.

L. M. Bernardo. ABCD matrix formalism of fractional Fourier optics. *Opt Eng*, 35:732–740, 1996.

L. M. Bernardo. Talbot self-imaging in fractional Fourier planes of real and complex orders. *Opt Commun*, 140:195–198, 1997.

L. M. Bernardo and O. D. D. Soares. Fractional Fourier transforms and optical systems. *Opt Commun*, 110:517–522, 1994.

L. M. Bernardo and O. D. D. Soares. Fractional Fourier transforms and imaging. *J Opt Soc Am A*, 11:2622–2626, 1994.

L. M. Bernardo and O. D. D. Soares. Optical fractional Fourier transforms with complex orders. *Appl Opt*, 35:3163–3166, 1996.

J. Bertrand and P. Bertrand. A tomographic approach to Wigner's function. *Foundations Phys*, 17:397–405, 1987.

Y. Bitran, D. Mendlovic, R. G. Dorsch, A. W. Lohmann, and H. M. Ozaktas. Fractional Fourier transform: simulations and experimental results. *Appl Opt*, 34:1329–1332, 1995.

Y. Bitran, Z. Zalevsky, D. Mendlovic, and R. G. Dorsch. Fractional correlation operation: Performance analysis. *Appl Opt*, 35:297–303, 1996.

N. G. de Bruijn. A theory of generalized functions, with applications to Wigner distribution and Weyl correspondence. *Nieuw Archief voor Wiskunde*, 21:205–280, 1973.

A. Bultan. A four-parameter atomic decomposition of chirplets. *IEEE Trans Signal Processing*, 47:731–745, 1999.

Ç. Candan. *The Discrete Fractional Fourier Transform.* MS thesis, Bilkent University, Ankara, 1998.

Ç. Candan, M. A. Kutay, and H. M. Ozaktas. The discrete fractional Fourier transform. In *Proc 1999 IEEE Int Conf Acoustics, Speech, Signal Processing*, IEEE, Piscataway, New Jersey, 1999. Pages 1713–1716.

Ç. Candan, M. A. Kutay, and H. M. Ozaktas. The discrete fractional Fourier transform. *IEEE Trans Signal Processing*, 48, 2000.

G. Cariolaro, T. Erseghe, P. Kraniauskas, and N. Laurenti. A unified framework for the fractional Fourier transform. *IEEE Trans Signal Processing*, 46:3206–3219, 1998.

E. U. Condon. Immersion of the Fourier transform in a continuous group of functional transformations. *Proc National Academy Sciences*, 23:158–164, 1937.

W. X. Cong and N. X. Chen. Optimization of Gaussian beam shaping in fractional Fourier transform domain. *Prog Natural Science*, 9:96–102, 1999.

W. X. Cong, N. X. Chen, and B. Y. Gu. Beam shaping and its solution with the use of an optimization method. *Appl Opt*, 37:4500–4503, 1998.

W. X. Cong, N. X. Chen, and B. Y. Gu. Recursive algorithm for phase retrieval in the fractional Fourier transform domain. *Appl Opt*, 37:6906–6910, 1998.

W. X. Cong, N. X. Chen, and B. Y. Gu. A new method for phase retrieval in the optical system. *Chinese Phys Lett*, 15:24–26, 1998.

W. X. Cong, N. X. Chen, and B. Y. Gu. A recursive method for phase retrieval in Fourier transform domain. *Chinese Science Bulletin*, 43:40–44, 1998.

G. Dattoli, A. Torre, and G. Mazzacurati. An alternative point of view to the theory of fractional Fourier transform. *IMA J Appl Math*, 60:215–224, 1998.

Z.-T. Deng, H. J. Caulfield, and M. Schamschula. Fractional discrete Fourier transforms. *Opt Lett*, 21:1430–1432, 1996.

X. Deng, Y. Li, Y. Qiu, and D. Fan. Diffraction interpreted through fractional Fourier transforms. *Opt Commun*, 131:241–245, 1996.

X. Deng, Y. Li, D. Fan, and Y. Qiu. A fast algorithm for fractional Fourier transforms. *Opt Commun*, 138:270–274, 1997.

B.-Z. Dong, Y. Zhang, B.-Y. Gu, and G.-Z. Yang. Numerical investigation of phase retrieval in a fractional Fourier transform. *J Opt Soc Am A*, 14:2709–2714, 1997.

R. G. Dorsch. Fractional Fourier transformer of variable order based on a modular lens system. *Appl Opt*, 34:6016–6020, 1995.

R. G. Dorsch and A. W. Lohmann. Fractional Fourier transform used for a lens design problem. *Appl Opt*, 34:4111–4112, 1995.

R. G. Dorsch, A. W. Lohmann, Y. Bitran, D. Mendlovic, and H. M. Ozaktas. Chirp filtering in the fractional Fourier domain. *Appl Opt*, 33:7599–7602, 1994.

D. Dragoman. Fractional Wigner distribution function. *J Opt Soc Am A*, 13:474–478, 1996.

D. Dragoman. The relation between light diffraction and the fractional Fourier transform. *J Mod Opt*, 45;2117–2124, 1998.

D. Dragoman and M. Dragoman. Time-frequency modeling of atomic force microscopy. *Opt Commun*, 140:220–225, 1997.

D. Dragoman and M. Dragoman. Near and far field optical beam characterization using the fractional Fourier tansform. *Opt Commun*, 141:5–9, 1997.

D. Dragoman and M. Dragoman. Temporal implementation of Fourier-related transforms. *Opt Commun*, 145:33–37, 1998.

D. Dragoman, K.-H. Brenner, M. Dragoman, J. Bahr, and U. Krackhardt. Hemispherical-rod microlens as a variant fractional Fourier transformer. *Opt Lett*, 23:1499–1501, 1998.

M. F. Erden. *Repeated Filtering in Consecutive Fractional Fourier Domains*. PhD thesis, Bilkent University, Ankara, 1997.

M. F. Erden, M. A. Kutay, and H. M. Ozaktas. Repeated filtering in consecutive fractional Fourier domains and its application to signal restoration. *IEEE Trans Signal Processing*, 47:1458–1462, 1999.

M. F. Erden, M. A. Kutay, and H. M. Ozaktas. Applications of the fractional Fourier transform to filtering, estimation and restoration. In *Proc 1999 IEEE-EURASIP Workshop Nonlinear Signal Image Processing*, Boğaziçi University, Bebek, Istanbul, 1999. Pages 481–485.

M. F. Erden and H. M. Ozaktas. Synthesis of general linear systems with repeated filtering in consecutive fractional Fourier domains. *J Opt Soc Am A*, 15:1647–1657, 1998.

M. F. Erden, H. M. Ozaktas, and D. Mendlovic. Propagation of mutual intensity expressed in terms of the fractional Fourier transform. *J Opt Soc Am A*, 13:1068–1071, 1996.

M. F. Erden, H. M. Ozaktas, and D. Mendlovic. Synthesis of mutual intensity distributions using the fractional Fourier transform. *Opt Commun*, 125:288–301, 1996.

M. Fatih Erden, Haldun M. Ozaktas, Aysegul Sahin, and David Mendlovic. Design of dy-

namically adjustable anamorphic fractional Fourier transformer. *Opt Commun*, 136:52–60, 1997.

J. R. Fonollosa. Positive time-frequency distributions based on joint marginal constraints. *IEEE Trans Signal Processing*, 44:2086–2091, 1996.

J. R. Fonollosa and C. L. Nikias. A new positive time-frequency distribution. In *Proc 1994 IEEE Int Conf Acoustics, Speech, Signal Processing*, IEEE, Piscataway, New Jersey, 1994. Pages IV-301–IV-304.

J. García, D. Mas, and R. G. Dorsch. Fractional-Fourier-transform calculation through the fast-Fourier-transform algorithm. *Appl Opt*, 35:7013–7018, 1996.

J. García, D. Mendlovic, Z. Zalevsky, and A. Lohmann. Space-variant simultaneous detection of several objects by the use of multiple anamorphic fractional-Fourier-transform filters. *Appl Opt*, 35:3945–3952, 1996.

J. García, R. G. Dorsch, A. W. Lohmann, C. Ferreira, and Z. Zalevsky. Flexible optical implementation of fractional Fourier transform processors. Applications to correlation and filtering. *Opt Commun*, 133:393–400, 1997.

C. Gómez-Reino, C. Bao and M. V. Pérez. GRIN optics, Fourier optics and optical connections. In *17th Congress ICO: Optics for Science and New Technology, SPIE Proc 2778*, SPIE, Bellingham, Washington, 1996. Pages 128–131.

F. Gori, M. Santarsiero, and V. Bagini. Fractional Fourier transform and Fresnel transform. *Atti della Fondazione Giorgio Ronchi*, IL:387–390, 1994.

S. Granieri, R. Arizaga, and E. E. Sicre. Optical correlation based on the fractional Fourier transform. *Appl Opt*, 36:6636–6645, 1997.

S. Granieri, E. E. Sicre, and W. D. Furlan. Performance analysis of optical imaging systems based on the fractional Fourier transform. *J Mod Opt*, 45:1797–1807, 1998.

S. Granieri, O. Trabocchi, and E. E. Sicre. Fractional Fourier transform applied to spatial filtering in the Fresnel domain. *Opt Commun*, 119:275–278, 1995.

S. Granieri, M. del Carmen Lasprilla, N. Bolognini, and E. E. Sicre. Space-variant optical correlator based on the fractional Fourier transform: implementation by the use of a photorefractive $Bi_{12}GeO_{2_a}$ (BGO) holographic filter. *Appl Opt*, 35:6951–6954, 1996.

S. Granieri, W. D. Furlan, G. Saavedra, and P. Andrés. Radon-Wigner display: a compact optical implementation with a single varifocal lens. *Appl Opt*, 36:8363–8369, 1997.

A. P. Guinand. Matrices associated with fractional Hankel and Fourier transformations. *Proc Glasg Math Assoc*, 2:185–192, 1956.

Ö. Güleryüz. *Feature Extraction with the Fractional Fourier Transform.* MS thesis, Bilkent University, Ankara, 1998.

H. Hamam and H. H. Arsenault. Fresnel transform-based correlator. *Appl Opt*, 36:7408–7414, 1997.

R. Howe. The oscillator semigroup. In *The Mathematical Heritage of Hermann Weyl: Proc Symp Pure Mathematics 48*, American Mathematical Society, Providence, Rhode Island, U.S.A., 1988. Pages 61–132.

J. Hua and L. Liu. Optical dual fractional Fourier and Fresnel self-imaging transform.

*Optik*, 105:47–50, 1997.

J. Hua, L. Liu, and G. Li. Observing the fractional Fourier transform by free-space Fresnel diffraction. *Appl Opt*, 36:512–513, 1997.

J. Hua, L. Liu, and G. Li. Scaled fractional Fourier transform and its optical implementation. *Appl Opt*, 36:8490–8492, 1997.

J. Hua, L. Liu, and G. Li. Extended fractional Fourier transforms. *J Opt Soc Am A*, 14:3316–3322, 1997.

J. Hua, L. Liu, and G. Li. Performing the fractional Fourier transform by one Fresnel diffraction and one lens. *Opt Commun*, 137:11–12, 1997.

J. W. Hua, L. R. Liu, and G. Q. Li. Imaginary angle fractional Fourier transform and its optical implementation. *Science in China Series E: Technological Sciences*, 40:374–378, 1997.

Y. Huang and B. Suter. The fractional wave packet transform. *Multidimensional Systems Signal Processing*, 9:399–402, 1998.

E. Jagoszewski. Fractional Fourier transform in optical setups. *Optica Applicata*, 28:227–237, 1998.

E. Jagoszewski. Conventional and fractional optical Fourier transform. *Optica Applicata*, 28:257–259, 1998.

D. F. V. James and G. S. Agarwal. Generalized Radon transform for tomographic measurement of short pulses. *J Opt Soc Am A*, 12:704–708, 1995.

D. F. V. James and G. S. Agarwal. The generalized Fresnel transform and its applications to optics. *Opt Commun*, 207–212, 1996.

Z. Jiang. Scaling laws and simultaneous optical implementation of various order fractional Fourier transforms. *Opt Lett*, 20:2408–2410, 1995.

Z. Jiang, Q. Lu, and Y. Zhao. Sensitivity of the fractional Fourier transform to parameters and its application in optical measurement. *Appl Opt*, 8455–8458, 1997.

Y. B. Karasik. Expression of the kernel of a fractional Fourier transform in elementary functions. *Opt Lett*, 19:769–770, 1994.

D. B. Karp. The fractional Hankel transform and its applications in mathematical physics. *Russian Acad Sci Dokl Math*, 50:179–185, 1995.

F. H. Kerr. A distributional approach to Namias' fractional Fourier transforms. *Proc Royal Soc Edinburgh*, 108A:133–143, 1988.

F. H. Kerr. Namias' fractional Fourier transforms on $L^2$ and applications to differential equations. *J Math Anal Applic*, 136:404–418, 1988.

F. H. Kerr. A fractional power theory for Hankel transforms in $L^2(\mathbf{R}^+)$. *J Math Anal Applic*, 158:114–123, 1991.

F. H. Kerr. Fractional powers of Hankel transforms in the Zemanian spaces. *J Math Anal Applic*, 166:65–83, 1992.

R. S. Khare. Fractional Fourier analysis of defocused images. *Opt Commun*, 12:386–388, 1974.

H. Kober. Wurzeln aus der Hankel-, Fourier- und aus anderen stetigen Transformationen.

*Quart J Math (Oxford)*, 10:45–59, 1939.

F. Kong and B. Lü. Implementation of the fractional Fourier transform with a paraboloidal reflective mirror. *Optik*, 106:42–43, 1997.

P. Kraniauskas, G. Cariolaro, and T. Erseghe. Method for defining a class of fractional operations. *IEEE Trans Signal Processing*, 46:2804–2807, 1998.

C. J. Kuo and Y. Luo. Generalized joint fractional Fourier transform correlators: a compact approach. *Appl Opt*, 37:8270–8276, 1998.

M. A. Kutay. *Optimal Filtering in Fractional Fourier Domains.* MS thesis, Bilkent University, Ankara, 1995.

M. A. Kutay. *Generalized Filtering Configurations with Applications in Digital and Optical Signal and Image Processing.* PhD thesis, Bilkent University, Ankara, 1999.

M. A. Kutay and H. M. Ozaktas. Optimal image restoration with the fractional Fourier transform. *J Opt Soc Am A*, 15:825–833, 1998.

M. A. Kutay, H. M. Ozaktas, L. Onural, and O. Arıkan. Optimal filtering in fractional Fourier domains. In *Proc 1995 IEEE Int Conf Acoustics, Speech, Signal Processing*, IEEE, Piscataway, New Jersey, 1995. Pages 937–940.

M. A. Kutay, H. M. Ozaktas, O. Arıkan, and L. Onural. Optimal filtering in fractional Fourier domains. *IEEE Trans Signal Processing*, 45:1129–1143, 1997.

M. A. Kutay, M. F. Erden, H. M. Ozaktas, O. Arıkan, Ö. Güleryüz, and Ç. Candan. Space-bandwidth-efficient realizations of linear systems. *Opt Lett*, 23:1069–1071, 1998.

M. A. Kutay, M. F. Erden, H. M. Ozaktas, O. Arıkan, Ç. Candan, and Ö. Güleryüz. Cost-efficient approximation of linear systems with repeated and multi-channel filtering configurations. In *Proc 1998 IEEE Int Conf Acoustics, Speech, Signal Processing*, IEEE, Piscataway, New Jersey, 1998. Pages 3433–3436.

M. A. Kutay, H. Özaktaş, M. F. Erden, H. M. Ozaktas, and O. Arıkan. Solution and cost analysis of general multi-channel and multi-stage filtering circuits. In *Proc 1998 IEEE-SP Int Symp Time-Frequency Time-Scale Analysis*, IEEE, Piscataway, New Jersey, 1998. Pages 481–484.

M. A. Kutay, H. Özaktaş, H. M. Ozaktas, and O. Arıkan. The fractional Fourier domain decomposition. *Signal Processing*, 77:105–109, 1999.

M. A. Kutay, H. Özaktaş, M. F. Erden, and H. M. Ozaktas. Optimization and cost analysis of multi-stage and multi-channel filtering configurations. Technical Report BU-CE-0004, Bilkent University, Department of Computer Engineering, Ankara, January 2000.

E. V. Labunets and V. G. Labunets. Fast fractional Fourier transform. In *9th European Signal Processing Conference: Signal Processing IX: Theories Applications*, Typorama Publications, Patras, Greece, 1998. Pages 1757–1760.

A. Lakhtakia. Fractal self-Fourier functions. *Optik*, 94:51–52, 1993.

A. Lakhtakia and H. J. Caulfield. On some mathematical and optical integral transforms of self-similar (fractal) functions. *Optik*, 91:131–133, 1992.

J. Lancis, T. Szoplik, E. Tajahuerce, V. Climent, and M. Fernández-Alonso. Fractional derivative Fourier plane filter for phase-change visualization. *Appl Opt*, 36:7461–7464,

1997.

K. G. Larkin. A beginner's guide to the fractional Fourier transform, part 1. *Australian Optical Society News*, June 1995, pages 18–21.

K. G. Larkin. A beginner's guide to the fractional Fourier transform, part 2: a brief history of time frequency distributions. *Australian Optical Society News*, December 1995, pages 13–17.

S.-Y. Lee and H. H. Szu. Fractional Fourier transforms, wavelet transforms, and adaptive neural networks. *Opt Eng*, 33:2326–2330, 1994.

Z. Y. Liu, X. Y. Wu, and D. Y. Fan. Collins formula in frequency-domain and fractional Fourier transforms. *Opt Commun*, 155:7–11. 1998.

S. Liu, J. Wu, and C. Li. Cascading the multiple stages of optical fractional Fourier transforms under different variable scales. *Opt Lett*, 20:1415–1417, 1995.

S. Liu, J. Zhang, and Y. Zhang. Properties of the fractionalization of a Fourier transform. *Opt Commun*, 133:50–54, 1997.

S. Liu, J. Xu, Y. Zhang, L. Chen, and C. Li. General optical implementation of fractional Fourier transforms. *Opt Lett*, 20:1053–1055, 1995.

S. Liu, J. Jiang, Y. Zhang, and J. Zhang. Generalized fractional Fourier transforms. *J Phys A*, 30:973–981, 1997.

S. Liu, H. Ren, J. Zhang, and Z. Zhang. Image-scaling problem in the optical fractional Fourier transform. *Appl Opt*, 36:5671–5674, 1997.

A. W. Lohmann. Concept and implementation of the fractional Fourier transform. In *Optical Computing 1993, Tech Digest Ser 7*, OSA, Washington DC, 1993. Pages 18–21.

A. W. Lohmann. Image rotation, Wigner rotation, and the fractional order Fourier transform. *J Opt Soc Am A*, 10:2181–2186, 1993.

A. W. Lohmann. Fourier Curios. In *Current Trends in Optics*, Academic Press, London, 1994. Chapter 11, pages 149–161.

A. W. Lohmann. A fake zoom lens for fractional Fourier experiments. *Opt Commun*, 115:437–443, 1995.

A. W. Lohmann and D. Mendlovic. Fractional Fourier transform: photonic implementation. *Appl Opt*, 33:7661–7664, 1994.

A. W. Lohmann and D. Mendlovic. Fractional joint transform correlator. *Appl Opt*, 36:7402–7407, 1997.

A. W. Lohmann, D. Mendlovic, and G. Shabtay. Coherence waves. *J Opt Soc Am A*, 16:359–363, 1999.

A. W. Lohmann, D. Mendlovic, and Z. Zalevsky. Fractional Hilbert transform. *Opt Lett*, 21:281–283, 1996.

A. W. Lohmann, D. Mendlovic, and Z. Zalevsky. Fractional transformations in optics. In *Progress in Optics XXXVIII*, Elsevier, Amsterdam, 1998. Chapter IV, pages 263–342.

A. W. Lohmann and B. H. Soffer. Relationships between two transforms: Radon-Wigner and fractional Fourier. In *1993 OSA Annual Meeting Tech Digest*, OSA, Washington DC, 1993. Page 109.

A. W. Lohmann and B. H. Soffer. Relationships between the Radon-Wigner and fractional Fourier transforms. *J Opt Soc Am A*, 11:1798–1801, 1994.

A. W. Lohmann, E. Tepichín, and J. G. Ramírez. Optical implementation of the fractional Hilbert transform for two-dimensional objects. *Appl Opt*, 36:6620–6626, 1997.

A. W. Lohmann, Z. Zalevsky, and D. Mendlovic. Synthesis of pattern recognition filters for fractional Fourier processing. *Opt Commun*, 128:199–204, 1996.

A. W. Lohmann, R. G. Dorsch, D. Mendlovic, Z. Zalevsky, and C. Ferreira. Space-bandwidth product of optical signals and systems. *J Opt Soc Am A*, 13:470–473, 1996.

A. W. Lohmann, D. Mendlovic, Z. Zalevsky, and R. G. Dorsch. Some important fractional transformations for signal processing. *Opt Commun*, 125:18–20, 1996.

A. W. Lohmann, Z. Zalevsky, R. G. Dorsch, and D. Mendlovic. Experimental considerations and scaling property of the fractional Fourier transform. *Opt Commun*, 146:55–61, 1998.

B. Lü and F. Kong. Synthesis and decomposition of fractional Fourier optical systems written in terms of ABCD matrix. *Optik*, 106:155–158, 1997.

B. Lü, F. Kong, and B. Zhang. Optical systems expressed in terms of fractional Fourier transforms. *Opt Commun*, 137:13–16, 1997.

L. F. Ludwig. General thin-lens action on spatial intensity distribution behaves as non-integer powers of Fourier transform. In *Spatial Light Modulators and Applications 1988*, OSA, Washington DC, 1988. Pages 173–176.

S. Mann and S. Haykin. 'Chirplets' and 'warblets': novel time-frequency methods. *Electronics Lett*, 28:114–116, 1992.

S. Mann and S. Haykin. The chirplet transform: physical considerations. *IEEE Trans Signal Processing*, 43:2745–2761, 1995.

M. E. Marhic. Roots of the identity operator and optics. *J Opt Soc Am A*, 12:1448–1459, 1995.

D. F. McAlister, M. Beck, L. Clarke, A. Meyer, and M. G. Raymer. Optical phase-retrieval by phase-space tomography and fractional-order Fourier transforms. *Opt Lett*, 20:1181–1183, 1995.

A. C. McBride and F. H. Kerr. On Namias's fractional Fourier transforms. *IMA J Appl Math*, 39:159–175, 1987.

D. Mendlovic. The fractional Fourier transform: a tutorial. In *Proc 1999 IEEE-EURASIP Workshop Nonlinear Signal Image Processing*, Boğaziçi University, Bebek, Istanbul, 1999. Pages 476–480.

D. Mendlovic and H. M. Ozaktas. Fractional Fourier transforms and their optical implementation: I. *J Opt Soc Am A*, 10:1875–1881, 1993. Reprinted in *Selected Papers on Fourier Optics, SPIE Milestone Ser 105*, SPIE, Bellingham, Washington, 1995. Pages 427–433.

D. Mendlovic and H. M. Ozaktas. Fourier transforms of fractional order and their optical interpretation. In *16th Congress ICO: Optics as a Key to High Technology, SPIE Proc 1983*, SPIE, Bellingham, Washington, 1993. Pages 387–388.

D. Mendlovic and H. M. Ozaktas. Fractional Fourier transform in optics. In *18th Congress ICO: Optics for the Next Millennium, SPIE Proc 3749*, SPIE, Bellingham, Washington, 1999. Pages 40–41.

D. Mendlovic, H. M. Ozaktas, and A. W. Lohmann. Fourier transforms of fractional order and their optical interpretation. In *Optical Computing 1993, Tech Digest Ser 7*, OSA, Washington DC, 1993. Pages 127–130.

D. Mendlovic, H. M. Ozaktas, and A. W. Lohmann. Graded-index fibers, Wigner-distribution functions, and the fractional Fourier transform. *Appl Opt*, 33:6188–6193, 1994.

D. Mendlovic, H. M. Ozaktas, and A. W. Lohmann. Self Fourier functions and fractional Fourier transforms. *Opt Commun*, 105:36–38, 1994.

D. Mendlovic, H. M. Ozaktas, and A. W. Lohmann. Fractional correlation. *Appl Opt*, 34:303–309, 1995.

D. Mendlovic, Z. Zalevsky, and N. Konforti. Computation considerations and fast algorithms for calculating the diffraction integral. *J Mod Opt*, 44:407–414, 1997.

D. Mendlovic, Z. Zalevsky, and H. M. Ozaktas. Applications of the fractional Fourier transform to optical pattern recognition. In *Optical Pattern Recognition*, Cambridge University Press, Cambridge, 1998. Chapter 4, pages 89–125.

D. Mendlovic, Y. Bitran, R. G. Dorsch, and A. W. Lohmann. Optical fractional correlation: experimental results. *Appl Opt*, 34:1665–1670, 1995.

D. Mendlovic, Y. Bitran, R. G. Dorsch, C. Ferreira, J. Garcia, and H. M. Ozaktaz. Anamorphic fractional Fourier transform: optical implementation and applications. *Appl Opt*, 34:7451–7456, 1995.

D. Mendlovic, Z. Zalevsky, N. Konforti, R. G. Dorsch, and A. W. Lohmann. Incoherent fractional Fourier transform and its optical implementation. *Appl Opt*, 34:7615–7620, 1995.

D. Mendlovic, Z. Zalevsky, R. G. Dorsch, Y. Bitran, A. W. Lohmann, and H. Ozaktas. New signal representation based on the fractional Fourier transform: definitions. *J Opt Soc Am A*, 12:2424–2431, 1995.

D. Mendlovic, R. G. Dorsch, A. W. Lohmann, Z. Zalevsky, and C. Ferreira. Optical illustration of a varied fractional Fourier-transform order and the Radon-Wigner display. *Appl Opt*, 35:3925–3929, 1996.

D. Mendlovic, Z. Zalevsky, A. W. Lohmann, and R. G. Dorsch. Signal spatial-filtering using the localized fractional Fourier transform. *Opt Commun*, 126:14–18, 1996.

D. Mendlovic, Z. Zalevsky, D. Mas, J. García, and C. Ferreira. Fractional wavelet transform. *Appl Opt*, 36:4801–4806, 1997.

D. Mendlovic, D. Mas, A. W. Lohmann, Z. Zalevsky, and G. Shabtay. Fractional triple correlation and its applications. *J Opt Soc Am A*, 15:1658–1661, 1998.

D. Mihovilovic and R. N. Bracewell. Adaptive chirplet representation of signals on time-frequency plane. *Electronics Lett*, 27:1159–1161, 1991.

D. Mihovilović and R. N. Bracewell. Whistler analysis in the time-frequency plane using

chirplets. *J Geophysical Research*, 97:17199–17204, 1992.

W. Miller, Jr. *Lie Theory and Special Functions.* Academic Press, New York, 1968.

D. Mustard. Lie group imbeddings of the Fourier transform. School of Mathematics Preprint AM87/13, The University of New South Wales, Kensington, Australia, 1987.

D. A. Mustard. The fractional Fourier transform and a new uncertainty principle. School of Mathematics Preprint AM87/14, The University of New South Wales, Kensington, Australia, 1987.

D. Mustard. Lie group imbeddings of the Fourier transform and a new family of uncertainy principles. In *Proc Miniconf Harmonic Analysis Operator Algebras*, Australian National University, Canberra, 1987. Pages 211–222,

D. Mustard. The fractional Fourier transform and the Wigner distribution. School of Mathematics Preprint AM89/6, The University of New South Wales, Kensington, Australia, 1989.

D. Mustard. Fractional convolution. School of Mathematics Preprint AM90/26, The University of New South Wales, Kensington, Australia, 1990.

D. Mustard. Uncertainty principles invariant under the fractional Fourier transform. *J Australian Mathematical Society B*, 33:180–191, 1991.

D. Mustard. The fractional Fourier transform and the Wigner distribution. *J Australian Mathematical Society B*, 38:209–219, 1996.

D. Mustard. Fractional convolution. *J Australian Mathematical Society B*, 40:257–265, 1998.

V. Namias. The fractional order Fourier transform and its application to quantum mechanics. *J Inst Maths Applics*, 25:241–265, 1980.

V. Namias. Fractionalization of Hankel transforms. *J Inst Maths Applics*, 26:187–197, 1980.

L. Onural, M. F. Erden, and H. M. Ozaktas. Extensions to common Laplace and Fourier transforms. *IEEE Signal Processing Lett*, 4:310–312, 1997.

L. Onural. Diffraction from a wavelet point of view. *Opt Lett*, 18:846–848, 1993.

L. Onural, M. Kocatepe. Family of scaling chirp functions, diffraction, and holography. *IEEE Trans Signal Processing*, 43: 1568–1578, 1995.

L. Onural, M. Kocatepe, and H. M. Ozaktas. A class of wavelet kernels associated with wave propagation. In *Proc 1994 IEEE Int Conf Acoustics, Speech, Signal Processing*, IEEE, Piscataway, New Jersey, 1994. Pages III-9–III-12.

Y. Owechko. Time-frequency distributions using rotated-window spectrograms. In *Proc 1998 IEEE-SP Int Symp Time-Frequency Time-Scale Analysis*, IEEE, Piscataway, New Jersey, 1998. Pages 301–304.

H. M. Ozaktas. The fractional fourier-transform in optics. In *1994 OSA Annual Meeting Prog*, OSA, Washington DC, 1994. Page 173.

H. M. Ozaktas. Repeated fractional Fourier domain filtering is equivalent to repeated time and frequency domain filtering. *Signal Processing*, 54:81–84, 1996.

H. M. Ozaktas and O. Aytür. Fractional Fourier domains. *Signal Processing*, 46:119–124, 1995.

H. M. Ozaktas, B. Barshan, and D. Mendlovic. Convolution and Filtering in Fractional Fourier Domains. *Optical Review*, 1:15–16, 1994.

H. M. Ozaktas and M. F. Erden. Relationships among ray optical, Gaussian beam, and fractional Fourier transform descriptions of first-order optical systems. *Opt Commun*, 143:75–86, 1997.

H. M. Ozaktas, M. F. Erden, and M. A. Kutay. Signal processing with repeated filtering in fractional Fourier domains. *J Optoelectronics. Laser, supplement: Optics for Information Infrastructure*, 9:160–162, 1998.

H. M. Ozaktas, N. Erkaya, and M. A. Kutay. Effect of fractional Fourier transformation on time-frequency distributions belonging to the Cohen class. *IEEE Signal Processing Lett*, 3:40–41, 1996.

H. M. Ozaktas and M. A. Kutay. Time-Order Signal Representations. Technical Report BU-CE-0005, Bilkent University, Department of Computer Engineering, Ankara, January 2000.

H. M. Ozaktas and M. A. Kutay. Wigner distributions, linear canonical transforms, and phase-space optics. Technical Report BU-CE-0006, Bilkent University, Department of Computer Engineering, Ankara, January 2000.

H. M. Ozaktas, M. A. Kutay, and D. Mendlovic. Introduction to the fractional Fourier transform and its applications. Technical Report BU-CEIS-9802, Bilkent University, Department of Computer Engineering and Information Science, Ankara, January 1998.

H. M. Ozaktas, M. A. Kutay, and D. Mendlovic. Introduction to the fractional Fourier transform and its applications. In *Advances in Imaging and Electron Physics 106*, Academic Press, San Diego, California, 1999. Pages 239–291.

H. M. Ozaktas and D. Mendlovic. Fourier transforms of fractional order and their optical interpretation. *Opt Commun*, 101:163–169, 1993.

H. M. Ozaktas and D. Mendlovic. Fractional Fourier transforms and their optical implementation: II. *J Opt Soc Am A*, 10:2522–2531, 1993. Reprinted in *Selected Papers on Fourier Optics, SPIE Milestone Ser 105*, SPIE, Bellingham, Washington, 1995. Pages 434–443.

H. M. Özaktaş and D. Mendlovic. Kesirli Fourier dönüşümleri, Wigner dağılımları ile ilgisi, ve optik sinyal işlemeye uygulamaları. In *Sinyal İşleme ve Uygulamaları*, Boğaziçi University, Bebek, Istanbul, 1993. Pages 189–191. (In Turkish: Fractional Fourier transforms, their relationship to Wigner distributions, and applications to optical signal processing. In *Signal Processing and its Applications.*)

H. M. Ozaktas and D. Mendlovic. Fractional Fourier transform as a tool for analyzing beam propagation and spherical mirror resonators. *Opt Lett*, 19:1678–1680, 1994.

H. M. Ozaktas and D. Mendlovic. Fractional Fourier optics. *J Opt Soc Am A*, 12:743–751, 1995.

H. M. Ozaktas and D. Mendlovic. The fractional Fourier transform and its applications in optics and signal processing. *Optical Processing and Computing: SPIE Int Tech Working Group Newsletter*, October 1995, page 5.

H. M. Ozaktas and D. Mendlovic. Every Fourier optical system is equivalent to consecutive fractional-Fourier-domain filtering. *Appl Opt*, 35:3167–3170, 1996.

H. M. Ozaktas and D. Mendlovic. Applications of the fractional Fourier transform in optics and signal processing—a review. In *17th Congress ICO: Optics for Science and New Technology, SPIE Proc 2778*, SPIE, Bellingham, Washington, 1996. Pages 414–417.

H. M. Ozaktas, D. Mendlovic, and A. W. Lohmann. About the Wigner distribution of a graded index medium and the fractional Fourier transform operation. In *16th Congress ICO: Optics as a Key to High Technology, SPIE Proc 1983*, SPIE, Bellingham, Washington, 1993. Pages 407–408.

H. M. Ozaktas and D. A. B. Miller. Digital Fourier optics. *Appl Opt*, 35:1212–1219, 1996.

H. M. Ozaktas, B. Barshan, D. Mendlovic, and L. Onural. Convolution, filtering, and multiplexing in fractional Fourier domains and their relation to chirp and wavelet transforms. *J Opt Soc Am A*, 11:547–559, 1994.

H. M. Ozaktas, B. Barshan, L. Onural, and D. Mendlovic. Filtering in fractional Fourier domains and their relation to chirp transforms. In *Proc 7th Mediterranean Electrotechnical Conference*, IEEE, Piscataway, New Jersey, 1994. Pages 77–79.

H. M. Ozaktas, B. Barshan, D. Mendlovic, and H. Urey. Space-variant filtering in fractional Fourier domains. In *Optical Computing, Inst Phys Conf Ser 139*, Institute of Physics Publishing, Bristol, U.K., 1995. Pages 285–288.

H. M. Ozaktas, O. Arikan, M. A. Kutay, and G. Bozdağı. Digital computation of the fractional Fourier transform. *IEEE Trans Signal Processing*, 44:2141–2150, 1996.

A. K. Özdemir and O. Arıkan. Fast computation of the ambiguity function and the Wigner distribution on arbitrary line segments. *IEEE Trans Signal Processing*, 48, 2000.

C. Palma and V. Bagini. Extension of the Fresnel transform to $ABCD$ systems. *J Opt Soc Am A*, 14:1774–1779, 1997.

A. L. Patterson. *Z Krist*, 112:22, 1959.

S.-C. Pei and M.-H. Yeh. Improved discrete fractional Fourier transform. *Opt Lett*, 22:1047–1049, 1997.

S.-C. Pei and M.-H. Yeh. Two dimensional discrete fractional Fourier transform. *Signal Processing*, 67:99–108, 1998.

S.-C. Pei and M.-H. Yeh. Discrete fractional Hilbert transform. In *Proc 1998 IEEE Int Symp Circuits Systems*, IEEE, Piscataway, New Jersey, 1998. Pages IV-506–IV-509.

S.-C. Pei and M.-H. Yeh. Discrete fractional Hadamard transform. In *Proc 1999 IEEE Int Symp Circuits Systems*, IEEE, Piscataway, New Jersey, 1999. Pages III-179–III-182.

S.-C. Pei, M.-H. Yeh, and C.-C. Tseng. Discrete fractional Fourier transform based on orthogonal projections. *IEEE Trans Signal Processing*, 47:1335–1348, 1999.

S.-C. Pei, C.-C. Tseng, M.-H. Yeh, and J.-J. Shyu. Discrete fractional Hartley and Fourier transforms. *IEEE Trans Circuits Systems II*, 45:665–675, 1998.

P. Pellat-Finet. Fresnel diffraction and the fractional-order Fourier transform. *Opt Lett*, 19:1388–1390, 1994.

P. Pellat-Finet. Transfert du champ électromagnétique par diffraction et transformation

de Fourier fractionnaire. *C R Acad Sci Paris*, 320:91–97, 1995.

P. Pellat-Finet and G. Bonnet. Fractional order Fourier transform and Fourier optics. *Opt Commun*, 111:141–154, 1994.

P. Pellat-Finet and Y. Torres. Image formation with coherent light: the fractional Fourier transform approach. *J Mod Opt*, 44:1581–1594, 1997.

A. M. Pons, A. Lorente, C. Illueca, D. Mas, and J. M. Artigas. Fresnel diffraction in a theoretical eye: a fractional Fourier transform approach. *J Mod Opt*, 46:1043–1050, 1999.

M. A. Rashid. A simple proof of the result that the Wigner transformation if of finite order. *J Math Phys*, 30:1999–2000, 1989.

I. Raveh and D. Mendlovic. New properties of the radon transform of the cross Wigner ambiguity distribution function. *IEEE Trans Signal Processing*, 47:2077–2080, 1999.

M. G. Raymer, M. Beck, and D. F. McAlister. Complex wave-field reconstruction using phase-space tomography. *Phys Rev Lett*, 72:1137–1140, 1994.

M. G. Raymer, M. Beck, and D. McAlister. Spatial and temporal optical field reconstruction using phase-space tomography. In *Quantum Optics VI*, Springer, Berlin, 1994.

M. S. Richman and T. W. Parks. Understanding discrete rotations. In *Proc 1997 IEEE Int Conf Acoustics, Speech, Signal Processing*, IEEE, Piscataway, New Jersey, 1997. Pages 2057–2060.

B. Ristic and B. Boashash. Kernel design for time-frequency signal analysis using the Radon transform. *IEEE Trans Signal Processing*, 41,1996–2008, 1993.

B. Ruiz and H. Rabal. Fractional Fourier transform description with use of differential operators. *J Opt Soc Am A*, 14:2905–2913, 1997.

A. Sahin. *Two-Dimensional Fractional Fourier Transform and its Optical Implementation.* MS thesis, Bilkent University, Ankara, 1996.

A. Sahin, M. A. Kutay, and H. M. Ozaktas. Nonseparable two-dimensional fractional Fourier transform. *Appl Opt*, 37:5444–5453, 1998.

A. Sahin, H. M. Ozaktas, and D. Mendlovic. Optical implementation of the two-dimensional fractional Fourier transform with different orders in the two dimensions. *Opt Commun*, 120:134–138, 1995.

A. Sahin, H. M. Ozaktas, and D. Mendlovic. Optical implementations of two-dimensional fractional Fourier transforms and linear canonical transforms with arbitrary parameters. *Appl Opt*, 37:2130–2141, 1998.

T.-H. Sang, W. J. Williams, and J. C. O'Neill. An algorithm for positive time-frequency distributions. In *Proc 1996 IEEE-SP Int Symp Time-Frequency Time-Scale Analysis*, IEEE, Piscataway, New Jersey, 1996. Pages 165–168.

B. Santhanam and J. H. McClellan. The DRFT—a rotation in time-frequency space. In *Proc 1995 IEEE Int Conf Acoustics, Speech, Signal Processing*, IEEE, Piscataway, New Jersey, 1995. Pages 921–924.

B. Santhanam and J. H. McClellan. The discrete rotational Fourier transform. *IEEE Trans Signal Processing*, 44:994–998, 1996.

O. Seger. *Model Building and Restoration with Applications in Confocal Microscopy.* PhD

thesis, Linköping University, Sweden, 1993.

O. Seger and R. Lenz. Modelling of the point-spread function of laser scanning microscopes using canonical transforms. *J Visual Communication Image Representation*, 3:364–380, 1992.

J. Shamir and N. Cohen. Root and power transformations in optics. *J Opt Soc Am A*, 12:2415–2423, 1995.

C. J. R. Sheppard. Free-space diffraction and the fractional Fourier transform. *J Mod Opt*, 45:2097–2103, 1998.

C. J. R. Sheppard and K. G. Larkin. Similarity theorems for fractional Fourier transforms and fractional Hankel transforms. *Opt Commun*, 154:173–178, 1998.

J. T. Sheridan. General fractal preserving linear and non-linear transformations. *Optik*, 103:27–30, 1996.

C.-C. Shih. Optical interpretation of a complex-order Fourier transform. *Opt Lett*, 20:1178–1180, 1995.

C.-C. Shih. Fractionalization of Fourier transform. *Opt Commun*, 118:495–498, 1995.

S. G. Shin, S. I. Jin, S. Y. Shin, and S. Y. Lee. Optical neural network using fractional Fourier transform, log-likelihood, and parallelism. *Opt Commun*, 153:218–222, 1998.

R. Simon and N. Mukunda. Iwasawa decomposition in first-order optics: universal treatment of shape-invariant propagation for coherent and partially coherent beams. *J Opt Soc Am A*, 15:2146–2155, 1998.

D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani. Measurement of the Wigner distribution and the density matrix of a light mode using optical homodyne tomography: application to squeezed states and the vacuum. *Phys Rev Lett*, 70:1244–1247, 1993.

S. H. Song, S. Park, E.-H. Lee, P. S. Kim, and C. H. Oh. Planar optical implementation of multichannel fractional Fourier transforms. *Opt Commun*, 137:219–222, 1997.

S. H. Song, J.-S. Jeong, S. Park, and E.-H. Lee. Planar optical implementation of fractional correlation. *Opt Commun*, 143:287–293, 1997.

L. Stankovic and I. Djurovic. Relationship between the ambiguity function coordinate transformations and the fractional Fourier transform. *Annals Telecommunications*, 53: 316–319, 1998.

Status Report on the Fractional Fourier Transform. A. W. Lohmann, D. Mendlovic, and Z. Zalevsky, editors. Tel-Aviv University, Faculty of Engineering, Tel-Aviv, Israel, 1995.

M. E. Taylor. Noncommutative microlocal analysis, part I. *Memoirs AMS*, volume 52, number 313, 1984.

R. Tripathi, G. S. Pati, and K. Singh. A simple technique for space-variant target detection using fractional correlation. In *International Conference on Optics and Optoelectronics '98, SPIE Proc 3729*, SPIE, Bellingham, Washington, 1999. Pages 45–50.

J. Tu and S. Tamura. Wave field determination using tomography of the ambiguity function. *Phys Rev E*, 55:1946–1949, 1997.

J. H. Tu and S. Tamura. Analytic relation for recovering the mutual intensity by means of intensity information. *J Opt Soc Am A*, 15:202–206, 1998.

S. B. Tucker, J. Ojeda-Castañeda, and W. T. Cathey. Matrix description of near-field diffraction and the fractional Fourier transform. *J Opt Soc Am A*, 16: 316–322, 1999.

J. Turski. Harmonic analysis on $SL(2, c)$ and projectively adapted pattern representation. *J Fourier Anal Appl*, 4:67–91, 1998.

J. Uozumi and T. Asakura. Fractal Optics. In *Current Trends in Optics*. Academic Press, London, 1994. Chapter 6, pages 83–93.

J. C. Várilly and J. M. Gracia-Bondía. The Wigner transformation is of finite order. *J Math Phys*, 28:2390–2392, 1987.

N. Ja. Vilenkin. *Special Functions and the Theory of Group Representations*. Izd. Nauka., Moscow, 1965 (in Russian). English translation: AMS Transl. vol. 22, American Mathematical Society, Providence, Rhode Island, U.S.A., 1968.

K. Vogel and H. Risken. Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase. *Phys Rev A*, 40:2847–2849, 1989.

W. Vogel and W. Schleich. Phase distribution of a quantum state without using phase states. *Phys Rev A*, 44:7642–7646, 1991.

M. Wang, A. K. Chan, and C. K. Chui. Linear frequency-modulated signal detection using Radon-ambiguity transform. *IEEE Trans Signal Processing*, 46:571–586, 1998.

X. E. Wang and J. Zhou. Scaled fractional Fourier transform and optical systems. *Opt Commun*, 147:341–348, 1998.

A. Wawrzyñczyk. On the infinitesimal generator of the Fourier transform. *An Inst Mat Univ Nac Autónoma México*, 30:83–88, 1990.

F. B. Weissler. Two-point inequalities, the Hermite semigroup, and the Gauss-Weierstrass semigroup. *J Functional Analysis*, 32:102–121, 1979.

H. Weyl. Quantenmechanik und Gruppentheorie. *Ztsch Physik*, 46:1–47, 1927.

H. Weyl. *The Theory of Groups and Quantum Mechanics*, second edition. Dover, New York 1930.

N. Wiener. Hermitian polynomials and Fourier analysis. *J Math Phys (MIT)*, 8:70–73, 1929.

K. B. Wolf. *Integral Transforms in Science and Engineering*. Plenum Press, New York, 1979.

K. B. Wolf and G. Krötzsch. Metaxial correction of fractional Fourier transformers. *J Opt Soc Am A*, 16:821–830, 1999.

K. B. Wolf and A. L. Rivera. Holographic information in the Wigner function. *Opt Commun*, 144:36–42, 1997.

J. C. Wood and D. T. Barry. Tomographic time-frequency analysis and its application toward time-varying filtering and adaptive kernel design for multicomponent linear-FM signals. *IEEE Trans Signal Processing*, 42:2094–2104, 1994.

J. C. Wood and D. T. Barry. Linear signal synthesis using the Radon-Wigner transform. *IEEE Trans Signal Processing*, 42:2105–2111, 1994.

J. C. Wood and D. T. Barry. Radon transformation of time-frequency distributions for analysis of multicomponent signals. *IEEE Trans Signal Processing*, 42:3166–3177, 1994.

X.-G. Xia. On bandlimited signals with fractional Fourier transform. *IEEE Signal Processing Lett*, 3:72–74, 1996.

X.-G. Xia, Y. Owechko, B. H. Soffer, and R. M. Matic. On generalized-marginal time-frequency distributions. *IEEE Trans Signal Processing*, 44:2882–2886, 1996.

İ. Ş. Yetik, M. A. Kutay, H. Özaktaş, and H. M. Ozaktas. Continuous and discrete fractional Fourier domain decomposition. In *Proc 2000 IEEE Int Conf Acoustics, Speech, Signal Processing*, IEEE, Piscataway, New Jersey, 2000.

İ. Ş. Yetik, H. M. Özaktaş, B. Barshan, and L. Onural. Perspektif projeksiyonları ve kesirli Fourier dönüşümü. In *Sinyal İşleme ve Uygulamaları*, Istanbul University, Istanbul, 2000. (In Turkish: Perspective projections and the fractional Fourier transform. In *Signal Processing and its Applications.*)

H. Yoshimura and T. Iwai. Properties of the Gaussian Schell-model source field in a fractional Fourier plane. *J Opt Soc Am A*, 14:3388–3393, 1997.

H. Yoshimura and T. Iwai. Effect of lens aperture on the average intensity in a fractional Fourier plane. *Pure Appl Opt.* 7:1133–1141, 1998.

L. Yu, Y. Y. Lu, X. M. Zeng, M. C. Huang, M. Z. Chen, W. D. Huang, and Z. Z. Zhu. Deriving the integral representation of a fractional Hankel transform from a fractional Fourier transform. *Opt Lett*, 23:1158–1160, 1998.

L. Yu, W. D. Huang, M. C. Huang, Z. Z. Zhu, Z. M. Zeng, and W. Ji. The Laguerre-Gaussian series representation of two-dimensional fractional Fourier transform. *J Phys A*, 31:9353–9357, 1998.

B. Yurke, W. Schleich, and D. F. Walls. Quantum superpositions generated by quantum nondemolition measurements. *Phys Rev A*, 42:1703–1711, 1990.

Z. Zalevsky. *Unconventional Optical Processors for Pattern Recognition and Signal Processing.* PhD thesis, Tel-Aviv University, Tel-Aviv, 1996.

Z. Zalevsky. Applications of the fractional Fourier transform to correlation, feature extraction and pattern recognition. In *Proc 1999 IEEE-EURASIP Workshop Nonlinear Signal Image Processing*, Boğaziçi University, Bebek, Istanbul, 1999. Pages 579–584.

Z. Zalevsky and D. Mendlovic. Fractional Wiener filter. *Appl Opt*, 35:3930–3936, 1996.

Z. Zalevsky and D. Mendlovic. Fractional Radon transform: definition. *Appl Opt*, 35:4628–4631, 1996.

Z. Zalevsky and D. Mendlovic. Light propagation analysis in graded index fiber—review and applications. *Fiber and Integrated Optics*, 16:55–61, 1997.

Z. Zalevsky, D. Mendlovic, and J. H. Caulfield. Localized, partially space-invariant filtering. *Appl Opt*, 36:1086–1092, 1997.

Z. Zalevsky, D. Mendlovic, and J. H. Caulfield. Fractional correlator with real-time control of the space-invariance property. *Appl Opt*, 36:2370–2375, 1997.

Z. Zalevsky, D. Mendlovic, and R. G. Dorsch. Gerchberg-Saxton algorithm applied in the fractional Fourier or the Fresnel domain. *Opt Lett*, 21:842–844, 1996.

Z. Zalevsky, D. Mendlovic, and A. W. Lohmann. The ABCD-Bessel transformation. *Opt Commun*, 147:39–41, 1998.

Z. Zalevsky, D. Mendlovic, and H. M. Ozaktas. Energetic efficient synthesis of general mutual intensity distribution. *J Opt A*, 2:83–87, 2000.

Z. Zalevsky, I. Raveh, G. Shabtay, D. Mendlovic, and J. Garcia. Single-output color pattern recognition using a fractional correlator. *Opt Eng*, 36:2127–2136, 1997.

A. I. Zayed. On the relationship between the Fourier and fractional Fourier transforms. *IEEE Signal Processing Lett*, 3:310–311, 1996.

A. I. Zayed. A convolution and product theorem for the fractional Fourier transform. *IEEE Signal Processing Lett*, 5:101–103, 1998.

A. I. Zayed. Hilbert transform associated with the fractional Fourier transform. *IEEE Signal Processing Lett*, 5:206–208, 1998.

A. I. Zayed. Fractional Fourier transform of generalized functions. *Integral Transforms Special Functions*, 7:299–312, 1998.

Y. Zhang and B.-Y. Gu. Rotation-invariant and controllable space-variant optical correlation. *Appl Opt*, 37:6256–6261, 1998.

Y. Zhang, B.-Y. Gu, and G.-Z. Yang. Generation of self-fractional Hankel functions. *J Phys A*, 31:9769–9772, 1998.

Y. Zhang, B.-Y. Gu, B.-Z. Dong, G.-Z. Yang, H. Ren, X. Zhang, and S. Liu. Fractional Gabor transform. *Opt Lett*, 22:1583–1585, 1997.

Y. Zhang, B.-Y. Gu, B.-Z. Dong, and G.-Z. Yang. Optical implementations of the radon-Wigner display for one-dimensional signals. *Opt Lett*, 23:1126–1128, 1998.

Y. Zhang, B.-Z. Dong, B.-Y. Gu, and G.-Z. Yang. Beam shaping in the fractional Fourier transform domain. *J Opt Soc Am A*, 15:1114–1120, 1998.

Y. Zhang, B.-Y. Gu, B.-Z. Dong, and G.-Z. Yang. A new kind of windowed fractional transforms. *Opt Commun*, 152:127–134, 1998.

Y. Zhang, B.-Y. Gu, B.-Z. Dong, and G.-Z. Yang. New optical configurations for implementing Radon-Wigner display: matrix analysis approach. *Opt Commun*, 160:292–300, 1999.

Y. Zhang, B.-Y. Gu, B.-Z. Dong, and G.-Z. Yang. Rotationally symmetric beam shaping in the fractional Fourier transform domain. *Optik*, 110: 61–65, 1999.

*Advanced Topics in Shannon Sampling and Interpolation Theory.* R. J. Marks II, editor. Springer, New York, 1993.

R. P. Agarwal. *Difference Equations and Inequality Theory: Methods and Applications.* Marcel Dekker, New York, 1992.

A. N. Akansu and R. A. Haddad. *Multiresolution Signal Decomposition: Transforms, Subbands, and Wavelets.* Academic Press, London, 1992.

R. Aldrovandi and D. Galetti. On the structure of quantum phase space. *J Math Phys*, 31:2987–2995, 1990.

M. A. Alonso and G. W. Forbes. Generalization of Hamilton's formalism for geometrical optics. *J Opt Soc Am A*, 12:2744–2752, 1995.

J. A. Arnaud. Hamiltonian theory of beam mode propagation. In *Progress in Optics XI*, Elsevier, Amsterdam, 1973. Chapter VI, pages 247–304.

J. A. Arnaud. *Beam and Fiber Optics.* Academic Press, New York, 1976. Chapter 4, pages 220–324.

N. M. Atakishiev and S. K. Suslov. Difference analogs of the harmonic oscillator. *Theoretical Mathematical Physics*, 85:1055–1062, 1991.

N. M. Atakishiyev, S. M. Chumakov, and K. B. Wolf. Wigner distribution function for finite systems. *J Math Phys*, 39:6247–6261, 1998.

G. G. Athanasiu and E. G. Floratos. Coherent states in finite quantum mechanics. *Nuclear Phys B*, 425:343–364, 1994.

H. Bacry and M. Cadilhac. Metaplectic group and Fourier optics. *Phys Rev A*, 23:2533–2536, 1981.

R. Balian and C. Itzykson. Observations sur la mechanique quantique finie. *Compte Rendus Acad Sci*, 303:773–778, 1986.

R. Bamler and H. Glünder. The Wigner distribution function of two-dimensional signals: Coherent optical generation and display. *Optica Acta*, 30:1789–1803, 1983.

H. H. Barrett. The Radon transform and its applications. In *Progress in Optics XXI*, Elsevier, Amsterdam, 1984. Chapter 3, pages 217–286.

H. O. Bartelt, K.-H. Brenner, and A. W. Lohmann. The Wigner distribution function and its optical production. *Opt Commun*, 32:32–38, 1980.

M. J. Bastiaans. The Wigner distribution function applied to optical signals and systems. *Opt Commun*, 25:26–30, 1978.

M. J. Bastiaans. Wigner distribution function and its application to first-order optics. *J Opt Soc Am*, 69:1710–1716, 1979.

M. J. Bastiaans. The Wigner distribution function and Hamilton's characteristics of a geometric-optical system. *Opt Commun*, 30:321–326, 1979.

M. J. Bastiaans. Transport equations for the Wigner distribution function. *Optica Acta*, 26:1265–1272, 1979.

M. J. Bastiaans. Transport equations for the Wigner distribution function in an inhomogeneous and dispersive medium. *Optica Acta*, 26:1333–1344, 1979.

M. J. Bastiaans. Gabor's expansion of a signal into Gaussian elementary signals. *Proc IEEE*, 68:538–539,1980.

M. J. Bastiaans. A sampling theorem for the complex spectrogram, and Gabor's expansion of a signal in Gaussian elementary signals. *Opt Eng*, 20:594–598, 1981.

M. J. Bastiaans. The Wigner distribution function of partially coherent light. *Optica Acta*, 28:1215–1224, 1981.

M. J. Bastiaans. Gabor's signal expansion and degrees of freedom of a signal. *Optica Acta*, 28:1223–1229, 1982.

M. J. Bastiaans. Gabor's signal expansion and its relation to sampling of the sliding-window spectrum. In *Advanced Topics in Shannon Sampling and Interpolation Theory*, Springer, New York, 1993. Chapter 1, pages 1–35.

M. J. Bastiaans. On the sliding-window representation in digital signal processing. *IEEE Trans on Acoustics, Speech, and Signal Processing*, 33:868–873, 1985.

M. J. Bastiaans. Application of the Wigner distribution function to partially coherent light. *J Opt Soc Am A*, 3:1227–1238, 1986.

M. J. Bastiaans. Uncertainty principle and informational entropy for partially coherent light. *J Opt Soc Am A*, 3:1243–1246, 1986.

M. J. Bastiaans. Propagation laws for the second-order moments of the Wigner distribution function in first-order optical systems. *Optik*, 82:173–181, 1989.

M. J. Bastiaans. Gabor's signal expansion applied to partially coherent light. *Opt Commun*, 86:14–18, 1991.

M. J. Bastiaans. Second-order moments of the Wigner distribution function in first-order optical systems. *Optik*, 88:163–168, 1991.

M. J. Bastiaans. Gabor's signal expansion and the Zak transform. *Appl Opt*, 33:5241–5255, 1994.

M. J. Bastiaans. Applications of the Wigner distribution function in optics. In *The Wigner Distribution: Theory and Applications in Signal Processing*, Elsevier, Amsterdam, 1997. pages 375–426.

D. Basu and K. B. Wolf. The unitary irreducible representations of $SL(2, R)$ in all subgroup reductions. *J Math Phys*, 23:189–205, 1982.

W. Beckner. Inequalities in Fourier analysis. *Annals Mathematics*, 102:159–182, 1975.

M. Bendinelli, A. Consortini, L. Ronchi, and B. R. Frieden. Degrees of freedom. and eigenfunctions, for the noisy image. *J Opt Soc Am*, 64:1498–1502, 1974.

J. Bertrand and P. Bertrand. A class of affine Wigner functions with extended covariance properties. *J Math Phys*, 33:2515–2527, 1992. Erratum, 34:885, 1993.

G. Birkhoff and G. C. Rota. *Ordinary Differential Equations*. Wiley, New York, 1989.

B. Boashash. Note on the use of the Wigner distribution for time-frequency signal analysis. *IEEE Trans Acoustics, Speech, and Signal Processing*, 36:1518–1521, 1988.

B. G. Boone. *Signal Processing Using Optics*. Oxford University Press, New York, 1997.

M. Born and E. Wolf. *Principles of Optics*, sixth edition. Pergamon Press, Oxford, 1980.

R. W. Boyd. *Radiometry and the Detection of Optical Radiation*. Wiley, New York, 1983.

R. N. Bracewell. *The Fourier Transform and its Applications*, second edition. McGraw-Hill, New York, 1986.

R. N. Bracewell. *Two-Dimensional Imaging*. Prentice-Hall, Englewood Cliffs, New Jersey, 1995.

R. N. Bracewell. *The Fourier Transform and its Applications*, third edition. McGraw-Hill, New York, 1999.

H. J. Butterweck. General theory of linear, coherent, optical data-processing systems. *J Opt Soc Am*, 67:60–70, 1977.

H. J. Butterweck. Principles of optical data-processing. In *Progress in Optics XIX*, Elsevier, Amsterdam, 1981. Chapter 4, pages 211–280.

D.-W. Byun. Inversions of Hermite semigroup. *Proc Amer Math Soc*, 118:437–445, 1993.

C. E. Campbell. Ray vector fields. *J Opt Soc Am A*, 11:618–622, 1994.

M. J. Caola. Self-Fourier functions. *J Phys A*, 24:L1143–L1144. 1991.

C. Caratheodory. *Calculus of Variations and Partial Differential Equations of the First Order*. Chelsea Publishing, New York, 1965.

H. J. Carmichael. *Statistical Methods in Quantum Optics I: Master Equations and Fokker-Planck Equations*. Springer, Berlin, 1999.

W. T. Cathey. *Optical Information Processing and Holography*. Wiley, New York, 1974.

H. I. Choi and W. J. Williams. Improved time-frequency representation of multicomponent signals using exponential kernels. *IEEE Trans Acoustics, Speech, and Signal Processing*, 37:862–871, 1989.

D. Choudhury, P. N. Puntambekar, A. K. Chakraborty. Optical synthesis of self-Fourier functions. *Opt Commun*, 119:279–282, 1995.

C. K. Chui. *An Introduction to Wavelets*. Academic Press, London, 1992.

G. Cincotti, F. Gori, and M. Santarsiero. Generalized self-Fourier functions. *J Phys A*, 25:L1191–L1194, 1992.

T. A. C. M. Claasen and W. F. G. Mecklenbräuker. The Wigner distribution—a tool for time-frequency signal analysis. Part I: continuous-time signals. *Philips J Res*, 35:217–250, 1980.

T. A. C. M. Claasen and W. F. G. Mecklenbräuker. The Wigner distribution—a tool for time-frequency signal analysis. Part II: discrete-time signals. *Philips J Res*, 35:276–300, 1980.

T. A. C. M. Claasen and W. F. G. Mecklenbräuker. The Wigner distribution—a tool for time-frequency signal analysis. Part III: relations with other time-frequency signal transformations. *Philips J Res*, 35:372–389, 1980.

T. A. C. M. Claasen and W. F. G. Mecklenbräuker. The aliasing problem in discrete-time Wigner distributions. *IEEE Trans Acoustics, Speech, and Signal Processing*, 31:1067-1072, 1983.

M. W. Coffey. Self-reciprocal Fourier functions. *J Opt Soc Am A*, 11:2453–2455, 1994.

L. Cohen. Generalized phase-space distribution functions. *J Math Phys*, 7:781–786, 1966.

L. Cohen. Quantization problem and variational principle in the phase space formulation of quantum mechanics. *J Math Phys*, 17:1863–1866, 1976.

L. Cohen. Time-frequency distributions—a review. *Proc IEEE*, 77:941–981, 1989.

L. Cohen. *Time-Frequency Analysis*. Prentice-Hall, Englewood Cliffs, New Jersey, 1995.

C. Cohen-Tannoudji, B. Diu, and F. Laloë. *Quantum Mechanics*. Wiley, New York, 1977. 2 volumes.

I. Daubechies. The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans Information Theory*, 36:961–1006, 1990.

I. Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1992.

L. Debnath and P. Mikusiński. *Introduction to Hilbert Spaces with Applications*. Academic Press, San Diego, 1990.

A. J. van Dekker and A. van den Bos. Resolution: a survey. *J Opt Soc Am A*, 14:547–557, 1997.

G. A. Deschamps. Ray Techniques in Electromagnetics. *Proc IEEE*, 60:1022–1035, 1972.

B. W. Dickinson and K. Steiglitz. Eigenvectors and functions of the discrete Fourier transform. *IEEE Trans Acoustics, Speech, and Signal Processing*, 30:25–31, 1982.

D. Dragoman. Higher-order moments of the Wigner distribution function in first-order optical systems. *J Opt Soc Am A*, 11:2643–2646, 1994.

D. Dragoman. The Wigner distribution function in optics and optoelectronics. In *Progress in Optics XXXVII*, Elsevier, Amsterdam, 1997. Chapter I, pages 1–56.

A. J. Dragt and J. M. Finn. Lie series and invariant functions for analytic symplectic maps. *J Math Phys*, 17:2215–2227, 1976.

A. J. Dragt. Lie algebraic theory of geometrical optics and optical aberrations. *J Opt Soc Am*, 72:372–379, 1982.

D. E. Dudgeon and R. M. Mersereau. *Multidimensional Digital Signal Processing.* Prentice-Hall, Englewood Cliffs, New Jersey, 1984.

H. Dym and H. P. McKean. *Fourier Series and Integrals.* Academic Press, San Diego, 1972.

R. L. Easton, Jr., A. J. Ticknor, and H. H. Barrett. Application of the Radon transform to optical production of the Wigner distribution function. *Opt Eng*, 23:738–744, 1984.

N. Engheta. A note on fractional calculus and the image method for dielectric spheres. *J Electromagnetic Waves Applic*, 9:1179–1188, 1995.

N. Engheta. On fractional calculus and fractional multipoles in electromagnetism. *IEEE Trans Antennas Propagation*, 44:554–566, 1996.

N. Engheta. Electrostatic "fractional" image methods for perfectly conducting wedges and cones. *IEEE Trans Antennas Propagation*, 44:1565–1574, 1996.

N. Engheta. Use of fractional integration to propose some "fractional" solutions for the scalar Helmholtz equation. *Prog Electromagnetics Research*, 12:107–132, 1996.

N. Engheta. On the role of fractional calculus in electromagnetic theory. *IEEE Antennas and Propagation Magazine*, August 1997, pages 35–46.

N. Engheta. Fractional curl operator in electromagnetics. *Microwave and Optical Tech Lett*, 17:86–91, 1998.

M. F. Erden and H. M. Ozaktas. Accumulated Gouy phase shift in Gaussian beam propagation through first-order optical systems. *J Opt Soc Am A*, 14:2190–2194, 1997.

M. Ferraro and T. M. Caelli. Relationship between integral transform invariances and Lie group theory. *J Opt Soc Am A*, 5:738–742, 1988.

P. Flandrin. *Temps-fréquence.* Hermès, Paris, 1993.

G. B. Folland. *Harmonic Analysis in Phase Space.* Princeton University Press, Princeton, New Jersey, 1989.

A. T. Friberg. Phase-space methods for partially coherent wavefields. In *Optics in Four Dimensions—1980*, American Institute of Physics, New York, 1981. Pages 313–331.

A. T. Friberg. Energy transport in optical systems with partially coherent light. *Appl Opt*, 25:4547–4556, 1986.

B. R. Frieden. Evaluation, design, and extrapolation methods for optical signals, based

on use of the prolate functions. In *Progress in Optics IX*, Elsevier, Amsterdam, 1971. Chapter 8, pages 311–407.

D. Gabor. Light and information. In *Progress in Optics I*, Elsevier, Amsterdam, 1961. Chapter 4, pages 109–153.

D. Galetti and A. F. R. de Toledo Piza. An extended Weyl-Wigner transformation for special finite spaces. *Physica A*, 149:267–282, 1988.

G. García-Calderón and M. Moshinsky. Wigner distribution functions and the representation of canonical transformations in quantum mechanics. *J Phys A*, 13:L185–L188, 1980.

C. W. Gardiner. *Quantum Noise*. Springer, Berlin, 1991.

J. D. Gaskill. *Linear Systems, Fourier Transforms, and Optics*. Wiley, New York, 1978.

A. Ghatak and K. Thyagarajan. Graded index optical waveguides: a review. In *Progress in Optics XVIII*, Elsevier, Amsterdam, 1980. Chapter 1, pages 1–126.

R. Gilmore. Baker-Campbell-Hausdorff formulas. *J Math Phys*, 15:2090–2092, 1974.

H. Goldstein. *Classical Mechanics*, second edition. Addison-Wesley, Reading, Massachusetts, 1980.

J. W. Goodman. *Statistical Optics*. Wiley, New York, 1985.

J. W. Goodman. *Introduction to Fourier Optics*, second edition. McGraw-Hill, New York, 1996.

F. Gori. Sampling in optics. In *Advanced Topics in Shannon Sampling and Interpolation Theory*, Springer, New York, 1993. Chapter 2, pages 37–83.

F. Gori. Why is the Fresnel transform so little known? In *Current Trends in Optics*, Academic Press, London, 1994. Chapter 10, pages 139–148.

F. Gori and G. Guattari. Effects of coherence on the degrees of freedom of an image. *J Opt Soc Am*, 61:36–39, 1971.

F. Gori and G. Guattari. Shannon number and degrees of freedom of an image. *Opt Commun*, 7:163–165, 1973.

F. Gori, M. Santarsiero, and G. Guattari. Coherence and the spatial distribution of intensity. *J Opt Soc Am A*, 10:673–679, 1993.

F. A. Grünbaum. The eigenvectors of the discrete Fourier transform: a version of the Hermite functions. *J Math Anal Applic*, 88:355–363, 1982.

V. Guillemin and S. Sternberg. The metaplectic representation, Weyl operators, and spectral theory. *J Funct Anal*, 42:128–225, 1981.

V. Guillemin and S. Sternberg. *Symplectic Techniques in Physics*. Cambridge University Press, Cambridge, 1984.

T. Hakioğlu. Finite-dimensional Schwinger basis, deformed symmetries, Wigner function, and an algebraic approach to quantum phase. *J Phys A*, 31:6975–6994, 1998.

T. Hakioğlu. Linear canonical transformations and quantum phase: a unified canonical and algebraic approach. *J Phys A*, 32:4111–4130, 1999.

G. Hazak. Comment on "Wave field determination using three-dimensional intensity information." *Phys Rev Lett*, 69:2874, 1992.

E. Hecht, A. Zajac, and K. Guardino. *Optics*, third edition. Addison-Wesley, Reading, Massachusetts, 1997.

F. B. Hildebrand. *Methods of Applied Mathematics*, second edition. Prentice-Hall, Englewood Cliffs, New Jersey, 1965.

F. Hildebrand. *Finite-Difference Equations and Simulations.* Prentice-Hall, Englewood Cliffs, New Jersey, 1968.

E. Hille and R. S. Phillips. *Functional Analysis and Semi-Groups*, revised edition. American Mathematical Society, Providence, Rhode Island, U.S.A., 1957.

M. Hillery, R. F. O'Connell, M. O. Scully, and E. P. Wigner. Distribution functions in physics: fundamentals. *Phys Rep*, 106:121–167, 1984.

F. Hlawatsch and G. F. Boudreaux-Bartels, Linear and quadratic time-frequency signal representations. *IEEE Signal Processing Magazine*, April 1992, pages 21–67.

F. Hlawatsch and W. Kozek. The Wigner distribution of a linear signal space. *IEEE Trans Signal Processing*, 41:1248–1258, 1993.

J. L. Horner. Light utilization in optical correlators. *Appl Opt*, 21:4511–4514, 1982.

J. L. Horner. Clarification of Horner efficiency. *Appl Opt*, 31:4629, 1992.

K. Iizuka. *Engineering Optics*, second edition. Springer, Berlin, 1987.

*Image Recovery: Theory and Applications.* H. Stark, editor. Academic Press, Orlando, 1987.

V. Yu. Ivanov, V. P. Sivokon, and M. A. Vorontsov. Phase retrieval from a set of intensity measurements: theory and experiment. *J Opt Soc Am A*, 9:1515—1524, 1992.

A. K. Jain. *Fundamentals of Digital Image Processing.* Prentice-Hall, Englewood Cliffs, New Jersey, 1989.

A. J. E. M. Janssen. The Zak transform: A signal transform for sampled time-continuous signals. *Philips J Res*, 43:23–69, 1988.

B. Javidi. Nonlinear joint power spectrum based optical correlation. *Appl Opt*, 28:2358–2367, 1989.

B. Javidi and J. L. Horner. *Real-Time Optical Information Processing*, Academic Press, San Diego, 1994.

F. A. Jenkins and H. E. White. *Fundamentals of Optics*, fourth edition. McGraw-Hill, New York, 1976.

M. Kauderer. Fourier-optics approach to the symplectic group. *J Opt Soc Am A*, 7:231–239, 1990.

W. G. Kelley and A. C. Peterson. *Difference Equations: An Introduction with Applications.* Academic Press, San Diego, 1991.

M. V. Klein and T. E. Furtak. *Optics*, second edition. Wiley, New York, 1986.

M. Kline and I. W. Kay. *Electromagnetic Theory and Geometrical Optics.* Wiley, New York, 1965.

B. H. Kolner and M. Nazarathy. Temporal imaging with a time lens. *Opt Lett*, 14:630–632, 1989. Erratum, 15:655, 1990.

B. V. K. V. Kumar and L. Hassebrook. Performance measures for correlation filters. *Appl*

*Opt*, 29:2997–3006, 1990.

A. Lakhtakia. Fractal self-Fourier functions. *Optik*, 94:51–52, 1993.

C. Lanczos. *The Variational Principles of Mechanics*, fourth edition. Dover, New York, 1970.

H. J. Landau. On the density of phase-space expansions. *IEEE Trans Information Theory*, 39:1152–1156, 1993.

H. J. Landau and H. O. Pollak. Prolate spheroidal wave functions, Fourier analysis and yncertainty—II. *The Bell System Technical J*, 40:65–84, 1961.

U. Leonhardt. Discrete Wigner function and quantum-state tomography. *Phys Rev A*, 53:2998–3015, 1996.

F. L. Lewis. *Optimal Estimation.* Wiley, New York, 1986.

Y. Li and Y. Sheng. Wavelets, optics, and pattern recognition. In *Optical Pattern Recognition*, Cambridge University Press, Cambridge, 1998. Chapter 3, pages 64–88.

*Lie Methods in Optics.* J. S. Mondragón and K. B. Wolf, editors. Springer, Berlin, 1986.

*Lie Methods in Optics II.* K. B. Wolf, editor. Springer, Berlin, 1989.

S. G. Lipson. Self-Fourier objects and other self-transform objects: comment. *J Opt Soc Am A*, 10:2088–2089, 1993.

S. G. Lipson and H. Lipson. *Optical Physics*, second edition. Cambridge University Press, Cambridge, 1981.

A. Lohmann. Ein neues Dualitätsprinzip in der Optik. *Optik*, 11:478–488, 1954. An English version appeared as: Duality in Optics. *Optik*, 89:93–97, 1992.

A. W. Lohmann. *Optical Information Processing*, lecture notes. Optik+Info, Post Office Box 51, Uttenreuth, Germany, 1986.

A. W. Lohmann and D. Mendlovic. Self-Fourier objects and other self-transform objects. *J Opt Soc Am A*, 9:2009–2012, 1992.

A. W. Lohmann and D. Mendlovic. An optical self-transform with odd cycles. *Opt Commun*, 93:25–26, 1992.

A. W. Lohmann and D. Mendlovic. Temporal filtering with time lenses. *Appl Opt*, 31:6212–6219, 1992.

A. W. Lohmann and D. Mendlovic. Image formation of a self-Fourier object. *Appl Opt*, 33:153–157, 1994.

A. W. Lohmann and H. W. Werlich. Incoherent matched filter with Fourier holograms. *Appl Opt*, 7:561–563, 1968.

A. Luis and J. Peřina. Discrete Wigner function for finite-dimensional systems. *J Phys A*, 31:1423–1441, 1998.

R. K. Luneburg. *Mathematical Theory of Optics.* University of California Press, Berkeley, 1964.

S. G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 11:674–693, 1989.

S. Mallat. *A Wavelet Tour of Signal Processing.* Academic Press, London, 1998.

L. Mandel and E. Wolf. *Optical Coherence and Quantum Optics.* Cambridge University

Press, Cambridge, 1995.

D. Marcuse. *Light Transmission Optics*, second edition. Krieger Publishing, Malabar, Florida, 1982.

R. J. Marks II. *Introduction to Shannon Sampling and Interpolation Theory.* Springer, New York, 1991.

J. H. McClellan. Comments on "Eigenvalue and eigenvector decomposition of the discrete Fourier transform." *IEEE Trans Audio Electroacoustics*, 21:65, 1973.

J. H. McClellan and T. W. Parks. Eigenvalue and eigenvector decomposition of the discrete Fourier transform. *IEEE Trans Audio Electroacoustics*, 20:66–74, 1972.

N. W. McLachlan. *Theory and Applications of Mathieu Functions.* Dover, New York, 1964.

D. Mendlovic, N. Konforti, and E. Marom. Shift and projection invariant pattern recognition using logarithmic harmonics. *Appl Opt*, 29:4784–4789, 1990.

D. Mendlovic and A. W. Lohmann. Space-bandwidth product adaptation and its application to superresolution: fundamentals. *J Opt Soc Am A*, 14:558–562, 1997.

D. Mendlovic, A. W. Lohmann, and Z. Zalevsky. Space-bandwidth product adaptation and its application to superresolution: examples. *J Opt Soc Am A*, 14:563–567, 1997. (Mendlovic and Ozaktas 1993c)

D. Mendlovic and H. M. Ozaktas. Optical-coordinate transformation methods and optical-interconnection architectures. *Appl Opt*, 32:5119–5124, 1993.

D. Mendlovic and Z. Zalevsky. Definition, properties and applications of the generalized temporal-spatial Wigner distribution function. *Optik*, 107:49–56, 1997.

D. A. B. Miller. Spatial channels for communicating with waves between volumes. *Opt Lett*, 23:1645–1647, 1998.

K. S. Miller and B. Ross. Fractional difference calculus. In *Univalent Functions, Fractional Calculus, and their Applications*, Ellis Horwood, Chichester, U.K., 1989. Chapter 12, pages 139–152.

K. D. Möller. *Optics.* University Science Books, Mill Valley, California, 1988.

M. Moshinsky and C. Quesne. Linear canonical transformations and their unitary representations. *J Math Phys*, 12:1772–1780, 1971.

M. Moshinsky, T. H. Seligman, and K. B. Wolf. Canonical transformations and the radial oscillator and Coulomb problem. *J Math Phys*, 13:901–907, 1972.

M. Navarro-Saad and K. B. Wolf. The group theoretical treatment of aberrating systems. I. Aligned lens systems in third aberration order. *J Math Phys*, 27:1449–1457, 1986.

M. Navarro-Saad and K. B. Wolf. Factorization of the phase-space transformation produced by an arbitrary refracting surface. *J Opt Soc Am A*, 3:340–346, 1986.

A. W. Naylor and G. R. Sell. *Linear Operator Theory in Engineering and Science.* Springer, New York, 1982.

M. Nazarathy and J. Shamir. Fourier optics described by operator algebra. *J Opt Soc Am*, 70:150–159, 1980.

M. Nazarathy and J. Shamir. Holography described by operator algebra. *J Opt Soc Am*, 71:529–541, 1981.

M. Nazarathy and J. Shamir. First-order optics—a canonical operator representation: lossless systems. *J Opt Soc Am*, 72:356–364, 1982.

M. Nazarathy and J. Shamir. First-order optics—operator representation for systems with loss or gain. *J Opt Soc Am*, 72:1398–1408, 1982.

M. Nazarathy, A. Hardy, and J. Shamir. Generalized mode propagation in first-order optical systems with loss or gain. *J Opt Soc Am*, 72:1409–1420, 1982.

M. Nazarathy, A. Hardy, and J. Shamir. Misaligned first-order optics: canonical operator theory. *J Opt Soc Am A*, 3:1360–1369, 1986.

K. A. Nugent. Wave field determination using three-dimensional intensity information. *Phys Rev Lett*, 68:2261–2264, 1992.

J. Ojeda-Castañeda and A. Noyola-Isgleas. Differential operators for scalar wave propagation. *J Opt Soc Am A*, 5:1605–1609, 1988.

J. C. O'Neill, P. Flandrin, and W. J. Williams. On the existence of discrete Wigner distributions. *IEEE Signal Processing Lett*, 6:304–306, 1999.

L. Onural and M. T. Özgen. Extraction of three-dimensional object location information directly from in-line holograms using Wigner analysis. *J Opt Soc Am A*, 9:252–260, 1992.

Alan V. Oppenheim and Ronald W. Shafer. *Discrete-Time Signal Processing.* Prentice-Hall, Englewood Cliffs, New Jersey, 1989.

V. F. Osipov. *Pochti periodicheskie funktsii Bora-Frenelya.* University of St. Petersburg, St. Petersburg, 1992. (In Russian: Bohr-Fresnel Almost-Periodic Functions.)

H. M. Ozaktas, K.-H. Brenner, and A. W. Lohmann. Interpretation of the space-bandwidth product as the entropy of distinct connection patterns in multifacet optical interconnection architectures. *J Opt Soc Am A*, 10:418–422, 1993.

H. M. Ozaktas and D. Mendlovic. Multi-stage optical interconnection architectures with least possible growth of system size. *Opt Lett*, 18:296–298, 1993.

H. M. Ozaktas and M. C. Nuss. Time-variant linear pulse processing. *Opt Commun*, 131:114–118, 1996.

H. M. Ozaktas and H. Urey. Space-bandwidth product of conventional Fourier transforming systems. *Opt Commun*, 104:29–31, 1993.

H. M. Ozaktas, H. Urey, and A. W. Lohmann. Scaling of diffractive and refractive lenses for optical computing and interconnections. *Appl Opt*, 33:3782–3789, 1994.

A. Papoulis. *Systems and Transformations with Applications in Optics.* McGraw-Hill, New York, 1968.

A. Papoulis. Ambiguity function in Fourier optics. *J Opt Soc Am*, 64:779–788, 1974.

A. Papoulis. *Signal Analysis.* McGraw-Hill, New York, 1977.

A. Papoulis. *Probability, Random Variables, and Stochastic Processes*, third edition. McGraw-Hill, New York, 1991.

A. Papoulis. Pulse-compression, fiber communications, and diffraction—a unified approach. *J Opt Soc Am A*, 11:3–13, 1994.

K. Patorski. The self-imaging phenomenon and its applications. In *Progress in Optics XXVII*, Elsevier, Amsterdam, 1989. Chapter 1, pages 1–108.

R. J. Pegis. The modern development of Hamiltonian optics. In *Progress in Optics I*, Elsevier, Amsterdam, 1961. Chapter VI, pages 1–29.

A. Perelomov. *Generalized Coherent States and Their Applications.* Springer, Berlin, 1986.

F. Peyrin and R. Prost. A unified definition for the discrete-time, discrete-frequency, and discrete-time/frequency Wigner distributions. *IEEE Trans Acoustics, Speech, and Signal Processing*, 34:858–866, 1986.

S. Qian and D. Chen. *Joint Time-Frequency Analysis.* Prentice-Hall, Englewood Cliffs, New Jersey, 1996.

C. Quesne and M. Moshinsky. Canonical transformations and matrix elements. *J Math Phys*, 12:1780–1783, 1971.

R. Rammal and J. Bellissard. An algebraic semi-classical approach to Bloch electrons in a magnetic field. *J Phys France*, 51:1803–1830, 1990.

S. Ramo, J. R. Whinnery, and T. Van Duzer. *Fields and Waves in Communication Electronics*, third edition. Wiley, New York, 1994.

P. J. Reardon and R. A. Chipman. Maximum power of refractive lenses: a fundamental limit. *Opt Lett*, 15:1409-1411, 1990.

G. O. Reynolds, J. B. DeVelis, G. B. Parrent, Jr., and B. J. Thompson. *The New Physical Optics Notebook: Tutorials in Fourier Optics.* SPIE, Bellingham, Washington, 1989.

W. T. Rhodes. Simple procedure for the analysis of coherent imaging systems. *Opt Lett*, 19:1559–1561, 1994.

W. T. Rhodes. The fallacy of single-lens coherent imaging theory. In *1998 OSA Annual Meeting Prog*, OSA, Washington DC, 1998. Page 117.

M. S. Richman, T. W. Parks, and R. G. Shenoy. Discrete-time, discrete-frequency, time-frequency analysis. *IEEE Trans Signal Processing*, 46:1517–1527, 1998.

O. Rioul and M. Vetterli. Wavelets and signal processing. *IEEE Signal Processing Magazine*, October 1991, pages 14–38.

S. Roman. *Advanced Linear Algebra.* Springer, New York, 1992.

B. E. A. Saleh and N. S. Subotic. Time-variant filtering of signals in the mixed time-frequency domain. *IEEE Trans Acoustics, Speech, and Signal Processing*, 33:1479–1485, 1985.

B. E. A. Saleh and M. C. Teich. *Fundamentals of Photonics.* Wiley, New York, 1991.

T. S. Santhanam and A. R. Tekumalla. Quantum mechanics in finite dimensions. *Foundations Physics*, 6:583-587, 1976.

W. P. Schleich, E. Mayr, and D. Krähmer. *Quantum Optics in Phase Space.* Wiley, New York, 1999.

I. E. Segal. Transforms for operators and symplectic automorphisms over a locally compact Abelian group. *Math Scand*, 13:31–43, 1963.

T. Sekiguchi and K. B. Wolf. The Hamiltonian formulation of optics. *Am J Phys*, 55:830–835, 1987.

J. Shamir. Cylindrical lens systems described by operator algebra. *Appl Opt*, 18:4195–4202, 1979.

A. E. Siegman. *Lasers.* University Science Books, Mill Valley, California, 1986.

R. Simon and K. B. Wolf. Structure of the set of paraxial optical systems. *J Opt Soc Am A*, 17:342–355, 2000.

D. Slepian. Prolate spheroidal wave functions, Fourier analysis and uncertainty—IV: extensions to many dimensions; generalized prolate spheroidal functions. *The Bell System Technical J*, 43:3009–3058, 1964.

D. Slepian. Prolate spheroidal wave functions, Fourier analysis and uncertainty—V. *The Bell System Technical J*, 57:1371–1430, 1978.

D. Slepian and H. O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty—I. *The Bell System Technical J*, 40:43–63, 1961.

D. Slepian and H. O. Pollack. Prolate spheroidal wave functions, Fourier analysis and uncertainty—III: the dimension of essentially time- and band-limited signals. *The Bell System Technical J*, 41:1295–1336, 1962.

S. Solimeno, B. Crosignani, and P. Di Porto. *Guiding, Diffraction, and Confinement of Optical Radiation.* Academic Press, Orlando, Florida, 1986.

O. N. Stavroudis. *The Optics of Rays, Wavefronts, and Caustics.* Academic Press, New York, 1972.

D. Stoler. Operator methods in physical optics. *J Opt Soc Am*, 71:334–341, 1981.

G. Strang. *Linear Algebra and its Applications*, third edition. Harcourt Brace Jovanovich, New York, 1988.

G. Strang and T. Nguyen. *Wavelets and Filter Banks.* Wellesley-Cambridge Press, Wellesley, Massachusetts, 1996.

B. W. Suter. *Multirate and Wavelet Signal Processing.* Academic Press, London, 1997.

J. L. Synge. *Geometrical Optics: An Introduction to Hamilton's Method.* Cambridge University Press, Cambridge, 1937.

J. L. Synge. *Geometrical Mechanics and de Broglie Waves.* Cambridge University Press, Cambridge, 1954.

K. Tanaka. Paraxial theory in optical design in terms of Gaussian brackets. In *Progress in Optics XXIII*, Elsevier, Amsterdam, 1986. Chapter 2, pages 63–111.

M. Testorf and J, Ojeda-Castañeda. Fractional Talbot effect: analysis in phase space. *J Opt Soc Am A*, 13:119–125, 1996.

G. Toraldo di Francia. Resolving power and information. *J Opt Soc Am*, 45:497–501, 1955.

G. Toraldo di Francia. Degrees of freedom of an image. *J Opt Soc Am*, 59:799–804, 1969.

*The Transforms and Applications Handbook*, second edition. A. Poularikas, editor. CRC Press, Boca Raton, Florida, 2000.

D. E. Vakman. *Sophisticated Signals and the Uncertainty Principle in Radar.* Springer, Berlin, 1968.

A. VanderLugt. Signal detection by complex spatial filtering. *IEEE Trans Information Theory*, 10:139–146, 1964.

A. VanderLugt. *Optical Signal Processing.* Wiley, New York, 1992.

M. Vetterli and J. Kovacevic. *Wavelets and Subband Coding.* Prentice-Hall, Englewood

Cliffs, New Jersey, 1995.

D. F. Walls and G. J. Milburn. *Quantum Optics.* Springer, Berlin, 1994.

G. G. Walter. *Wavelets and Other Orthogonal Systems with Applications.* CRC Press, Boca Raton, Florida, 1994.

A. Walther. Gabor's theorem and energy transfer through lenses. *J Opt Soc Am*, 57:639–644, 1967.

A. Walther. Radiometry and coherence. *J Opt Soc Am*, 58:1256–1259, 1968.

*Wavelets: Mathematics and Applications.* J. J. Benedetto and M. W. Frazier, editors. CRC Press, Boca Raton, Florida, 1993.

C. S. Weaver and J. W. Goodman. A technique for optically convolving two functions. *Appl Opt*, 5:1248–1249, 1966.

A. Weil. Sur certains groupes d'opérateurs unitaires. *Acta Mathematica*, 111:143–211, 1964.

P. B. Wiegmann and A. V. Zabrodin. Algebraization of difference eigenvalue equations related to $U_q(sl_2)$. *Nuclear Phys B*, 451:699–724, 1995.

N. Wiener. *The Fourier Integral and Certain of its Applications.* Cambridge University Press, Cambridge, 1933.

E. Wigner. On the quantum correction for thermodynamic equilibrium. *Phys Rev*, 40:749–759, 1932.

*The Wigner Distribution: Theory and Applications in Signal Processing.* W. Mecklenbräuker and F. Hlawatsch, editors. Elsevier, Amsterdam, 1997.

*Wigner Distributions and Phase Space in Optics 2000.* G. W. Forbes, V. I. Man'ko, H. M. Ozaktas, R. Simon, and K. B. Wolf, editors. Feature issue of *J Opt Soc Am A*, 17 (12), December 2000.

R. M. Wilcox. Exponential operators and parameter differentiation in quantum physics. *J Math Phys*, 8:962–982, 1967.

J. H. Wilkinson. *The Algebraic Eigenvalue Problem.* Oxford University Press, Oxford, 1988.

J. T. Winthrop. Propagation of structural information in optical wave fields. *J Opt Soc Am*, 61:15–30, 1971.

K. B. Wolf. Canonical transforms. I. Complex linear transforms. *J Math Phys*, 15:1295–1301, 1974.

K. B. Wolf. Canonical transforms. II. Complex radial transforms. *J Math Phys*, 15:2102–2111, 1974.

K. B. Wolf. Canonical transforms. separation of variables, and similarity solutions for a class of parabolic differential equations. *J Math Phys*, 17:601–613, 1976.

K. B. Wolf. On self-reciprocal functions under a class of integral transforms. *J Math Phys*, 18:1046–1051, 1977.

K. B. Wolf. *Integral Transforms in Science and Engineering.* Plenum Press, New York, 1979.

K. B. Wolf. Symmetry in Lie optics. *Annals of Physics*, 172:1–25, 1986.

K. B. Wolf. The group theoretical treatment of aberrating systems. II. Axis-symmetric

inhomogeneous systems and fiber optics in third aberration order. *J Math Phys*, 27:1458–1465, 1986.

K. B. Wolf. Fourier transform in metaxial geometric optics. *J Opt Soc Am A*, 8:1399—1403, 1991.

K. B. Wolf. Relativistic aberration of optical phase space. *J Opt Soc Am A*, 10:1925–1934, 1993.

K. B. Wolf. Wigner distribution function for paraxial polychromatic optics. *Opt Commun*, 132:343–352, 1996.

K. B. Wolf and G. Krötzsch. Geometry and dynamics in refracting systems. *Eur J Phys*, 16:14–20, 1995.

K. B. Wolf and E. V. Kurmyshev. "Squeezed states" in Helmholtz optics. *Phys Rev A*, 47:3365–3370, 1993.

E. M. Wright and J. C. Garrison. Path-integral derivation of the complex $ABCD$ Huygens integral. *J Opt Soc Am A*, 4:1751–1755, 1987.

Y. Yamamoto and A. İmamoğlu. *Mesoscopic Quantum Optics*. Wiley, New York, 1999.

A. Yariv. *Quantum Electronics*, third edition. Wiley, New York, 1989.

A. Yariv. *Optical Electronics in Modern Communications*, fifth edition. Oxford University Press, Oxford, 1997.

R. Yarlagadda. A note on the eigenvectors of DFT matrices. *IEEE Trans Acoustics, Speech, and Signal Processing*, 25:586–589, 1977.

K. Yosida. *Operational Calculus: A Theory of Hyperfunctions*. Springer, New York, 1984.

F. T. S. Yu. *Optical Information Processing*. Wiley, New York, 1983.

F. T. S. Yu and S. Jutamulia. *Optical Signal Processing, Computing, and Neural Networks*. Wiley, New York, 1992.

*Optical Pattern Recognition*. F. T. S. Yu and S. Jutamulia, editors. Cambridge University Press, Cambridge, 1998.

Y. M. Zhu, F. Peyrin, and R. Goutte. Equivalence between two-dimensional analytic and real signal Wigner distribution. *IEEE Trans Acoustics, Speech, and Signal Processing*, 37:1631–1634, 1989.