Field Guide to
Linear Systems in Optics

J. Scott Tyo
Andrey S. Alenin

SPIE Terms of Use: This SPIE eBook is DRM-free for your convenience. You may install this eBook on any device you own, but not post it publicly or transmit it to others. SPIE eBooks are for personal use only. For details, see the SPIE Terms of Use. To order a print version, visit SPIE.

SPIE.

Field Guide to

Linear Systems in Optics

J. Scott Tyo
Andrey S. Alenin

SPIE Field Guides
Volume FG35

John E. Greivenkamp, Series Editor

SPIE PRESS
Bellingham, Washington USA

SPIE.

Introduction to the Series

Welcome to the *SPIE Field Guides* — a series of publications written directly for the practicing engineer or scientist. Many textbooks and professional reference books cover optical principles and techniques in depth. The aim of the *SPIE Field Guides* is to distill this information, providing readers with a handy desk or briefcase reference that provides basic, essential information about optical principles, techniques, or phenomena, including definitions and descriptions, key equations, illustrations, application examples, design considerations, and additional resources. A significant effort will be made to provide a consistent notation and style between volumes in the series.

Each *SPIE Field Guide* addresses a major field of optical science and technology. The concept of these *Field Guides* is a format-intensive presentation based on figures and equations supplemented by concise explanations. In most cases, this modular approach places a single topic on a page, and provides full coverage of that topic on that page. Highlights, insights, and rules of thumb are displayed in sidebars to the main text. The appendices at the end of each *Field Guide* provide additional information such as related material outside the main scope of the volume, key mathematical relationships, and alternative methods. While complete in their coverage, the concise presentation may not be appropriate for those new to the field.

The *SPIE Field Guides* are intended to be living documents. The modular page-based presentation format allows them to be easily updated and expanded. We are interested in your suggestions for new *Field Guide* topics as well as what material should be added to an individual volume to make these *Field Guides* more useful to you. Please contact us at **fieldguides@SPIE.org**.

John E. Greivenkamp, *Series Editor*
Optical Sciences Center
The University of Arizona

The Field Guide Series

Keep information at your fingertips with the SPIE Field Guides:

Field Guide to Linear Systems in Optics

The College of Optical Sciences (OSC) at the University of Arizona has long offered a course called "OPTI512R: Fourier Transforms, Linear Systems, and Optics" in its graduate program. The course was initiated and designed by Prof. Jack Gaskill, and was taught largely out of a textbook by the same name that was published in 1978. When Prof. Tyo joined OSC in 2006, he was asked to take over the course, as Prof. Gaskill had retired some years earlier.

Tyo came to the class with an electrical engineer's classic understanding of linear systems in time and frequency. Tyo quickly came to realize that, at that time, Prof. Gaskill's textbook was the only one written from the perspective of an optical engineer who needs to take 2D spatial Fourier transforms instead of 1D temporal ones. This difference gives rise to several subtle but important stylistic requirements that Prof. Gaskill captured well in his text. As with most instructors, Tyo began to add his own take on the material over the years.

Andrey Alenin joined his group in 2010, and he showed a strong interest in both the pedagogy and the presentation of the course material; the two authors have since worked together to refine the presentation. As of the writing of this *Field Guide*, Prof. Gaskill's text is still the primary reference in the class. However, when John Greivenkamp discussed with us the possibility of writing a *Field Guide* on this topic, he gave the authors the opportunity to go through the notes and reorganize them into a sequence more suited for this handbook format.

The process is, of course, circular. During the current semester of teaching OPTI512R, while completing this *Field Guide*, the authors have realized that the entire structure of the course will need to be revised going forward. The efforts undertaken to write this book have provided a new perspective on the classic course content.

Field Guide to Linear Systems in Optics

J. Scott Tyo
College of Optical Sciences
The University of Arizona

Andrey S. Alenin
College of Optical Sciences
The University of Arizona

Table of Contents

Table of Contents

Table of Contents

Glossary

Functions in this *Field Guide* are functions of spatial variables x and y unless noted otherwise. Lowercase letters are used to denote functions of the spatial variables $(f(x), g(x))$, whereas capital letters represent their Fourier transforms $(F(\xi), G(\xi))$.

Sequences of discrete samples of a function are denoted with the subscript k (f_k, g_k), and samples of the corresponding DFTs are denoted with subscript n (F_n, G_n).

Primed variables $(x', y', \xi', \eta', \text{etc.})$ denote variables of integration.

BPF	Bandpass filter
CTF	Coherent transfer function
D	Pupil diameter
d_o, d_i	Object and image distances
$\mathcal{D}\{f_k\}$	Discrete Fourier transform of sequence f_k
$\mathbf{E}$	Complex vector electric field
$E[f(x)]$	Expected value of $f(x)$
f	Focal length
$f/\#$	F-number
$f/\#_w$	Working F-number
$f_s(x)$	Ideally sampled function $f(x)$
$\mathcal{F}\{f(x)\}$	Fourier transform of $f(x)$
$h(x)$	Impulse response
$H(\xi)$	Transfer function
$\mathcal{H}(\xi)$	Optical transfer function
HPF	High-pass filter
$J_0(x)$	Zeroth-order Bessel function of first kind
$\mathbf{k}$	Wave vector
L	Spatial extent of a function
$\mathcal{L}$	Linear shift invariant operator
$\mathcal{L}\{f(t)\}$	Laplace transform of $f(t)$
LPF	Low-pass filter
$m_n(f(x))$	n^{th} moment of $f(x)$
MTF	Modulation transfer function
OTF	Optical transfer function
PSD	Power spectra density
PSF	Point spread function

Glossary

r	Polar coordinate radius
$\mathbf{r}$	2D vector $x\hat{\mathbf{x}} + y\hat{\mathbf{y}}$
$\mathscr{S}$	Mathematical operator
SNR	Signal to noise ratio
T	Temporal period
$t(x, y)$	Transmission function
u	Complex scalar optical field amplitude
W	Spatial frequency bandwidth
$W(x, y)$	Wavefront aberration function
X	Spatial period
x, y	Cartesian coordinates
x_s	Spatial sampling interval
$\mathcal{Z}\{f_k\}$	Z-transform of sequence f_k
$\delta(x - x_0)$	Impulse function at $x = x_0$
Δx	Sampling resolution in the space domain
$\Delta \xi$	Sampling resolution in the frequency domain
$\gamma_{fg}(x)$	Correlation between functions $f(x)$ and $g(x)$
$\gamma_x, \gamma_y, \gamma_z$	Direction cosines
η	Spatial frequency in y
λ	Wavelength
ν	Temporal frequency
θ	Polar coordinate angle
ρ	Radial distance in frequency plane
$\boldsymbol{\rho}$	$\xi\hat{\boldsymbol{\xi}} + \eta\hat{\boldsymbol{\eta}}$
υ	Normalized frequency ξ/ξ_s
ξ	Spatial frequency in x

Complex Numbers and Complex Plane

Complex analysis is an extension of mathematics on the real number line that is valuable in many areas. To introduce complex numbers, first define the **imaginary number**

$$i = \sqrt{-1}$$

A **complex number** u is the superposition of a completely **real number** and a completely **imaginary number**:

$$u = \Re[u] + i\Im[u] = x + iy$$

When the complex number is written in terms of the real and imaginary parts, it is called **Cartesian** notation.

Because the real and imaginary parts of a complex number cannot cancel each other, they can be treated as orthogonal and be used as basis vectors in a 2D plane called the **complex plane**. After a complex number is plotted in this 2D plane, it is often useful to define a **polar** coordinate system:
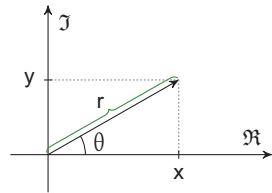
$$x = r\cos\theta \qquad\qquad y = r\sin\theta$$

$$\theta = \arctan(y/x) \qquad\qquad r = \sqrt{x^2 + y^2}$$

Euler's identity follows from the Taylor series expansion of the exponential using the arithmetic rules on the following page. The identity allows us to relate polar and Cartesian notation in terms of a **complex exponential**:

$$u = re^{i\theta} = r\cos\theta + ir\sin\theta = x + iy$$

The complex exponential can be used in all computations, and it obeys all of the rules of exponentials. This *Field Guide* relies heavily on this representation.

Complex Arithmetic

When performing **complex arithmetic**, it is necessary to determine whether to operate in Cartesian or polar coordinates. One system is generally easier to work in than the other for any particular computation.
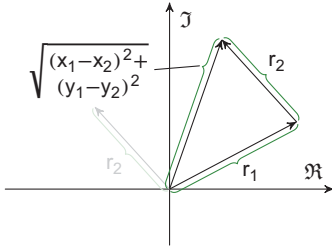
Addition and Subtraction

The orthogonal nature of the real and imaginary parts of the complex number point to a graphical method of performing addition. It is necessary to convert to Cartesian coordinates to perform these operations as

$$u_1 + u_2 = (x_1 + x_2) + i(y_1 + y_2)$$
$$u_1 - u_2 = (x_1 - x_2) + i(y_1 - y_2)$$

Multiplication and Division

Multiplication and division are usually easier to compute when working in **polar coordinates**, but there are rules for working in **Cartesian coordinates**, as well:

$$u_1 u_2 = (x_1 x_2 - y_1 y_2) + i(x_1 y_2 + x_2 y_1)$$
$$= r_1 r_2 e^{i(\theta_1 + \theta_2)}$$

$$\frac{u_1}{u_2} = \frac{(x_1 x_2 + y_1 y_2) - i(x_1 y_2 - x_2 y_1)}{x_2{}^2 + y_2{}^2} = \frac{r_1}{r_2} e^{i(\theta_1 - \theta_2)}$$

Integer Powers of Complex Numbers
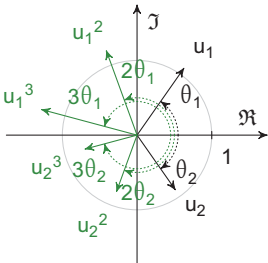
In Cartesian notation, binomial expansion rules are used for powers. In polar notation, rules for raising exponents to a power are employed:

$$u^n = (x + iy)^n = r^n e^{in\theta}$$

Specialized Complex Operations

The **complex conjugate** of u is denoted u^*. The conjugate is defined as

$$u = re^{i\theta} \qquad\qquad u^* = re^{-i\theta}$$
$$u = x + iy \qquad\qquad u^* = x - iy$$

The complex conjugate satisfies

$$|u| = |u^*| \qquad\qquad \angle u = -\angle u^*$$

It has special use when added to, subtracted from, or multiplied by the original complex number:

$$u + u^* = 2x \qquad u - u^* = i2y \qquad uu^* = r^2$$

The **magnitude** of a complex number is the length r of the corresponding vector in the complex plane. It is denoted by the absolute value symbol:

$$|u| = \sqrt{uu^*} = r = \sqrt{x^2 + y^2}$$

The **argument** of a complex number is the angle made with the real axis in the complex plane. Because the functions sine and cosine are **periodic**, the phase angle is only determined to within 2π radians:

$$\angle u = \angle re^{i(\theta + 2n\pi)} = \theta + 2n\pi$$

where n is an arbitrary integer.

The fact that the argument of the complex number is not unique leads to multiple values for the root. Consider

$$\sqrt[k]{u} = u^{1/k}$$
$$= \sqrt[k]{r}e^{i(\theta/k + (2\pi n/k))}$$

This results in k unique roots of u with $n = 0, \dots, k - 1$. It is important to select the proper root when multiple solutions exist, as dictated by the physics of a particular problem.
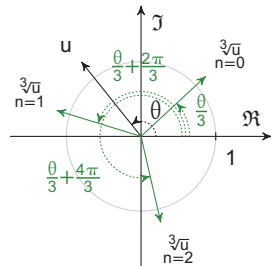
Complex Sinusoidal Functions and Phasors

One of the most important uses of complex numbers in this *Field Guide* is the use of complex exponentials to represent **time harmonic functions**. These functions vary periodically in time with period $T = 1/\nu$, where ν is the temporal frequency of oscillation in [Hz].

Consider a function that varies harmonically in both time and space as

$$f(x;t) = F(x)\cos\left[2\pi\nu t - \phi(x)\right]$$

where the phase constant $\phi(x)$ is expressed in **radians**, and 2π radians is equivalent to 360°. It is common in optics to express this function as

$$f(x;t) = \mathrm{Re}\left[F(x)e^{i2\pi\nu t - \phi(x)}\right] = \mathrm{Re}\left[\tilde{f}*(x)e^{i2\pi\nu t}\right]$$

The **complex phasor** is defined as

$$\tilde{f}(x) = F(x)e^{i\phi(x)}$$

Phasors bring a significant computational advantage when working with trigonometric functions such as sine and cosine. When adding two sinusoids that have the same frequency ν but different phases,

$$u_1(t) + u_2(t) = A_1\cos(2\pi\nu t - \phi_1) + A_2\cos(2\pi\nu t - \phi_2)$$

The variables u_1 and u_2 can be represented by the phasors

$$\tilde{u}_1 = A_1 e^{i\phi_1} \qquad\qquad \tilde{u}_2 = A_2 e^{i\phi_2}$$

It is not necessary to resort to sum- and difference-angle trigonometric identities. By using phasors and capitalizing on the linearity of complex addition, it is possible to write

$$u_1(t) + u_2(t) = \mathrm{Re}\left[(\tilde{u}_1 + \tilde{u}_2)\,e^{i2\pi\nu t}\right]$$

The phasors can be manipulated directly before considering the harmonic nature of the functions.

Idealized Models and the Unit Step Function

In the study of optics, it is often useful to model a physical process with an idealized mathematical function. A good example of this is an opaque edge that blocks all light for $x < 0$ and passes all light for $x > 0$. In reality, no true physical edge could satisfy this ideal, as there would be irregularities in the edge and some sort of a transition region between light and shadow. Those detailed features may be important for certain optical systems, such as near-field optical imaging. However, for many optical processes, the difference between the true transmission and the idealized model is insignificant. In such cases the ideal edge is modeled as the **unit step function**:

$\text{step}(x) = \begin{cases} 1 & x > 0 \\ \frac{1}{2} & x = 0 \\ 0 & x < 0 \end{cases}$	
$\text{ramp}(x) = \begin{cases} x & x \geqslant 0 \\ 0 & x < 0 \end{cases}$	
$\text{sgn}(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}$	
$\text{rect}(x) = \begin{cases} 1 & |x| < \frac{1}{2} \\ \frac{1}{2} & |x| = \frac{1}{2} \\ 0 & |x| > \frac{1}{2} \end{cases}$	

The latter three can be defined in terms of the step function:
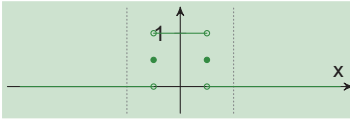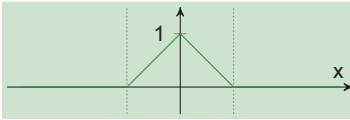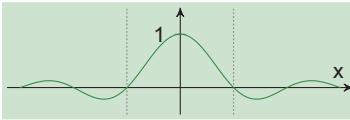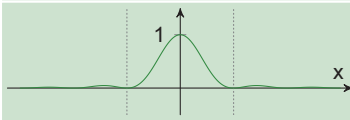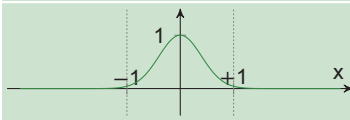
$$\text{ramp}(x) = \int_{-\infty}^{x} \text{step}(x')\, dx'$$

$$\text{sgn}(x) = \text{step}(x) - \text{step}(-x)$$

$$\text{rect}(x) = \text{step}(x + \tfrac{1}{2}) - \text{step}(x - \tfrac{1}{2})$$

In general, the step, ramp, and sgn functions are used to modify other functions, such as sine and cosine. They can represent edges, apertures, and other physical processes that are important in optics, imaging, and diffraction.

Pulse-Like Functions

Pulse-like functions are localized in one or more independent dimensions. There are many processes in optics that are well modeled by these functions, including a burst of optical energy from a pulsed laser (temporal localization), the aperture of a lens (spatial localization), or a band-pass spectral filter (wavenumber localization). Furthermore, many periodic structures like diffraction gratings or lens arrays can be built up using a pulse-like function as the unit cell.

This *Field Guide* relies heavily on five pulse-like functions:

$\text{rect}(x) = \begin{cases} 1 & \lvert x \rvert < \frac{1}{2} \\ \frac{1}{2} & \lvert x \rvert = \frac{1}{2} \\ 0 & \lvert x \rvert > \frac{1}{2} \end{cases}$	
$\text{tri}(x) = \begin{cases} 1 - \lvert x \rvert & \lvert x \rvert < 1 \\ 0 & \lvert x \rvert \geqslant 1 \end{cases}$	
$\text{sinc}(x) = \dfrac{\sin(\pi x)}{\pi x}$	
$\text{sinc}^2(x) = \dfrac{\sin^2(\pi x)}{(\pi x)^2}$	
$\text{Gaus}(x) = e^{-\pi x^2}$	

The functions defined here are used throughout the literature with slightly different notation conventions. This *Field Guide* uses the selected definitions so that all functions are symmetric about $x = 0$ and integrate to unity:

$$\int_{-\infty}^{\infty} p(x)\, dx = 1$$

where $p(x)$ is any of the pulse functions defined above.

Impulse Function

The pulse-like functions introduced previously need to be scaled and shifted in order to represent a broad range of physical processes. This scaling/shifting is generally denoted as

$$p(x) \rightarrow \frac{1}{|b|} p\left(\frac{x - x_0}{b}\right)$$

The effect of a positive shift x_0 shifts the center of the function to the right, and the effect of a positive scaling factor $b > 1$ widens the pulse. If the original pulse-like function is normalized as $\int_{-\infty}^{\infty} p(x) = 1$, then the scaled function also integrates to unity as
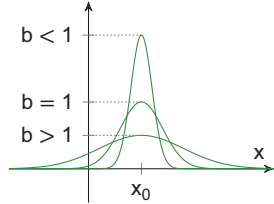
$$\frac{1}{|b|} \int_{-\infty}^{\infty} p\left(\frac{x - x_0}{b}\right) dx = 1$$

Examining the above scaling and shifting properties, it is possible to look at the limit as $b \rightarrow 0$. The **impulse function** is defined as

$$\delta(x - x_0) = \lim_{b \to 0} \frac{1}{|b|} p\left(\frac{x - x_0}{b}\right)$$

As the magnitude of b goes to zero, the pulse becomes narrower with an increasing peak value, while still integrating to unity. In the limit, the impulse function can be thought of as having zero-width and infinite height, but remaining integrable. It is conventional to denote the impulse function $a\delta(x - x_0)$ graphically as an arrow located at $x = x_0$, with height proportional to a.
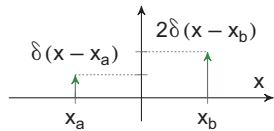
When the variables are discrete, as in sampled systems, the **Kronecker** δ-function is used:

$$\delta_{n,m} = \begin{cases} 1 & n = m \\ 0 & n \neq m \end{cases}$$

Impulse Function Properties

The **impulse function** is especially important when it appears in the integrand, either alone or multiplied by an arbitrary function $f(x)$. Two important properties:

Unit Area	$\int_a^b \delta(x - x_0)\, dx = \begin{cases} 1 & a < x_0 < b \\ 0 & \text{else} \end{cases}$
Sifting	$\int_{-\infty}^{\infty} f(x)\,\delta(x - x_0)\, dx = f(x_0)$

There are several other properties that derive either from simple changes of variables or from the definition of the impulse function on the previous page. For all of these properties, the equals sign should really be interpreted as "is equivalent to" because these properties are really applicable in the sense of using $\delta(x - x_0)$ as an operator.

Scaling	$\delta\left(\frac{x - x_0}{b}\right) = |b|\delta(x - x_0)$
	$\delta(ax - x_0) = \frac{1}{|a|}\delta\left(x - \frac{x_0}{a}\right)$
Even Symmetry	$\delta(-x + x_0) = \delta(x - x_0)$

The properties of the impulse function can be used to simplify expressions that include a product of the impulse function with another function:

Sifting Simplification	$f(x)\delta(x - x_0) = f(x_0)\delta(x - x_0)$	
	$\delta(x)\delta(x - x_0) = 0$	$x_0 \neq 0$
	$\delta(x - x_0)\delta(x - x_0)$	undefined

The impulse function can also be defined in terms of the integral of a complex exponential. This property is a consequence of the **Fourier transform**, which will be treated later in this *Field Guide*:

$$\delta(x - x_0) = \int_{-\infty}^{\infty} e^{i 2\pi \xi (x - x_0)}\, d\xi$$

Integrals and Derivatives of the Delta Function

The function $\delta(x - x_0)$ can be used to define several other special functions.

Unit Step Function
Consider the partial integral of the δ-function:

$$\int_{-\infty}^{x} \delta(x' - x_0)\, dx' = \left\{ \begin{array}{cl} 1 & x > x_0 \\ \frac{1}{2} & x = x_0 \\ 0 & x < x_0 \end{array} \right. = \text{step}(x - x_0)$$

In order to get the value of $\frac{1}{2}$ at $x = x_0$, resort to the definition of the impulse function and evaluate the integral before taking the limit. This result also implies that

$$\delta(x - x_0) = \frac{d}{dx}\text{step}(x - x_0)$$

The Doublet Function
The derivative of the δ-function is also important and denoted as $\delta'(x - x_0)$. Higher-order derivatives can be denoted either with multiple primes or a superscripted number in parentheses.
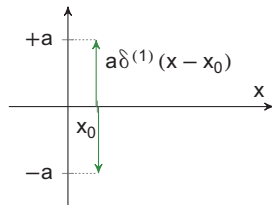
For example,

$$\delta'''(x - x_0) = \delta^{(3)}(x - x_0)$$

denotes the third derivative of the δ-function. It is conventional to depict the doublet function graphically as two impulses—one positive and one negative—at approximately the same location. By using integration by parts, it is straightforward to show that

$$\int_{-\infty}^{\infty} f(x')\,\delta^{(n)}(x' - x_0)\, dx' = (-1)^n \left. \frac{d^n f(x)}{dx^n}\right|_{x=x_0}$$

Whereas the δ-function sifts out a value at a particular location, the doublet function and its higher-order versions sift out the derivative of the function on which they operate.

Comb Function

In many applications, an array of delta functions that are evenly spaced is useful. Such a function is often referred to as the **comb function**:

$$\text{comb}(x) = \sum_{n=-\infty}^{\infty} \delta(x-n).$$
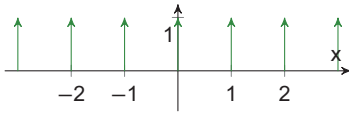
If an application calls for something other than integer spacing, the scaling properties of the δ-function can be used to write

$$\text{comb}\left(\frac{x}{b}\right) = |b| \sum_{n=-\infty}^{\infty} \delta(x-nb)$$

These δ-functions are separated by b, and each one has an integrated area of $|b|$.

The following expression follows directly from the properties of the δ-function. It is important in developing the theory of ideal sampling:

$$f(x)\left[\frac{1}{|b|}\text{comb}\left(\frac{x}{b}\right)\right] = \sum_{n=-\infty}^{\infty} f(nb)\,\delta(x-nb)$$

The following integral is defined as a **convolution integral** between an arbitrary function $p(x)$ and the comb function. The result is a periodic function with period X:

$$f_p(x) = \int_{-\infty}^{\infty} p(x')\frac{1}{|b|}\text{comb}\left(\frac{x-x'}{b}\right)dx' = \sum_{n=-\infty}^{\infty} p(x-nb)$$

The following property is a consequence of the Fourier series that will be considered later in this *Field Guide*:

$$\text{comb}(x) = \sum_{n=-\infty}^{\infty} e^{i2\pi nx}$$

Orthonormal Basis Functions

Basis functions are used to represent more complicated functions of interest. A set of basis functions $\{\psi(x)_n\}_{n=1}^N$ usually has some canonical properties that are important for describing a certain process. The set can be finite, countably infinite ($N = \infty$), or the parameter n can even be continuous. The following table shows the two kinds of **superposition of basis functions**; the bottom row represents the **orthonormality condition**.

Discrete Sum	Continuous Integral
$f(x) = \sum_{n=1}^{N} c_n \psi_n(x)$	$f(x) = \int_{-\infty}^{\infty} c(\nu) \psi_\nu(x)\, d\nu$
$\langle \psi_n(x), \psi_m(x) \rangle = \delta_{n,m}$	$\langle \psi_\nu(x), \psi_\mu(x) \rangle = \delta(\nu - \mu)$

The notation $\langle \psi_n(x), \psi_m(x) \rangle$ represents an **inner product**, which is a generalization of the **dot product** to complex and continuous functions.

Discrete Inner Product	Continuous Inner Product
$\sum_{n=a}^{b} f_n^* g_n$	$\int_a^b f^*(x) g(x)\, dx$

The values of a and b define the valid range of the functions, which can be infinite when appropriate. The weights c_n and $c(\nu)$ are computed using the orthonormality condition:

$$c_n = \langle \psi_n(x), f(x) \rangle$$

The following table lists three important basis sets.

Continuous set of δ-functions shifted by x_0	
$\psi_{x_0}(x) = \delta(x - x_0)$	$-\infty < x_0 < \infty$
Continuous set of complex sinusoids of frequency ξ	
$\psi_\xi(x) = e^{i2\pi\xi x}$	$-\infty < \xi < \infty$
Discrete set of harmonics n of fundamental frequency ξ_0	
$\psi_n(x) = e^{i2\pi n \xi_0 x}$	$n \in (-\infty, \ldots, 0, \ldots, \infty)$

Harmonic Analysis and Fourier Series

Periodic functions are important in a number of optical, diffraction, and imaging problems. An X-periodic function of the spatial variable x has **period** X and **fundamental frequency** ξ_0, where

$$\xi_0 = \frac{1}{X}$$

A periodic function $f_X(x)$ obeys

$$f_X(x + X) = f_X(x) \quad \forall x$$

> The **harmonics** of a periodic functions are the integer multiples of the fundamental frequency. The convention is that the *first* harmonic is the fundamental frequency, the *second* harmonic is $2\xi_0$, and so on. Note that while the higher harmonics have periods given by X/n, each of the harmonics is also X-periodic.

If several X-periodic functions are added together, then the resulting sum is also X-periodic. The summation known as the **Fourier series**,

$$f_X(x) = \sum_{n=-\infty}^{\infty} c_n e^{i2\pi n \xi_0 x}$$

is X-periodic. The harmonic exponentials act as the **orthonormal basis** $\psi_n(x) = e^{i2\pi n \xi_0 x}$ because

$$\left\langle e^{i2\pi n \xi_0 x}, e^{i2\pi m \xi_0 x} \right\rangle = \frac{1}{X} \int_{x'=x_0}^{x_0+X} e^{i2\pi(m-n)\xi_0 x'} dx' = \delta_{m,n}$$

The coefficients of the Fourier series can then be computed using the orthogonality relationship

$$c_n = \left\langle \psi_n(x), f_X(x) \right\rangle$$
$$= \frac{1}{X} \int_{x'=x_0}^{x_0+X} f_X(x) e^{-i2\pi n \xi_0 x'} dx'$$

The choice of the starting point x_0 for the integral is arbitrary, and the integral covers exactly one period X of the original function $f_X(x)$.

Square Wave and Truncated Fourier Series

An important function is the **square wave**, which is a good model for diffraction gratings, crystal structures, and other physical processes important in optics:

$$f_{sw}(x) = \sum_{n=-\infty}^{\infty} \text{rect}\left(\frac{x - 2nX}{2aX}\right)$$
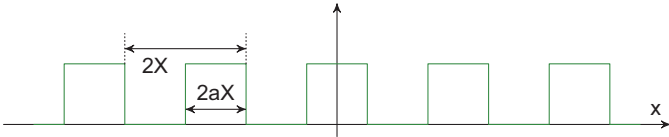
The Fourier series coefficients are

$$c_n = \frac{1}{2X} \int_{-aX}^{aX} e^{i\pi \frac{nx'}{X}} dx' = a\text{sinc}(na)$$

The Fourier series for the square wave with $a = \frac{1}{2}$ contains only **odd harmonics**, and the zero-frequency term $c_0 = a$, corresponding to the average value of the square wave.

The definition of the Fourier series includes all harmonics from $-\infty$ to ∞. In practical systems, some finite number N of positive and negative harmonics will result in a **truncated Fourier series**. Such a truncation in the Fourier series leads to the **Gibbs phenomenon**, which manifests itself as ringing in the partially reconstructed signal.
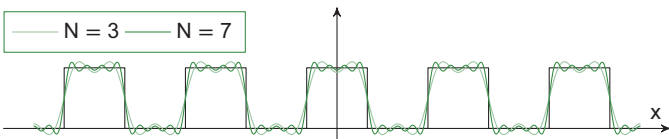
This plot shows the effect of truncating the Fourier series for the square wave with $a = \frac{1}{2}$ at $N = 3$ and $N = 7$. The overshoot and undershoot apparent in the plot is always present for any finite N.

Fourier Transform

As the period X of a periodic signal increases, the corresponding fundamental frequency $\xi_0 = 1/X$ decreases. The spacing between harmonic frequencies also decreases. In the limit $X \to \infty$, the signal becomes **aperiodic**, and the discrete set of harmonic frequencies becomes a continuum of frequencies $\xi \in (-\infty, \infty)$.

The **Fourier transform** of a function $f(x)$ is defined as

$$\mathcal{F}\{f(x)\} = F(\xi) = \int_{-\infty}^{\infty} f(x)e^{-i2\pi\xi x}dx$$

The inverse Fourier transform of $F(\xi)$ is defined as

$$\mathcal{F}^{-1}\{F(\xi)\} = f(x) = \int_{-\infty}^{\infty} F(\xi)e^{+i2\pi\xi x}d\xi$$

Another way of thinking of the Fourier transform is that the function $f(x)$ is a weighted superposition of complex exponentials $e^{i2\pi\xi x}$, where the weight $F(\xi)$ is determined by taking the inner product $\left\langle e^{i2\pi\xi x}, f(x) \right\rangle$.

The Fourier transform can be abbreviated using one of several notations:

$$\begin{aligned} f(x) &\xrightarrow{\mathcal{F}} F(\xi) \\ f(x) &\leftrightarrow F(\xi) \\ \mathcal{F}\{f(x)\} &= F(\xi) \\ \mathcal{F}^{-1}\{F(\xi)\} &= f(x) \end{aligned}$$

The table shows important Fourier transform pairs on which this *Field Guide* relies heavily. Many more pairs are included in the transform tables at the end of the book.

$f(x)$	$F(\xi)$
$\delta(x - x_0)$	$e^{-i2\pi x_0 \xi}$
$e^{i2\pi\xi_0 x}$	$\delta(\xi - \xi_0)$
$\text{rect}\left(\frac{x}{b}\right)$	$\lvert b \rvert \text{sinc}(b\xi)$
$\text{tri}\left(\frac{x}{b}\right)$	$\lvert b \rvert \text{sinc}^2(b\xi)$
$e^{-x}\text{step}(x)$	$\frac{1}{1+i2\pi}$
$\lvert x \rvert^{-1/2}$	$\lvert \xi \rvert^{-1/2}$
$e^{-\pi(x/b)^2}$	$\lvert b \rvert e^{-\pi(b\xi)^2}$
$e^{\pm i\pi(x^2 - 1/8)}$	$e^{\mp i\pi(\xi^2 - 1/8)}$
$\cos(2\pi\xi_0 x)$	$\frac{1}{2}\left[\delta(\xi - \xi_0) + \delta(\xi + \xi_0)\right]$
$\sin(2\pi\xi_0 x)$	$\frac{1}{2i}\left[\delta(\xi - \xi_0) - \delta(\xi + \xi_0)\right]$

Fourier Transform Properties
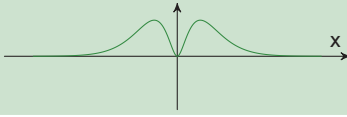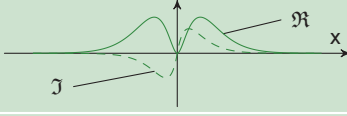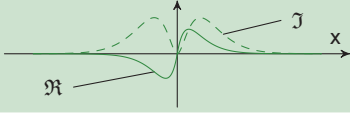
Fourier transform tables are collections of known transforms. (One is included at the end of this *Field Guide*.) The following properties of the Fourier transform, in conjunction with a Fourier transform table, allow for a greater range of pairs to be determined. For all of the cases that follow, assume that the transform pairs $f(x) \leftrightarrow F(\xi)$ and $g(x) \leftrightarrow G(\xi)$ are known.

	$f(x)$	$F(\xi)$
Coordinate Reversal	$f(\pm x)$	$F(\pm \xi)$
Complex Conjugates	$f^*(\pm x)$	$F^*(\mp \xi)$
Repeated Transform	$F(\pm x)$	$f(\mp \xi)$
Shifting	$f(x \pm x_0)$	$e^{\pm i 2\pi x_0 \xi} F(\xi)$
Modulation	$e^{\pm i 2\pi \xi_0 x} f(x)$	$F(\xi \mp \xi_0)$
Scaling	$f\left(\frac{x}{b}\right)$	$|b| F(b\xi)$
Derivative	$f^{(k)}(x)$	$(i 2\pi \xi)^k F(\xi)$
	$(-i 2\pi x)^k f(x)$	$F^{(k)}(\xi)$
Integral	$\int_{-\infty}^{x} f(x') dx'$	$\frac{F(\xi)}{i 2\pi \xi} + \delta(\xi) F(0)$
Linearity	$a f(x) + b g(x)$	$a F(\xi) + b G(\xi)$

The **Heisenberg Uncertainty Principle** is an important consequence of the scaling property of Fourier transforms. If the arguments of two distributions $f(x_1)$ and $F(\xi_1)$ correspond to Fourier conjugate variables—such as position and momentum in quantum mechanics—then it is not possible to observe the value of both arguments simultaneously with infinite precision. As the scaling factor $b \rightarrow 0$, and as $f(x/b)$ gets narrower corresponding to greater precision in the predictability of one quantity, then the distribution $F(b\xi)$ gets wider at the same rate corresponding to less precision in the predictability of the dual variable. This is true for all conjugate variables, regardless of their connection to a physical process.

Symmetry of Functions and Fourier Transforms

Functional symmetry can be very important when dealing with Fourier transforms. The following symmetries are commonly noted:

Even $f_e(x) = f_e(-x)$	
Odd $f_o(x) = -f_o(-x)$	
Hermitian $f_H(x) = f_H{}^*(-x)$	
Anti-Hermitian $f_A(x) = -f_A{}^*(-x)$	

An arbitrary function with no symmetry can be written as a superposition of two symmetric functions:

$$f(x) = f_e(x) + f_o(x)$$
$$f(x) = f_H(x) + f_A(x)$$

The following relationships can be demonstrated from these symmetries and the Fourier transform properties:

$f(x)$	$F(\xi)$
Real	Hermitian
Imaginary	Anti-Hermitian
Hermitian	Real
Anti-Hermitian	Imaginary
Real and Even	Real and Even
Real and Odd	Imaginary and Odd

Parseval's Theorem and Moment Theorem

Parseval's theorem is a restatement of conservation of energy. Mathematically, it is written as

$$\int_{-\infty}^{\infty} |f(x)|^2 \, dx = \int_{-\infty}^{\infty} |F(\xi)|^2 \, d\xi$$

It is also known as the **Rayleigh energy theorem**. Its most general case is referred to as **Plancherel's theorem**.

Parseval's theorem states that the total energy is the same no matter whether it is computed position by position or frequency by frequency, and it is straightforward to prove from the properties of the Fourier transform.

> Many of the physical quantities that are used with the Fourier transform have units that are equivalent to the square root of power or energy. For example, in optics the quantity of interest is often the electric field E, which has units of volts per meter. The power density $|E|^2/\eta$ has units of volts-squared per ohm per meter-squared, equivalent to watts per square meter. The integral of the square of a quantity over infinite limits is therefore often equivalent to the total energy in the field quantity.

In probability and statistics, the Fourier transform of a probability distribution function (PDF) $f(x)$ is known as a characteristic function $F(\xi)$. The statistical moments of a PDF are computed as

$$m_n[f(x)] = \int_{-\infty}^{\infty} x^n f(x) \, dx$$

The moment theorem follows directly from the derivative property of the Fourier transform:

$$m_n[f(x)] = \frac{1}{(-i2\pi)^n} \left. \frac{d^n F(\xi)}{d\xi^n} \right|_{\xi=0}$$

It is straightforward to show the corollary:

$$\int_{-\infty}^{\infty} f(x) \, dx = F(0)$$

When $f(x)$ is a PDF, $F(0) = 1$.

Laplace Transform

This *Field Guide* uses the Fourier transform and other transformations based on it to describe signals and systems. In some disciplines of linear systems, the **Laplace transform** is preferred because of its improved mathematical stability and natural ability to describe damped resonances.

The Laplace transform generalizes to complex frequencies. Because the Laplace transform is typically used with causal signals that are functions of time, the **complex frequency** is defined as

$$s = \sigma + i2\pi\nu$$

The Laplace transform is

$$F(s) = \int_{-\infty}^{\infty} f(t)e^{-st}dt$$

A Laplace transform has a **region of convergence**, which is the portion of the s-plane for which the transform can be evaluated. Several important pairs are shown below.

$f(t)$	$F(s)$	R.O.C.
$\delta(t - t_0)$	$e^{-t_0 s}$	all s
$\text{step}(t)$	$\frac{1}{s}$	$\text{Re}[s] > 0$
$e^{-\alpha t}\cos(2\pi\nu_0 t)\text{step}(t)$	$\frac{s+\alpha}{(s+\alpha)^2+(2\pi\nu_0)^2}$	$\text{Re}[s] > -\alpha$

One of the most important aspects of the Laplace transform is the ability to use it to solve differential equations. The Laplace transform of a derivative is

$$\mathcal{L}\{f'(t)\} = sF(s) - f(0^-)$$

where $f(0^-)$ is the initial condition of the function. Higher-order derivatives produce

$$\mathcal{L}\{f^{(n)}(t)\} = s^n F(s) - s^{n-1}f(0^-) - \ldots - f^{(n-1)}(0^-)$$

This property of the Laplace transform allows differential equations to be converted to algebraic equations.

2D Functions

Functions of the two independent variables x and y are especially important in optics and imaging because optical systems generally map planes to planes. A **2D function** is generally denoted as $f(x, y)$.

Separable functions are a special class of 2D functions that can be written as a product of two 1D functions:

$$f(x, y) = f_1(x)f_2(y)$$

Separability will be leveraged in many aspects of linear systems later in this *Field Guide*. All of the special functions defined earlier can be used to create 2D special functions.

tri(x, y)   Gaus(x, 1.5y)
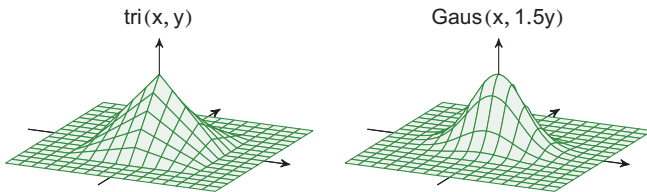
Functions can be separable in either Cartesian coordinates, as above, or in **cylindrical coordinates**, as

$$g\left(r, \theta\right) = g_1\left(r\right)g_2\left(\theta\right)$$

The class of **azimuthally symmetric** functions is often useful in optics, which can be written as

$$h\left(r, \theta\right) = h(r)$$

The Gaussian function has the special property of being separable in both Cartesian and cylindrical coordinates, as well as being azimuthally symmetric:

$$\text{Gaus}\left(x, y\right) = e^{-\pi x^2}e^{-\pi y^2} = e^{-\pi\left(x^2+y^2\right)} = e^{-\pi r^2} = \text{Gaus}\left(r\right)$$

Impulse Functions in Two Dimensions

The **impulse function's** characteristics change when going to two or more dimensions. In Cartesian coordinates, it is simplest to think of the impulse located at $\mathbf{r}_0 = (x_0, y_0)$ as a separable function:

$$\delta(\mathbf{r} - \mathbf{r}_0) = \delta(x - x_0)\delta(y - y_0)$$

The 2D impulse function integrates to unit area:

$$\iint_{-\infty}^{\infty} \delta(\mathbf{r} - \mathbf{r}_0)dA = \int_{-\infty}^{\infty} \delta(x - x_0)dx \int_{-\infty}^{\infty} \delta(y - y_0)dy = 1$$

Because the differential area element $dA = dxdy$ has units of m^2, each of the 1D impulse functions must carry units of inverse meters.

The two 1D impulse functions can also be plotted in two dimensions: $\delta(x - x_0)$ produces a 1D impulse for all locations along the line $x = x_0$, and $\delta(y - y_0)$ does the same for all locations

$\delta(x - x_0, y - y_0)$

along the perpendicular line $y = y_0$. Only at the point of intersection is the product of the two 1D impulses non-zero, creating the 2D impulse function.

$\delta(r - r_0, \theta - \theta_0)$
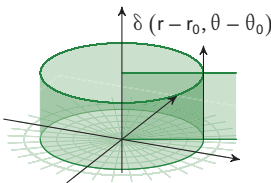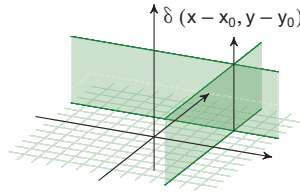
In cylindrical coordinates, the area element becomes $dA = rdrd\theta$ and the point $\mathbf{r}_0$ is (r_0, θ_0). The 1D impulse functions are

$$\delta(r - r_0) \qquad \frac{1}{r_0}\delta(\theta - \theta_0)$$

This *Field Guide* only considers 1D and 2D impulses. However, other areas of optics use higher-dimensional impulse functions for a variety of reasons. Examples include locations in a crystal lattice and point sources in 3D space or 4D space-time.

Fourier Transforms of 2D Functions

The 2D Fourier transform and its inverse are defined in much the same way as the 1D case:

$$\mathcal{F}\{f(x,y)\} = F(\xi, \eta) = \iint_{-\infty}^{\infty} f(x,y)e^{-i2\pi(\xi x + \eta y)}dxdy$$

$$\mathcal{F}^{-1}\{F(\xi, \eta)\} = f(x,y) = \iint_{-\infty}^{\infty} F(\xi, \eta)e^{+i2\pi(\xi x + \eta y)}d\xi d\eta$$

The spatial frequency variable ξ refers to cycles per unit length in the x-direction, and the variable η refers to cycles per unit length in the y-direction. These equations do not simplify in general. However, in the special case of separable functions, the 2D Fourier transform factors into a product of two 1D Fourier transforms:

$$\mathcal{F}\{f_1(x)f_2(y)\} = F_1(\xi)F_2(\eta)$$

All of the properties of 1D Fourier transforms have their analogues in 2D whether or not the functions are separable. Some of the important properties are listed here with the assumption that $f(x,y) \leftrightarrow F(\xi, \eta)$ is known.

	$f(x,y)$	$F(\xi, \eta)$
Shifting	$f(x \pm x_0, y \pm y_0)$	$e^{i2\pi(\pm x_0\xi \pm y_0\eta)}F(\xi, \eta)$
Modulation	$e^{i2\pi(\pm\xi_0 x \pm \eta_0 y)}f(x,y)$	$F(\xi \mp \xi_0, \eta \mp \eta_0)$
Scaling	$f(x/b, y/d)$	$\lvert bd \rvert F(b\xi, d\eta)$
	$f(ax, cy)$	$\frac{1}{\lvert ac \rvert}F(\xi/a, \eta/c)$
Derivative	$\frac{d^{(n+m)}f(x,y)}{(i2\pi)^{n+m}dx^n dy^m}$	$\xi^n \eta^m F(\xi, \eta)$
	$x^n y^m f(x,y)$	$\frac{d^{(n+m)}F(\xi, \eta)}{(-i2\pi)^{n+m}d\xi^n d\eta^m}$
Integral	$\int_{-\infty}^{x}\int_{-\infty}^{\infty}f(x',y')dx'dy'$	$\frac{F(0,0)\delta(\xi,\eta)}{2} + \frac{F(\xi,0)\delta(\eta)}{i2\pi\xi}$
	$\int_{-\infty}^{\infty}\int_{-\infty}^{y}f(x',y')dx'dy'$	$\frac{F(0,0)\delta(\xi,\eta)}{2} + \frac{F(0,\eta)\delta(\xi)}{i2\pi\eta}$
	$\int_{-\infty}^{\infty}f(x,y')dy'$	$F(\xi,0)\delta(\eta)$
	$\int_{-\infty}^{\infty}f(x',y)dx'$	$F(0,\eta)\delta(\xi)$

Hankel Transform

There is a special case of the 2D Fourier transform when the function is azimuthally symmetric, i.e., $f(x,y) = f(r)$. In this case, the variables of integration need to be converted from Cartesian to cylindrical variables:

$$\mathcal{F}\{f(r)\} = \int_{r=0}^{\infty} \int_{\theta=0}^{2\pi} f(r)e^{-i2\pi r(\xi \cos\theta + \eta \sin\theta)} r \, dr \, d\theta$$

The spatial frequency variables ξ and η also form a pair of orthogonal axes, and a cylindrical coordinate system can be defined as

$$\xi = \rho \cos\theta \qquad\qquad \eta = \rho \sin\theta$$
$$\rho = \sqrt{\xi^2 + \eta^2} \qquad\qquad \phi = \tan\frac{\eta}{\xi}$$

The variable ρ denotes cycles per unit length in the **radial** direction, and the variable ϕ represents cycles per 2π radians. After these variables are inserted into the 2D Fourier transform, the Fourier transform becomes

$$F(\rho) = \int_{r=0}^{\infty} f(r)J_0(2\pi\rho)r \, dr$$

This special case of the 2D Fourier transform is known as the **Hankel transform**. It is very important to remember that the Hankel transform is a special case of the 2D Fourier transform when the functions are azimuthally symmetric. It is not a different transformation with its own properties.

The function $J_n(x)$ is the **n^{th}-order Bessel function** of the first kind, and it can be loosely considered as serving the role of a sinusoidal function in the radial direction.

$$J_n(\rho) = \frac{1}{2\pi} \int_{\theta=0}^{2\pi} e^{-i(n\theta - \rho \sin\theta)} d\theta$$
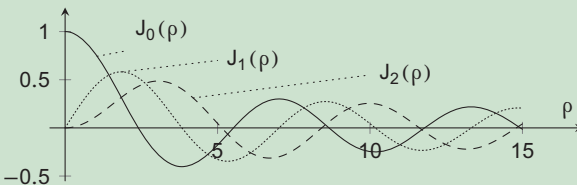
Hankel Transform Pairs and Properties

Two functions simplify analysis of azimuthally symmetric optical systems: $\mathrm{somb}(\rho)$ and $\gamma_{\mathrm{cyl}}\left(\rho/d, a\right)$. The "sombrero" function was coined by Gaskill and is also referred to in literature as $\mathrm{jinc}(\rho)$ and $\mathrm{besinc}(\rho)$. It plays the role of the sinc function in cylindrical coordinates:

$$\mathrm{somb}(\rho) = \frac{2J_1(\pi\rho)}{\pi\rho}$$

The second new function is $\gamma_{\mathrm{cyl}}\left(\rho/d, a\right)$. This function is the cross-correlation between $\mathrm{cyl}(\rho/d)$ and $\mathrm{cyl}(\rho/a)$. Several important pairs are shown in the following table.
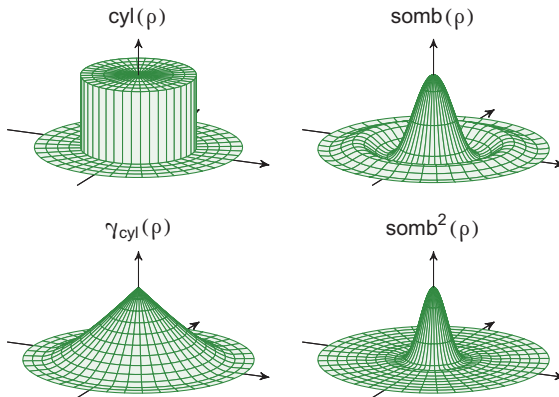
$f(r)$	$F(\rho)$
$\delta(r - r_0)$	$2\pi r_0 J_0(2\pi r_0 \rho)$
$\exp\left[\pm i \frac{r^2}{\lambda z}\right]$	$\pm i\lambda z e^{\mp i\lambda z \rho^2}$
$\mathrm{cyl}(r)$	$\frac{\pi}{4}\mathrm{somb}(\rho)$
$\gamma_{\mathrm{cyl}}\left(\frac{r}{d}, a\right)$	$\frac{\pi a^2 d^2}{4}\mathrm{somb}(a\rho)\mathrm{somb}(ad\rho)$
e^{-r}	$2\pi\left(4\pi\rho^2 + 1\right)^{2/3}$
$\cos(\pi r^2)$	$\sin(\pi\rho^2)$
r^{-1}	ρ^{-1}

Skew Functions

In many optical applications it is useful to define a 2D function in terms of a **skew transformation** of a Cartesian coordinate system. Consider the following mapping:

$$f(x,y) \rightarrow f(a_1x + b_1y + c_1, a_2x + b_2y + c_2)$$

As a relatively simple example, consider the function

$$\text{rect}(x, x-y) = \text{rect}(x)\text{rect}(x-y) = \begin{cases} 1, & \begin{matrix} |x| < \frac{1}{2} \\ |x-y| < \frac{1}{2} \end{matrix} \\ 0, & \begin{matrix} |x| > \frac{1}{2} \\ |x-y| > \frac{1}{2} \end{matrix} \end{cases}$$
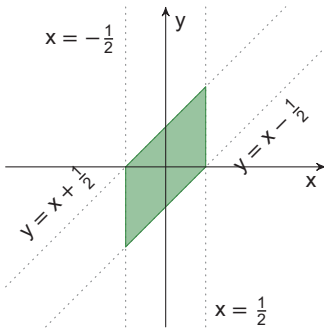
The function is equal to $\frac{1}{2}$ on the edges and $\frac{1}{4}$ at the corners. The principal axes of this function are the non-orthogonal lines $x = 0$ and $y = x$. It is much more convenient to represent the function as a separable function of these two mixed variables.

The Fourier transform of the skew function is related to the Fourier transform of the original function as follows:

$f(x,y)$	$f(a_1x + b_1y + c_1, a_2x + b_2y + c_2)$		
$F(\xi, \eta)$	$\dfrac{e^{-i2\pi(x_0\xi + y_0\eta)}}{|D|} F\left(\dfrac{b_2\xi - a_2\eta}{D}, \dfrac{a_1\eta - b_1\xi}{D} \right)$		
	$D = a_1b_2 - a_2b_1$	$x_0 = \dfrac{b_1c_2 - b_2c_1}{D}$	$y_0 = \dfrac{a_2c_1 - a_1c_2}{D}$

While this might seem specialized, the skew transformation is important in numerous applications in optics, including

- Non-rectangular sampling, such as in the Bayer filter
- Periodic media such as as natural crystals and artificial metamaterials
- Non-rectangular arrays that are common in recent generations of very large telescopes

Operators and LSI Systems

An **operator** is a mathematical description of how a system operates on or transforms an input (such as an optical field distribution in an aperture plane) to an output (such as the diffracted irradiance pattern). Mathematically, a general system operator is denoted as

$$\mathscr{S}\{f(x)\} = g(x)$$

> Both $f(x)$ and $g(x)$ are functions of the same independent variable, which differentiates operators from **transformations** such as the Fourier transform.

A subset of operators is of special note—those that are **linear**, **shift-invariant**, or both.

Given that $\mathscr{S}\{f_1(x)\} = g_1(x)$ and $\mathscr{S}\{f_2(x)\} = g_2(x)$, a linear operator obeys the **principle of superposition**:

$$\mathscr{S}\{a_1 f_1(x) + a_2 f_2(x)\} = a_1 g_1(x) + a_2 g_2(x)$$

Linearity allows a function to be written as a superposition of a known **basis set**, and an operator can act on the individual basis functions as

$$\mathscr{S}\left\{\sum_n a_n \psi_n(x)\right\} = \sum_n a_n \mathscr{S}\{\psi_n(x)\}$$

Given that $\mathscr{S}\{f(x)\} = g(x)$, a shift-invariant operator has the property

$$\mathscr{S}\{f(x - x_0)\} = g(x - x_0)$$

The tools of **convolution** and the **transfer function** introduced in the coming pages rely on these properties.

> An operator that is both linear and shift-invariant, also called LSI, is denoted in this *Field Guide* by the symbol $\mathscr{L}\{f(x)\}$. All LSI operators have the complex exponentials as **eigenfunctions**
>
> $$\mathscr{L}\left\{e^{i2\pi\xi x}\right\} = H(\xi)e^{i2\pi\xi x}$$

Convolution and Impulse Response

The **impulse response** of an LSI system is the output that is produced when the input to the system is an impulse:

$$\mathcal{L}\{\delta(x)\} = h(x)$$

When the system operates on a specific input $f(x)$,

$$\mathcal{L}\{f(x)\} = \mathcal{L}\left\{\int_{-\infty}^{\infty} f(x')\delta(x-x')dx'\right\} = \int_{-\infty}^{\infty} f(x')h(x-x')dx'$$

This mathematical operation is known as a **convolution integral**, and it is denoted as

$$f(x) * h(x) = \int_{-\infty}^{\infty} f(x')h(x-x')dx'$$

The significance of convolution is that any LSI system is completely described by its impulse response. Once $h(x)$ is known, the output for an arbitrary input can be computed.

The following tables present the properties of the convolution given $f(x) * h(x) = g(x)$.

Commutative	$f(x) * h(x) = h(x) * f(x)$
Associative	$(f(x) * h_1(x)) * h_2(x) = f(x) * (h_1(x) * h_2(x))$
Distributive	$f(x) * (h_1(x) + h_2(x)) = f(x) * h_1(x) + f(x) * h_2(x)$
Shift Invariance	$f(x - x_0) * h(x) = g(x - x_0)$
δ-function	$f(x) * \delta(x - x_0) = f(x - x_0)$
$\dfrac{d^n \delta(x)}{dx^n}$	$f(x) * \delta^{(n)}(x - x_0) = f^{(n)}(x - x_0)$

$f(x) * h(x)$	$g(x)$
$\text{step}(x) * \text{step}(x)$	$\text{ramp}(x)$
$\text{rect}(x) * \text{rect}(x)$	$\text{tri}(x)$
$\text{step}(x) * e^{-x}\text{step}(x)$	$(1 - e^{-x})\,\text{step}(x)$
$e^{-x}\text{step}(x) * e^{-x}\text{step}(x)$	$xe^{-x}\text{step}(x)$

Causality

Causality is a property that is most widely applicable to time-domain systems. A causal system has no output until the input is active. Mathematically, this places the following limitation on the impulse response:

$$h(t) = 0 \quad \forall t < 0$$

Causality is important for physically realizable transient systems. However, for functions of position, there is generally no practial importance to restricting the output for $x < 0$. For example, imaging systems with negative magnification produce outputs for $x < 0$, even if the input is zero for $x < 0$.

Kramers–Krönig Relationships

The requirement that $h(t) = 0$ for $t < 0$ for causal systems places restrictions on the relationship between the real and imaginary parts of the Fourier transform. An analytic function $h(t)$ is causal, and its Fourier transform is

$$\mathcal{F}\{f(t)\} = F(\nu) = F_r(\nu) + iF_i(\nu)$$

The real and imaginary parts of $F(\nu)$ satisfy

$$F_r(\nu) = \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty}\frac{F_i(\nu')}{\nu' - \nu}d\nu'$$

$$F_i(\nu) = -\frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty}\frac{F_r(\nu')}{\nu' - \nu}d\nu'$$

The symbol $\mathcal{P}$ denotes the **Cauchy principal value** of the integral, which dictates how to deform the contour of integration due to the pole at $\nu' = \nu$.

The Kramers–Krönig relationships place important restrictions on physical parameters of optical problems, such as the complex index of refraction of a medium. When dealing with an optical problem in the frequency domain, care must be taken to ensure that the Kramers–Krönig relationships are satisfied in order to obtain physically meaningful results.

Graphical Convolution

The following recipe allows the mathematical process of convolution to be visualized graphically. The convolution integral is

$$f(x) * h(x) = g(x) = \int_{-\infty}^{\infty} f(x')h(x - x')dx'$$

Consider the convolution between the two functions:

$f(x) = \cos(2\pi x)\,\text{step}(x)$	$h(x) = e^{-x}\text{step}(x)$
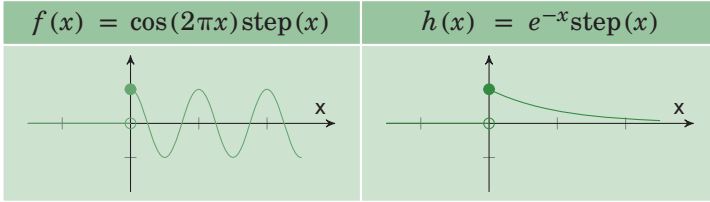
1. Choose a value of x and plot $f(x')$ and $h(x - x')$. Note that $h(x - x')$ is flipped with respect to the variable of integration x' and shifted by x.
2. Compute the multiplication $f(x')h(x - x')$.
3. Compute the integrated area of the product.
4. Choose a new value of x and repeat steps 2–4.

$x < 0$	$x > 0$
	f(x')h(x − x')
0	$\dfrac{\cos(2\pi x)+2\pi\,\sin(2\pi x)-e^{-x}}{1+4\pi^2}$
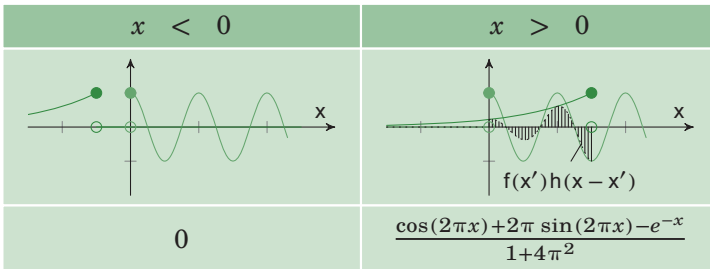
The two cases can be combined by introducing step(x)

$$g(x) = \frac{\cos(2\pi x)+2\pi\,\sin(2\pi x)-e^{-x}}{1+4\pi^2}\text{step}(x)$$
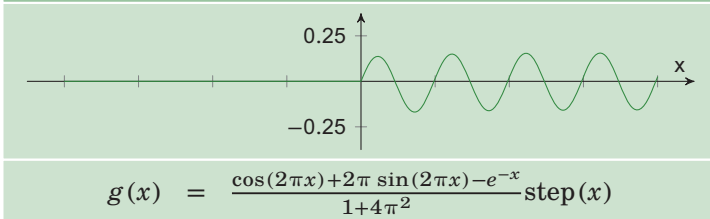
Convolution Theorem

The **convolution theorem** relates the Fourier transform of the output of an LSI system to the Fourier transforms of the input and the impulse response:

$$g(x) = f(x) * h(x) \qquad G(\xi) = F(\xi)H(\xi)$$

These two expressions suggest that a LSI system can be equivalently evaluated in either the space domain or the spatial frequency domain, depending on which is most convenient for any particular analysis.
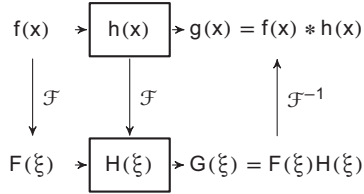
Amplitude-Modulated Signals

An important class of signals are those that are modulated by a carrier frequency. Examples in optics include analysis of interference fringe patterns and modulation in optical telecommunications systems. An amplitude-modulated signal and its corresponding Fourier transform are
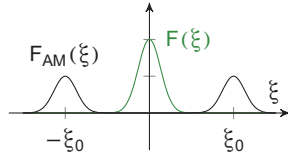
$$f_{AM}(x) = f(x)\cos(2\pi\xi_0 x)$$

$$F_{AM}(\xi) = \frac{F(\xi - \xi_0) + F(\xi + \xi_0)}{2}$$

Periodic Signals

An arbitrary periodic signal and its Fourier transform can be written in terms of a pulse-like function $p(x)$ as
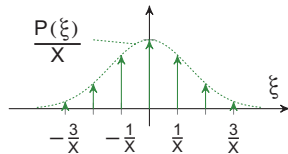
$$f_X(x) = p(x) * \frac{1}{|X|}\mathrm{comb}\left(\frac{x}{X}\right) \quad F_X(\xi) = P(\xi)\mathrm{comb}(X\xi)$$

The result is a **discrete spectrum**. The amplitude of the delta functions is determined by $P(\xi)$ and is equal to the Fourier series coefficients of the corresponding harmonics, $c_n = P(n\xi_0)$.

Correlation

Correlation is a measure of similarity between two functions. Define the **cross-correlation** between two complex functions $f(x)$ and $g(x)$ as

$$\gamma_{fg}(x) = f(x) \star g(x) = \int_{-\infty}^{\infty} f(x')g^*(x'-x)dx'$$

This operation closely resembles the convolution integral and can be written as

$$\gamma_{fg}(x) = f(x) * g^*(-x)$$

The argument of the cross-correlation can be thought of as a shift between the two functions that are being assessed for similarity. For real functions, the cross-correlation can be positive or negative. Negatively correlated functions are termed **anticorrelated**.

As a special case, the **autocorrelation function** is

$$\gamma_f(x) = \int_{-\infty}^{\infty} f(x')f^*(x'-x)dx' = f(x) * f^*(-x)$$
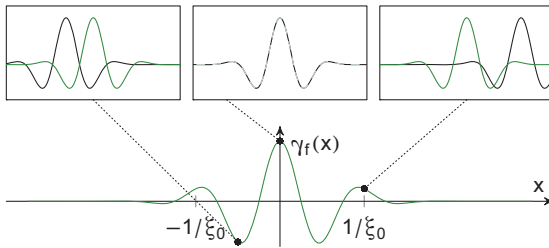
Given $f(x) = \mathrm{Gaus}(x/b)\cos(2\pi\xi_0 x)$,

$$\gamma_f(x) = \frac{b}{\sqrt{2}}\mathrm{Gaus}\left(\frac{x}{\sqrt{2}b}\right)\cos(2\pi\xi_0 x)$$

It is straightforward to use the Cauchy–Schwarz inequality to demonstrate that

$$\gamma_f(x) \le \gamma_f(0)$$

A function is most similar with itself when it is unshifted, and the equality holds only for periodic functions.

Convolution and Correlation in Two Dimensions

Convolution of functions of two dimensions is denoted as

$$g(x,y) = f(x,y) * h(x,y)$$

> Note that some publications distinguish 1D convolution from 2D convolution through the use of a double asterisks ($f(x,y) ** h(x,y)$), but this notation is not always used, and readers should be prepared to determine the dimensionality of the convolution from the context and the arguments.

As in the 1D case, the 2D convolution integral is written as

$$g(x,y) = \iint_{-\infty}^{\infty} f(x',y')h(x-x',y-y')dx'dy'$$

All of the properties derived for 1D convolution apply to 2D convolution, including the commutative, associative, and distributive properties, as well as the convolution theorem:

$$\mathcal{F}\{f(x,y) * h(x,y)\} = F(\xi,\eta)H(\xi,\eta)$$

For the special case where the two functions that are to be convolved are separable ($f(x,y) = f_1(x)f_2(y)$, $h(x,y) = h_1(x)h_2(y)$), the resulting function is also separable:

$$g(x,y) = f(x,y) * h(x,y)$$
$$g_1(x)g_2(y) = (f_1(x) * h_1(x))\,(f_2(y) * h_2(y))$$

Correlation is encountered in the computation of the optical transfer function. For example, $\gamma_{cyl}(r;2)$ is a result of correlating a cylinder that has radius 1 with a second cylinder that has radius 2.
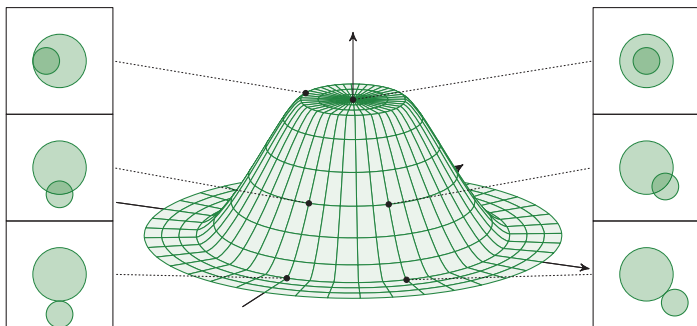
Band-Limited and Space-Limited Functions

The concepts of **band-limited** and **space-limited** functions are important in sampling theory. The implications of the two are apparent in this section.

A space-limited function has finite support of length L in the spatial domain. For convenience, the interval of support is defined as $x \in [0, L)$, but this choice is arbitrary:
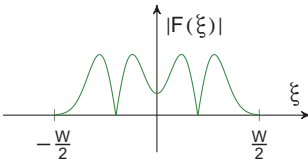
$$f(x) = \begin{cases} 0 & x < 0 \text{ or } x \geqslant L \\ f(x) & 0 \leqslant x < L \end{cases}$$

A band-limited function has finite support in the frequency domain. The full bandwidth W is contained within $|\xi| \leqslant \frac{W}{2}$:

$$F(\xi) = \begin{cases} 0 & |\xi| > \frac{W}{2} \\ F(\xi) & |\xi| \leqslant \frac{W}{2} \end{cases}$$

The properties of the Fourier transform make it so that a function cannot be simultaneously band limited and space limited (Heisenberg uncertainty principle). The only w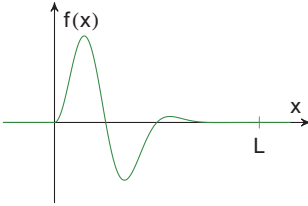ay that a function can be described by a finite spatial range of length L and a finite frequency range of width W is if the function is **simultaneously periodic** in both domains. This concept is leveraged in the definition of the **discrete Fourier transform** later in this section.
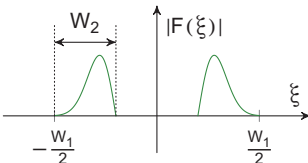
An alternate sampling strategy can be developed with $\xi_s < W_1$ when the signal technically occupies bandwidth W_1, yet can be confined to W_2.

Ideal Sampling

Signals in nature are generally continuous, while state-of-the-art detectors are composed of a discrete number of sensors. **Ideal sampling** refers to the approximation of the detector with a finite series of perfect δ-functions:

$$f_s(x) = f(x)\frac{1}{x_s}\mathrm{comb}\left(\frac{x}{x_s}\right) = \sum_{k=-\infty}^{\infty} f(kx_s)\delta(x - kx_s)$$

$$F_s(\xi) = F(\xi) * \mathrm{comb}(\xi x_s) = \xi_s \sum_{k=-\infty}^{\infty} F(\xi - k\xi_s)$$

In order to prevent **aliasing**, the signal needs to be sampled at a sufficiently high frequency. This threshold is often referred to as the **Nyquist** sampling frequency and is twice the signal's bandwidth, $\xi_N = 2\xi_w$. The following figures show an ideally sampled signal exactly at that threshold.
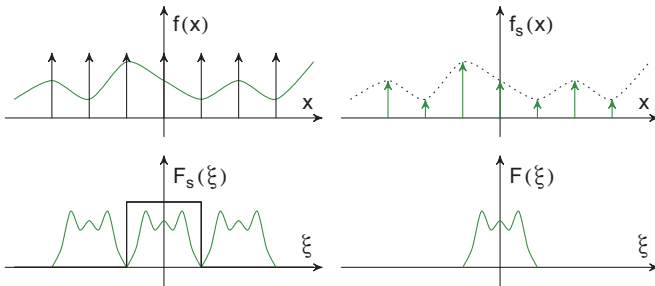
The rect function of width W represents an idealized **low-pass filter** (LPF), which can be written as

$$H(\xi) = \mathrm{rect}\left(\frac{\xi}{W}\right) \qquad F(\xi) = F_s(\xi)H(\xi)$$

$$h(x) = W\mathrm{sinc}(Wx) \qquad f(x) = f_s(x) * h(x)$$

The entire chain can be seen in the following figure.
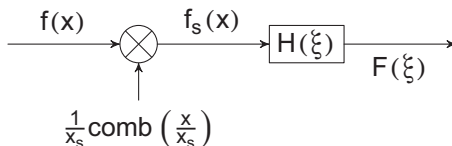
Sampling in Two Dimensions

Imaging applications, amongst others, require sampling in two spatial dimensions. The simplest 2D sampling strategy is a rectangular sample array with sampling intervals x_s and y_s:

$$f_s(x,y) = \sum_{k,m=-\infty}^{\infty} f(kx_s, my_s) \delta(x - kx_s, y - my_s)$$

$$F_s(\xi, \eta) = \xi_s \eta_s \sum_{k,m=-\infty}^{\infty} F(\xi - k\xi_s, \eta - m\eta_s)$$
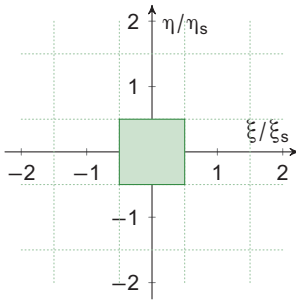
The Fourier transform of the sampled 2D signal is periodic with (ξ_s, η_s), and the rectangle bounded by the lines $\xi = \pm\xi_s/2$ and $\eta = \pm\eta_s/2$ defines the **first Nyquist zone**.

More-complicated sampling patterns are often useful. For example, the green channel in a Bayer color filter array omits every other sample in the grid. An equivalent sampling function is

$$f_s = f(x,y) \frac{1}{2}\left(1 + e^{i\pi\left(\frac{x}{x_s} + \frac{y}{y_s}\right)}\right) \frac{1}{x_s y_s} \text{comb}\left(\frac{x}{x_s}, \frac{y}{y_s}\right)$$

$$F_s = \left[F(\xi, \eta) + F\left(\xi - \frac{\xi_s}{2}, \eta - \frac{\eta_s}{2}\right)\right] * \text{comb}\left(\frac{\xi}{\xi_s}, \frac{\eta}{\eta_s}\right)$$

This result has its first Nyquist zone bounded by

$$\xi + \eta = \pm\frac{1}{2} \qquad\qquad \xi - \eta = \pm\frac{1}{2}$$
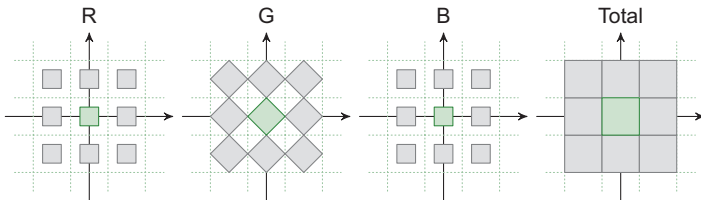
Non-Ideal Sampling

The concept of ideal point sampling is a useful mathematical tool, but real systems cannot sample at an infinitesimal point in space or time. Real sampling systems, such as focal plane arrays in cameras or digital sampling oscilloscopes, have some finite integration profile $p(x)$ associated with their physical response, and the output of the kth sample located at position x_k is
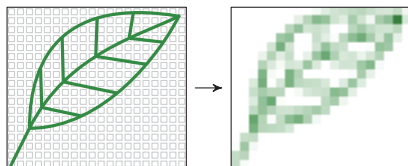
$$f_k = \int_{-\infty}^{\infty} f(x)p(x - x_k)dx$$

It is understood that x_k might be a position, time, wavelength, or other independent variable, and that the sampling could be in more than one dimension.
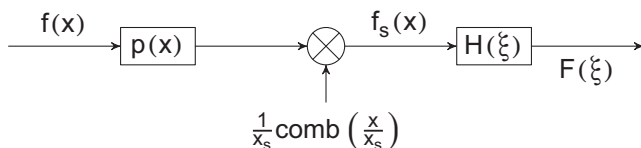
The definition of f_k indicates that **non-ideal sampling** can be dealt with by inserting a LPF with impulse response $p(x)$ in the flow diagram. The closer $p(x)$ is to $\delta(x)$, the closer the system is to ideal sampling. A common form is

$$p(x) = \text{rect}\left(\frac{x}{x_0}\right)$$

which represents a square pixel, a finite integration time, or other forms of non-ideal signal sampling. A rectangular sampling function introduces a LPF with transfer function

$$P(\xi) = x_0 \text{sinc}(x_0 \xi)$$

The implications of this transfer function are discussed later in the context of applications of linear systems.

Field Guide to Linear Systems in Optics

Aliasing

Aliasing is an artifact that occurs when the sampling rate is lower than the rate specified by the Nyquist condition. Consider the signal

$$f(x) = \frac{1}{2} + \frac{1}{2}\cos(2\pi x \xi_w)$$

$$F(\xi) = \frac{1}{2}\delta(\xi) + \frac{1}{4}\left[\delta(\xi - \xi_w) + \delta(\xi + \xi_w)\right]$$

and sample it at $\xi_s > 2\xi_w$ and at $\xi_s < 2\xi_w$:
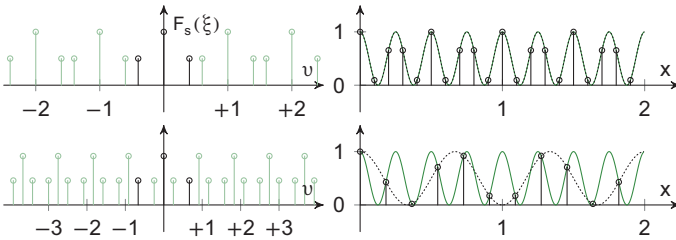
In the undersampled case ($\xi_s < 2\xi_w$), the contents of $F(\xi)$ are replicated too closely to each other, leading to high frequencies from the neighboring side lobes contributing to the central lobe that is used for reconstruction. This behavior yields a signal of a lower frequency than before.

Given that nothing is known about the signal *a priori*, the damage from aliasing is unrecoverable. The only way to "fix" the signal is to discard frequencies from the overlapping lobes. This can either be done before or after the sampling takes place. In either case, a loss in resolution is incurred, but it is better to discard to-be-aliased frequencies preemptively to avoid additional loss. In imaging, the effect is often referred to as **moiré pattern**.
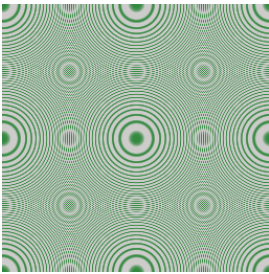
Band-Limited Reconstruction

Sampling is the process of converting a continuous signal $f(x)$ into a discrete sequence $f_k = f(kx_s)$. This process is practically implemented in an **analog-to-digital converter**. The resulting sequence can be manipulated in its digital form in a microprocessor, but if the end product is to be an audio signal, an image, etc., then the sequence must be converted back to a continuous observable.

There are essentially an infinite number of choices in terms of how the data will be interpolated between the sample points. Piecewise constant output (CD players), linear interpolation (most plotting routines), and spline interpolation are all common in different applications. **Band-limited reconstruction** uses a rectangular LPF with cutoff frequency $\xi_c/2$ to interpolate between the samples. In the space domain, the reconstructed signal $f_r(x)$ is

$$f_r(x) = f_s(x) * \xi_c \operatorname{sinc}(\xi_c x) = \sum_{k=-\infty}^{\infty} f_k \operatorname{sinc}\left(\xi_c(x - kx_s)\right)$$

When the signal is **oversampled**, then any choice of ξ_c that isolates $F(\xi)$ from $F_s(\xi)$ is acceptable, as shown in the bottom plot. The ultimate choice of bandwidth will depend on the signal-to-noise ratio (SNR) and the desired convergence. As seen in the figure, a choice of $\xi_c = \xi_s$ converges most rapidly. Narrower-bandwidth filters converge more slowly overall (more terms in the summation are required) because the sinc interpolator has larger spatial extent. Wider-bandwidth filters converge more slowly because of the more-rapid oscillations of the corresponding sinc function.
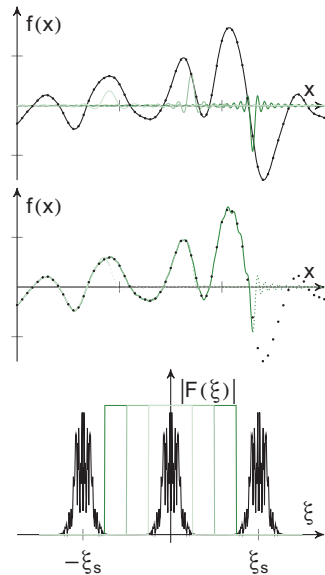
Discrete-Space Fourier Transform (DSFT)

Consider the direct computation of the Fourier transform of the ideally sampled signal:

$$F_s(\xi) = \mathcal{F}\{f_s(x)\} = \int_{-\infty}^{\infty} \sum_{k=-\infty}^{\infty} f(kx_s)\delta(x - kx_s)e^{-i2\pi\xi x}dx$$
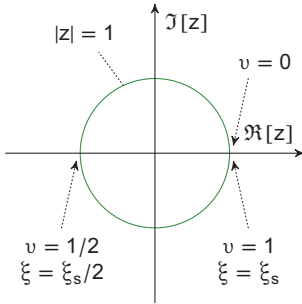
Using the properties of the δ-function, this becomes

$$F_s(\xi) = \sum_{k=-\infty}^{\infty} f_k e^{-i2\pi\xi kx_s}$$

where $f_k = f(kx_s)$ represents the discrete sequence of samples. A normalized frequency $\upsilon = \xi x_s$ can be defined that ranges from 0 to 1 as ξ varies from 0 to ξ_s. The series resulting from this change can be termed the **discrete-space Fourier transform** (DSFT).

$$F\left(e^{i2\pi\upsilon}\right) = \sum_{k=-\infty}^{\infty} f_k e^{-i2\pi\upsilon k} = F_s\left(\upsilon\xi_s\right)$$

The DSFT is periodic in υ by definition with period 1, which agrees with the periodicity required of $F_s(\xi)$. The notation $F\left(e^{i2\pi\upsilon}\right)$ highlights the link between the DSFT and the **z-transform** discussed on the following page.

It is often desired to design a **discrete filter** h_k that can be used to operate on the sequence f_k. The corresponding transfer function is

$$H\left(e^{i2\pi\upsilon}\right) = \sum_{k=-\infty}^{\infty} h_k e^{-i2\pi\upsilon k}$$

Defining the filter using the DSFT makes it inherently periodic, as well. Discrete convolution is defined as:

$$g_k = (f * h)_k = \sum_{k'=-\infty}^{\infty} f_{k'} h_{k-k'}$$

A discrete convolution theorem also holds:

$$G\left(e^{i2\pi\upsilon}\right) = F\left(e^{i2\pi\upsilon}\right) H\left(e^{i2\pi\upsilon}\right)$$

z-Transform

The **z-transform** is the discrete analog to the Laplace transform. Whereas the DSFT only considers real normalized frequencies $\upsilon \in [0, 1)$, the z-transform considers complex frequencies and is defined from the DSFT as

$$F(z) = \mathcal{Z}\{f_k\} = \sum_{n=-\infty}^{\infty} f_k z^{-k}$$

The variable z can assume all values in the complex plane, and the z-transform reduces to the DSFT when $z = e^{i2\pi\upsilon}$. The z-transform has a **region of convergence** like the Laplace transform, which includes the portions of the complex z-plane for which the transform is defined.
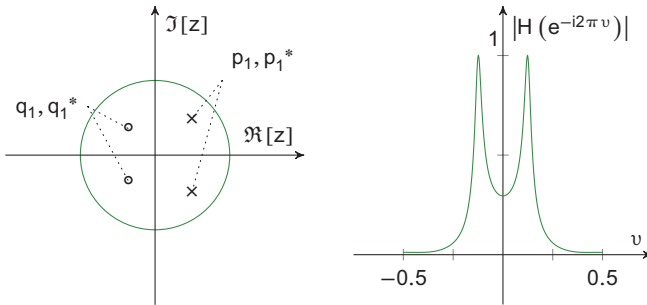
The z-transform is essential in digital signal processing, which is beyond the scope of this *Field Guide*. One of its key properties is the ability to design digital filters that have particular band-pass properties by placing **poles** and **zeroes** in the z-plane and defining the resulting transfer function in terms of the rational polynomial

$$H(z) = \frac{\prod_{m=1}^{M}(1 - q_m z^{-1})}{\prod_{n=1}^{N}(1 - p_n z^{-1})}$$

The N poles of the function in the z-plane are denoted as p_n and the M zeroes as q_m.

The z-transform has similar properties to its other Fourier cousins, including linearity, scaling, shifting, convolution, differentiation, conjugate, and reversal properties.

Discrete Fourier Transform (DFT)

The original function $f(x)$ was assumed to be band limited to satisfy

$$F(\xi) = 0, \ |\xi| \geqslant \frac{W}{2}$$

Now add the restriction that

$$f(x) = 0, \ x \leqslant 0, \ x \geqslant L$$

The sequence f_k therefore has $N = L/x_s$ samples. The Fourier transform of the ideally sampled function $F_s(\xi)$ is periodic with period ξ_s.
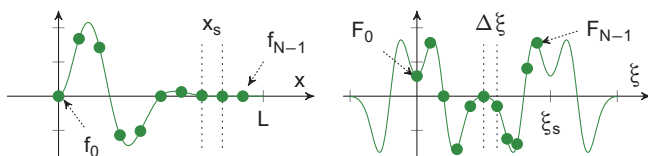
Define the interval in the frequency domain as

$$\Delta\xi = \frac{\xi_s}{N} = \frac{1}{L}$$

The following N-periodic sequence of samples provides a discrete representation of the Fourier transform of the ideally sampled functions:

$$F_n = F_s(n\Delta\xi) = \sum_{k=0}^{N-1} f_k e^{-i2\pi n(\Delta\xi)kx_s}$$

The product of the sampling intervals gives $(\Delta\xi)x_s = 1/N$, which leads to the **discrete Fourier transform** (DFT) and its inverse:

$$\mathcal{D}\{f_k\} = F_n = \sum_{k=0}^{N-1} f_k e^{-i2\pi\frac{nk}{N}}$$

$$\mathcal{D}^{-1}\{F_n\} = f_k = \frac{1}{N}\sum_{n=0}^{N-1} F_n e^{+i2\pi\frac{nk}{N}}$$

DFT Properties

The inverse DFT operation expresses the sequence f_k as a superposition of **complex exponential sequences**:

$$f_k = \frac{1}{N} \sum_{n=0}^{N-1} F_n e^{i2\pi \frac{kn}{N}}$$

The inner product between the kth and l^{th} sequences is

$$\langle \psi_k, \psi_l \rangle = \sum_{n=0}^{N-1} e^{i2\pi(k-l)n} = \frac{1 - e^{i2\pi(k-l)}}{1 - e^{i2\pi\left(\frac{k-l}{N}\right)}}$$

This ratio is the periodic analog to the sinc function, and evaluating it on the integers produces the **Kronecker delta**.

$$\delta_{k,l} = \begin{cases} 1 & k = l \\ 0 & k \neq l \end{cases}$$
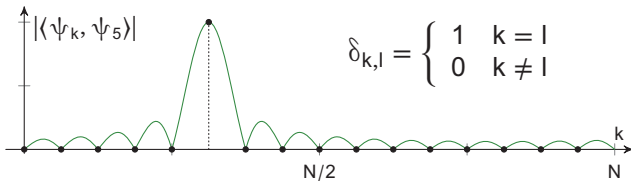
Linearity/Superposition	
$\mathcal{D}\{af_k + bg_k\} = aF_n + bG_n$	
Periodicity in Both Domains	
$f_{k+N} = f_k$	$F_{n+N} = F_n$
Convolution Theorem	
$f_k g_k \xrightarrow{\mathcal{D}} \sum_{n'=0}^{N} F_{n'} G_{n-n'}$	$F_n G_n \xrightarrow{\mathcal{D}^{-1}} \sum_{k'=0}^{N} f_{k'} g_{k-k'}$
Kronecker Delta	
$\delta_{k,k_0} \xrightarrow{\mathcal{D}} \exp\left[-i2\pi\frac{k_0 n}{N}\right]$	$\delta_{n,n_0} \xrightarrow{\mathcal{D}^{-1}} \exp\left[i2\pi\frac{n_0 k}{N}\right]$
Shifted	Modulated
$f_{k-k_0} \xrightarrow{\mathcal{D}} \exp\left[-i2\pi\frac{k_0 n}{N}\right] F_n$	$\exp\left[i2\pi\frac{n_0 k}{N}\right] f_k \xrightarrow{\mathcal{D}^{-1}} F_{n-n_0}$

DFT Evaluation

In many instances in optics, the DFT is used for convenience to approximate the true Fourier transform of the underlying continuous function using a computer. When this is the case, care must be taken in choosing the parameters of the approximation.

The spatial sampling interval x_s, frequency resolution $\Delta\xi$, and sequence length N are all closely related:

$$x_s = \frac{1}{\xi_s} = \frac{L}{N} \qquad \Delta\xi = \frac{1}{L} = \frac{\xi_s}{N} = \frac{1}{Nx_s}$$

A technique known as **zero padding** is commonly used to affect the frequency resolution of a DFT. By choosing a larger value of L, the windowed function $g(x)$ can be made to more closely resemble the true function $f(x)$. Adding more zeroes at the end of a sequence f_n does not alter the computation of the DSFT; it only increases the frequency resolution by making $\Delta\xi$ smaller.

The analysis above implies that

$$F_n = F(n\Delta\xi)$$

when the window of length L does not alter the function $f(x)$. However, the original assumption that went into ideal sampling was that the function $f(x)$ had to be **band limited**. The presence of the window means that $g(x)$ is **space limited**. The properties of the Fourier transform dictate that a function cannot be both simultaneously. The implied periodicity of the DFT resolves this matter because

$$F_{n+N} = F_n \ \text{ and } \ f_{k+N} = f_k$$

This makes it clear that the function is neither band limited nor space limited but is instead **periodic** in both domains. The periodicity property of the DFT leads to **frequency leakage** and other effects that can be mitigated by zero padding and the use of window functions.

Continuous and Discrete Fourier Domains

$\xi \backslash x$	Continuous	Discrete
Continuous	**Fourier Transform** $f(x) \rightarrow$ aperiodic $F(\xi) \rightarrow$ aperiodic $f(x) = \int_{-\infty}^{\infty} F(\xi)e^{+i2\pi\xi x}d\xi$ $F(\xi) = \int_{-\infty}^{\infty} f(x)e^{-i2\pi\xi x}dx$ 	**DSFT** $f_k \rightarrow$ aperiodic $F\left(e^{i2\pi\upsilon}\right) \rightarrow$ periodic $\upsilon = \xi/\xi_s$ $F\left(e^{i2\pi\upsilon}\right) = \sum f_k e^{i2\pi\upsilon}$
Discrete	**Fourier Series** $f(x) \rightarrow$ periodic $F(\xi) = \sum c_n \delta(\xi - n\xi_0)$ $f(x) = \sum_{-\infty}^{\infty} c_n e^{i2\pi n\xi_0 x}$ $c_n = \frac{1}{X}\int_0^X f(x)e^{-i2\pi n\xi_0 x}dx$ 	**DFT** $f_k \rightarrow$ periodic $F_n \rightarrow$ periodic $f_k = \frac{1}{N}\sum_{n=0}^{N-1} F_n e^{+i2\pi\frac{nk}{N}}$ $F_n = \sum_{k=0}^{N-1} f_k e^{-i2\pi\frac{nk}{N}}$

Gibbs Phenomenon and Frequency Leakage

The implied periodicity of the DFT has consequences when sampling actual, continuous signals. Consider an arbitrary sinusoidal signal

$$f(x) = \cos\left(2\pi\xi_0 x\right)$$

If this signal is sampled at rate ξ_s, the sampled sequence is

$$f_k = \cos\left(2\pi k\xi_0/\xi_s\right), \ k = 0, \dots, N-1$$

The implied periodicity of the DFT requires that $f_N = f_0$, but the underlying signal $f(x)$ has this same periodicity only if

$$\cos\left(2\pi N\xi_0/\xi_s\right) = 1$$

This condition is only satisfied if $\xi_0 = m\left(\xi_s/N\right)$, where m is an arbitrary integer.

$$f(x) = \cos(6\pi x), \ N = 18$$

$$g(x) = \cos(4.8\pi x), \ N = 18$$
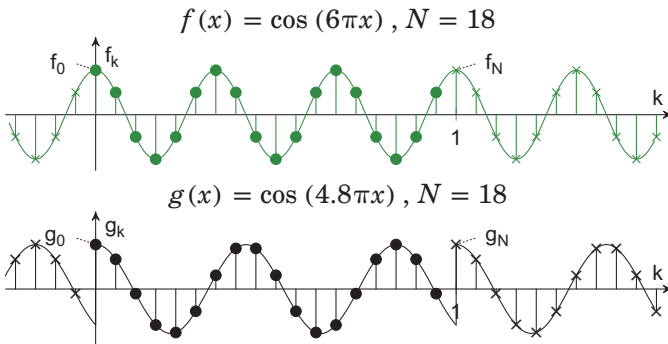
The artificial discontinuities introduced at the end of the sampling interval result in the **Gibbs phenomenon** in the frequency domain that broadens the DFT and is sometimes termed **frequency leakage** because the energy that should be at a single DFT point "leaks" into neighboring samples.
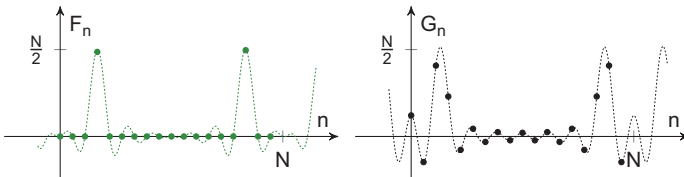
Windowing of Sequences

The fact that a finite interval $0 \leq x < L$ is used in computing the DFT creates a windowed version of the function $f(x)$:

$$g(x) = f(x)\,\text{rect}\left(\frac{x - \frac{L}{2}}{L}\right)$$

$$G(\xi) = F(\xi) * Le^{-i2\pi\frac{L}{2}\xi}\text{sinc}\left(L\xi\right)$$

The DSFT and the DFT produce discrete estimates of the Fourier transform of $g(x)$. The highest fidelity is generally achieved with a wider window ($L \to \infty$, $L\text{sinc}(L\xi) \to \delta(\xi)$).

Because of this property, specialized **window functions** $w(x)$ have been developed for discrete signal processing that have particular properties in either the space or frequency domain. In the following examples, $\bar{x} = 2\pi x/L$. The effect of each window in the space and frequency domains is

$$g(x) = w(x)f(x) \qquad G(\xi) = W(\xi) * F(\xi)$$

Space	Frequency
Gaussian — $w(x) = \text{Gaus}(x/b)$ Maximally smooth in both domains 	
Hanning — $w(x) = 0.5\,(1 - \cos\bar{x})$ Most basic member of raised cosine family 	
Hamming — $w(x) = 0.54 - 0.46\cos\bar{x}$ Similar to Hanning with first side lobe nulled 	
Flat Top — $w(x) = 1 - 1.93\cos\bar{x} + 1.29\cos\bar{x} - 0.388\cos\bar{x} + 0.028\cos\bar{x}$ Designed to prevent frequency leakage 	

Fast Fourier Transform (FFT)

The number of operations in the traditional calculation of the DFT scales with $O(N^2)$. The **fast Fourier transform** (FFT) provides a speedup by leveraging a more-efficient algorithm that scales with $O(N \log N)$.

The most-basic implementation is the Radix-2 decimation-in-time (DIT) method, depicted below for $N = 16$. The illustration demonstrates how grouping the inputs together can combine N contributions in fewer steps than when they are added one-by-one.
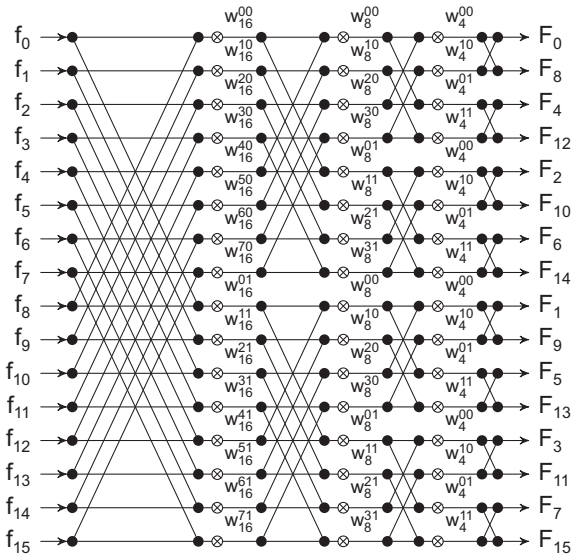
The $\otimes$ indicates multiplication by twiddle factors that are often precalculated for a given problem size N:

$$w_N^{nk} = e^{-i2\pi \frac{kn}{N}}$$

The optimization arises from the recursive splitting of the DFT into two parts, allowing for a $N / \log_2 N$ speedup. This optimization requires that $\log_2 N$ is an integer. This limitation is easily overcome with zero padding or other philosophically equivalent optimizations written for more-abstractly factored input sizes.

Discrete Convolution

Define **discrete convolution** of two sequences $f * h$ to be

$$g_k = (f * h)_k = \sum_{m=-\infty}^{\infty} f_m h_{k-m}$$

For finite sequences, summation limits can be narrowed to only accommodate the effective support of both sequences:

$$g_k = (f * h)_k = \sum_{m=-M}^{M} f_m h_{k-m}$$

An alternative way of calculating convolution is

$$G_n = F_n H_n \quad \rightarrow \quad g_k = \mathcal{D}^{-1}\left\{F_n H_n\right\}$$

which is faster to compute given the optimizations available for DFT. The caveat is that the DFT method introduces an implied periodicity, and as a result, the beginning of one period may interact with the end of the other. This calculation is denoted as **circular convolution**, as opposed to the previously introduced **linear convolution**.
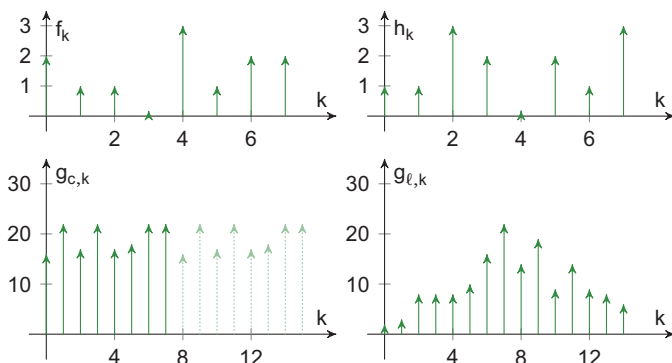
Two sequences of size N_1 and N_2 need to be padded with $\Delta N_1 = N_2 - 1$ and $\Delta N_2 = N_1 - 1$ zeroes, respectively, in order for circular and linear convolutions to produce the same results.

Interpolation and Decimation

Interpolation and **decimation** alter the number of samples in a discrete sequence. They are different from **upsampling** and **downsampling** in that they include bandlimited filtering to prevent aliasing.

Denote **interpolation** and **decimation** factors with L and M, respectively. The original signal $f(x)$ is sampled at the rate ξ_s to produce the sequence $f_{1,k}$. The following table depicts the processing steps involved in resampling that signal: interpolating it first ($\ell = 1 \rightarrow \ell = 3$) and decimating it second ($\ell = 3 \rightarrow \ell = 5$).

ℓ	$f_{\ell,k}$	$F_{\ell,s}$
1	$\sum_{k=-\infty}^{\infty} f(x)\delta(x - kx_s)$	$\xi_s \sum_{k=-\infty}^{\infty} F_1(\xi - k\xi_s)$
2	$\sum_{m=-\infty}^{\infty} f_{1,m}\delta(k - mL)$	$\sum_{k=-\infty}^{\infty} f_{2,k}e^{-i2\pi\xi kx_s}$
3	$\sum_{m=-\infty}^{\infty} f_{2,k}\operatorname{sinc}\left(\frac{k-Lm}{L}\right)$	$F_{2,s}(\xi)H_{\text{int}}(\xi)$
4	$\sum_{k=-\infty}^{\infty} f_{3,k}\operatorname{sinc}\left(\frac{k}{M}\right)$	$F_{3,s}(\xi)H_{\text{dec}}(\xi)$
5	$\sum_{m=-\infty}^{\infty} f_{4,k}\delta(m - kM)$	$\sum_{k=-\infty}^{\infty} f_{5,k}e^{-i2\pi\xi kx_s}$

These steps are graphically illustrated in the figures that follow using

$$f(x) = \text{Gaus}(x)\sin(2\pi\xi_o x)$$

$$F(\xi) = \frac{1}{2i}\left[\text{Gaus}(\xi - \xi_o) - \text{Gaus}(\xi + \xi_o)\right]$$

with $\xi_o = 2$ and an original sample rate $\xi_s = 12$. The function is then interpolated and decimated with $L = 2$ and $M = 3$. With higher decimation factors, the resulting sequence will have a lower band limit than the original signal, making resampling an inherently lossy process.

The post-filter $H_{\text{int}}(\xi)$ and the pre-filter $H_{\text{dec}}(\xi)$ are often assumed to be perfect LPFs. The choice of filter will enable further control over the information's frequency retention, which might be preferable for some tasks.

Interpolation and Decimation (cont.)

$f_{\ell,k}$	$\Im\{F_\ell\}$	$\Im\{F_{\ell,s}\}$

$\ell = 1$

↑ L   Adds L−1 zero samples after each sample.

$\ell = 2$

H_{int} Filters the sequence to discard Fourier side lobes.

$\ell = 3$

H_{dec} Filters frequencies not supported by decimation.

$\ell = 4$

↓ M   Discards M−1 samples from every group of M.

$\ell = 5$
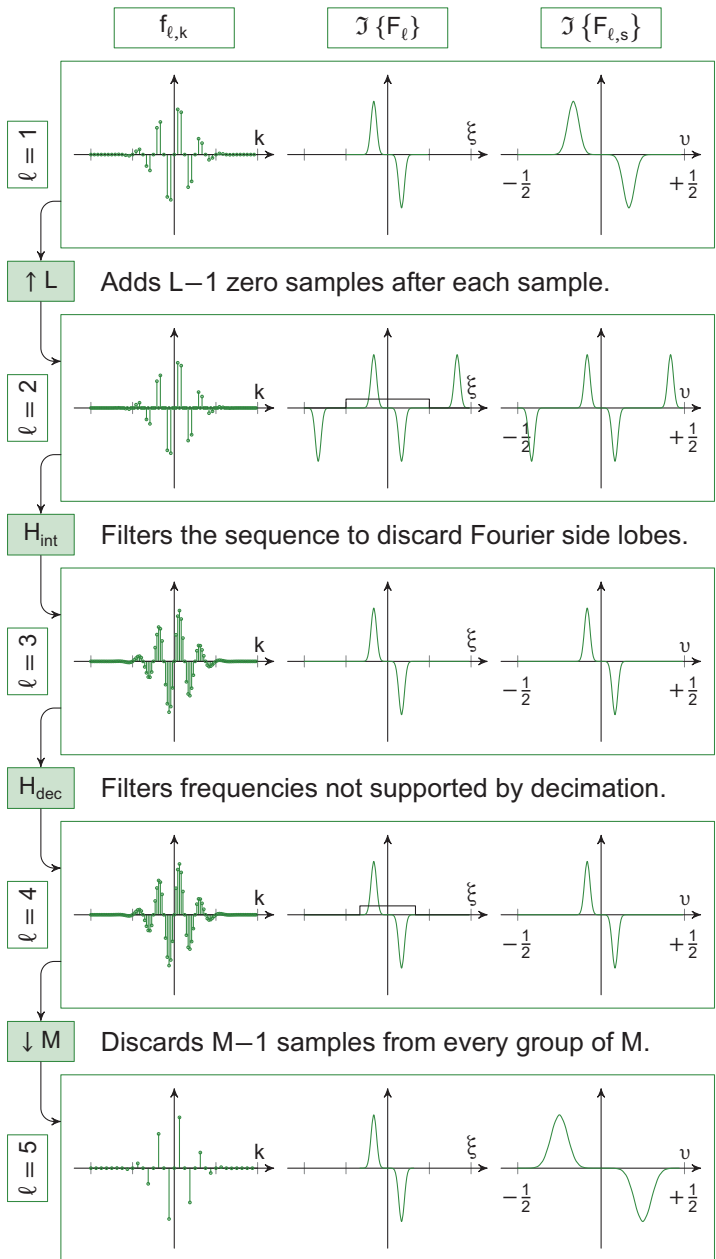
Filters

A **filter** is an operator that is used to shape or alter a function in a controlled manner. When the filter is LSI, it is defined in terms of its impulse response $h(x)$ or $h(t)$ and its transfer function $H(\xi)$ or $H(\nu)$.

All physical processes act as filters, and important examples encountered in optical applications include the following:

- Optical filters that are used to select a specific range of wavelengths of the optical spectrum. Such filters can be interference filters based on the **Fabry–Pérot** effect or traditional colored-glass filters that preferentially absorb particular wavelength ranges.

- An optical imaging system can be described with an **optical transfer function** (OTF) that generally acts as a LPF, ultimately limiting system resolution.

- Optical detectors used in telecommunications have a non-ideal response to an impulse in time, and this acts as a filter on the transient signal that is being measured.

- The traditional tool of **optical signal processing** uses the Fourier transforming properties of lens-based optical systems to instantaneously perform filtering operations in a $4f$-system. This is the basis of optical character recognition, which has largely been replaced by using DFTs in digital computers.

- The finite-size pixels in a focal plane array act as a LPF that limits the resolution of the imaging system. The pixel response and the OTF of the optics must be considered together for optimum performance.

- Turbulence in the atmosphere alters the phase of optical radiation and impacts the performance of astronomical instruments. Modern **adaptive optics** systems compensate for this by monitoring the phase filtering effects of the atmosphere and canceling them.

Amplitude-Only Filters

A general filter can be mathematically decomposed as a cascade of an **amplitude-only filter** and a phase-only filter:

$$H(\xi) = H_A(\xi)H_\Phi(\xi)$$

The amplitude-only filter attenuates the relative amounts of the various frequencies, but does not alter the respective phase relationships among them. Three important classes of such filters are **LPFs**, which preferentially attenuate high frequencies, **high-pass filters** (HPFs), which preferentially attenuate low frequencies, and **bandpass filters** (BPFs), which preferentially allow a range of frequencies to pass. Note that a BPF can be obtained as a cascade of a LPF and an HPF; however, in the example shown below, the BPF is not related to the LPF and HPF.
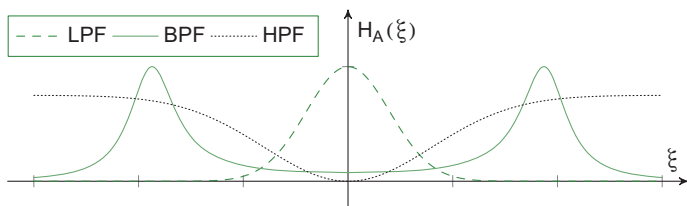
LPFs tend to smooth out the sharp transitions in a sequence, whereas HPFs tend to only allow a signal to pass near edges. Because BPFs prefer a range of frequencies, those tend to dominate the output. The following plot shows the effect of each of the above filters on an ideal square wave. Note that because the HPF and BPF attenuate the $\xi = 0$ component, the resulting outputs have both positive and negative values, unlike the square-wave input.
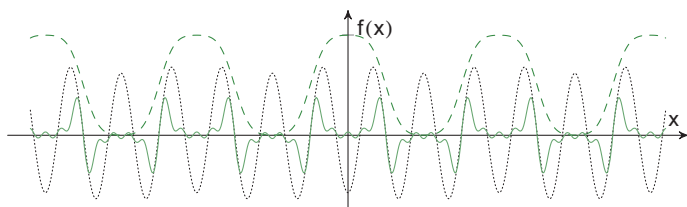
Phase-Only Filters

Phase-only filters have the special property that they do not alter the energy contained at any given frequency, but instead alter the relative phase relationship among the frequencies that make up a given signal. This alteration is often referred to as **distortion**. Although the power spectral density (PSD) may be unchanged, the actual waveform could be unrecognizable.

As an example, consider the truncated Fourier series:

$$f(x) = \frac{1}{2} + \frac{1}{2} \sum_{n=1}^{9} \text{sinc}\left(\frac{n}{2}\right) \cos\left(2\pi n \xi_0 x - \phi_0\right)$$

When $\phi_0 = 0$, this is an approximation to a square wave, but other values lead to distortion.
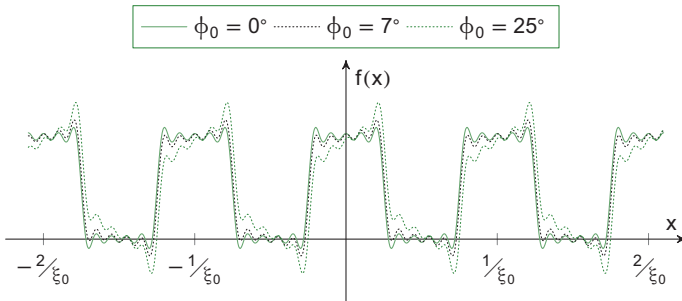
This behavior is produced by a filter that possesses a constant phase offset (in degrees) that is equal and opposite for positive and negative frequencies so that the symmetry properties of a real function are obeyed.

$$H_\Phi(\xi) = e^{-i2\pi\phi_0 \text{sgn}(\xi)}$$
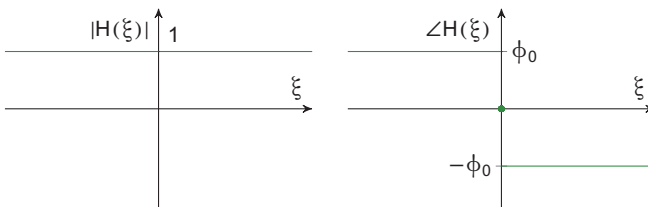
Special Classes of Phase Filters

An important subclass of phase-only filters is the **linear-phase-only filter**, which has a transfer function with phase that increases linearly with frequency:

$$H(\xi) = e^{i2\pi x_0 \xi} \qquad h(x) = \delta(x - x_0)$$

The properties of Fourier transforms and the impulse function mean that for linear-phase-only filters

$$f(x) * h(x) = f(x - x_0)$$

A linear-phase-only filter can be combined with an arbitrary filter to produce a shifted offset.

Consider the phase-only filter with transfer function

$$H(\xi) = e^{i\Phi(\xi)}$$

where the phase angle is small, i.e., $\Phi(\xi) \ll 1$. In this case, the transfer function can be approximated using the Taylor series for an exponential as

$$H(\xi) \approx 1 + i\Phi(\xi)$$

The corresponding impulse response for this filter is

$$h(x) = \delta(x) + i\phi(x)$$

Such **weak phase-only filters** have special importance in optics since small changes in the optical quality of materials and surfaces can lead to phase perturbations that alter the performance of an optical system. Examples of weak phase-only filters in optics include the following:

- Aberrations in optical systems perturb the phase across the aperture.
- Roughness in optical surfaces leads to a random alteration of optical phase.
- Small variations in the index of refraction in a medium act as weak phase filters and can limit system performance.
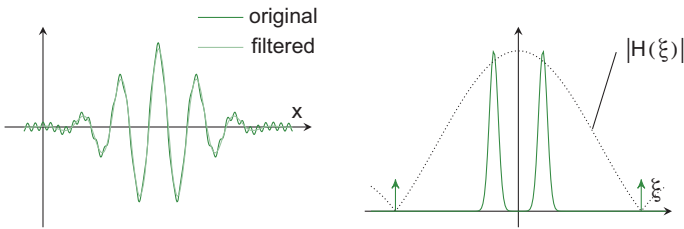
Equalization

A physical measurement process such as a camera taking a picture, a microphone recording an audio signal, or an oscilloscope measuring an electronic transient corrupts the true signal due to the effects of the measurement system. When that measurement can be modeled as LSI, the measurement of the signal $s(x)$ by a system with impulse response $h(x)$ is

$$f(x) = s(x) * h(x)$$

The process of **equalization**, or **deconvolution**, attempts to remove the effects of the system from the measurement to produce an estimate $\hat{s}(x) = h^{-1}(x) * f(x)$ that is as close to the true signal as possible. The operator $h^{-1}(x)$ is an **inversion operator**. Mathematically, it is trivial to write

$$h^{-1}(x) = \mathcal{F}^{-1}\left\{\frac{1}{H(\xi)}\right\}$$

In practice, there are a number of items that come into consideration for this operation to work. Even in the absence of noise, a real system will have certain frequencies for which the transfer function amplitude is so low that the attempted inversion is ill-posed.

Consider the example of a rectangular moving average:

$$f(x) = s(x) * \text{rect}(x)$$

In the provided illustration the transfer function nulls out the higher-frequency components of the signal, which can never be recovered by equalization. In real applications, the transfer function does not have to completely vanish to cause problems. Once the SNR at any frequency drops below some threshold, the data are unrecoverable.

Matched Filtering

A classic filter for detecting known targets is the **matched filter**. This filter was originally developed to detect radar returns from targets, but it has been extended to many other fields of target detection.

Consider the problem of detecting a known signal $s(x)$ in the presence of noise $n(x)$. Both signals have unit energy so that

$$\|s(x)\|^2 = \int_{-\infty}^{\infty} |s(x)|^2 dx = \|n(x)\|^2 = 1$$

Parseval's theorem can be used to write the **signal-to-noise ratio** (SNR) from a filter with impulse response $h(x)$ as

$$\text{SNR} = \frac{\|s(x) * h(x)\|^2}{\|n(x) * h(x)\|^2} = \frac{\|S(\xi)H(\xi)\|^2}{\|N(\xi)H(\xi)\|^2}$$

An important relationship known as the Cauchy–Schwarz inequality states that

$$\|A(x)B(x)\|^2 \leqslant \|A(x)\| \|B(x)\|$$

The equality is only obtained for $A(x) = B^*(x)$. The numerator is thus maximized by choosing the matched filter

$$H(\xi) = S^*(\xi) \qquad h(x) = s^*(-x)$$

The output of the filter for input $f(x)$ is

$$g(x) = s^*(-x) * f(x) = f(x) \star s(x) = \gamma_{fs}(x)$$

For this reason, the matched filter is also referred to as a **correlation-based filter**. There are extensions to the matched filter for use when there are interferers or other sources of non-white noise, and it can be shown that the matched filter is optimum when the noise is Gaussian. Matched filtering can be used in optical character recognition (OCR), where a character is correlated with a generic character template.
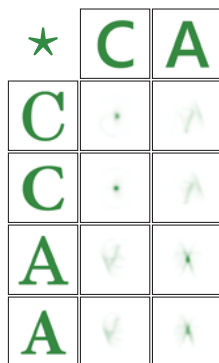
Projection-Slice Theorem

Consider a 2D function $f(x, y)$ and some physical operation that takes a **projection** through the function

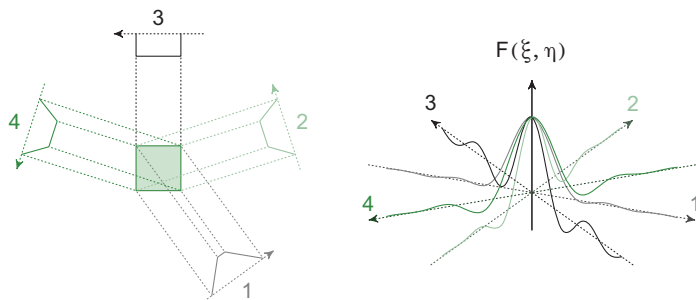$$p_f(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

The projection operation can be written in terms of the 2D convolution

$$p_f(x) = f(x, y) * \delta(x)$$

The 2D Fourier transform of $p_f(x)$ and the 2D Fourier transform of $f(x, y)$ are related as

$$\mathcal{F}\{p_f(x)\}(\xi, \eta) = F(\xi, 0)\delta(\eta)$$

The Fourier transform of the **projection** of $f(x, y)$ has produced a **slice** of the 2D Fourier transform $F(\xi, \eta)$.

This example could equally well have been in terms of a projection in the x-direction or in an arbitrary direction, for that matter. The **projection-slice theorem** is one of the foundations of **computed tomography** (CT). In CT systems, a set of projections is taken through the object at different angles, each of which produces a slice of the 2D Fourier transform of the object. These slices can then be compiled into a sampled version of the 2D Fourier transform, allowing $f(x, y)$ to be computed even though it may not be possible to measure it directly.

Random Functions and Sequences

In most practical applications, the signals are not known exactly; instead, they are **random signals**. A continuous function $f(x)$ may be random, and this section also considers **random sequences**, i.e., sampled versions of random signals. A very simple example of a random sequence is one due to the addition of a **deterministic signal** $s(x)$ with random noise $n(x)$:

$$f(x) = s(x) + n(x)$$

The probability that a particular sample f_k will lie between $z_0 - \Delta z/2$ and $z_0 + \Delta z/2$ is given by the **probability distribution function** (PDF):

$$\text{PDF} = p_f(z_0|k)\,dz$$

- An **ensemble** is a collection of **instantiations** or observations of a random process that capture its full statistical behavior.
- The PDF is defined to have unit area

$$\int_{-\infty}^{\infty} p_f(z)\,dz = 1$$

- The **expected value** of a function of the sequence is often referred to as the **mean**, ensemble average, or first moment:

$$E\left[g(f_k)\right] = \int_{-\infty}^{\infty} g(z) p_f(z|k)\,dz$$

- The **variance** is the second central moment:

$$Var\left[f_k\right] = E\left[\left(f_k - E\left[f_k\right]\right)^2\right] = E\left[f_k{}^2\right] - E\left[f_k\right]^2$$

- A **stationary** sequence has statistics that are independent of k. The sequence is **wide-sense stationary** if at least its mean and covariance are independent of k.
- A sequence is **ergodic** over some parameter if the expectation operation can be replaced by an average over the parameter

$$E\left[f(x)\right] = \lim_{X \to \infty} \frac{1}{X} \int_{-X/2}^{X/2} f(x)\,dx$$

A signal can simultaneously be ergodic in one parameter (e.g., space) and not in another (e.g., time).
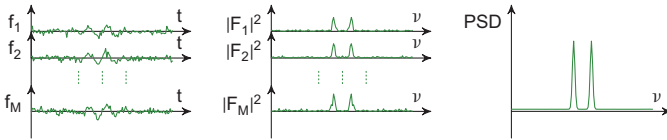
Power Spectral Density (PSD) Function

Consider an amplitude-modulated waveform with profile $p(t)$ corrupted by Gaussian random noise:

$$f(t) = f_0 \sin(2\pi\nu_0 t)\, p(t) + n(t)$$

Each observation produces a single **instantiation** of the process. The Fourier transform of the mth instantiation is

$$F_m(\nu) = \frac{f_0}{2i}\left(P(\nu - \nu_0) - P(\nu + \nu_0)\right) + N_m(\nu)$$

The **power spectral density** (PSD) is equivalent to the **energy spectral density** (ESD) per unit time. It can be defined in terms of the statistics of the amplitude of the Fourier transform as

$$\text{PSD}(\nu) = E\left[|F(\nu)|^2\right]$$

This quantity is always real, non-negative, and it has units proportional to the power or energy per unit frequency, depending on the nature of the field quantity. The phase information has been lost, which is common in random processes.

The properties of the Fourier transform relate the PSD to the autocorrelation function

$$\mathcal{F}^{-1}\{F(\nu)F^*(\nu)\} = f(t) * f^*(-t) = \gamma_f(t)$$

When $f(t)$ is a random process that is **ergodic** and **wide-sense stationary**, its statistical autocorrelation is related to the PSD by the **Wiener–Khinchin Theorem**:

$$r_{ff}(\tau) = E\left[f(t)f^*(t-\tau)\right] = \int_{-\infty}^{\infty} f(t)f^*(t-\tau)dt = \gamma_f(\tau)$$

A common use of the PSD in optics is the measurement of the **spectrum** of light. This is illustrated later in the context of Fourier transform spectroscopy.

Field Guide to Linear Systems in Optics

Filtering Random Signals

A random process $f(x)$ can be passed through a deterministic filter with impulse response $h(x)$. The specific output will depend on the exact details of the mth instantiation of the input

$$g_m(x) = f_m(x) * h(x)$$

Even though little can be said *a priori* about the specific output, the effect of the filter on the PSD of the signal can be easily predicted. In the frequency domain, the output spectrum can be written as

$$\text{PSD}_G(\xi) = E\left[F(\xi)H(\xi)F^*(\xi)H^*(\xi)\right] = \left|H(\xi)\right|^2 \text{PSD}_F(\xi)$$

Note that when working with random processes, the square magnitude of the transfer function is indicative of average performance, and the phase information drops out.

The SNR compares the relative amounts of power or energy in the desired signal and the undesired noise. It is often specified in terms of the PSDs:

$$\text{SNR} = \frac{\text{PSD}(f(x))}{\text{PSD}(n(x))}$$

In many applications, SNR is expressed in decibels, which is a logarithmic power ratio:

$$\text{SNR}_{\text{dB}} = 10\log_{10}\text{SNR}$$

Equalization can, in principle, be performed at any frequency where $H(\xi) \neq 0$. However, in real measurements, noise is present, and equalization should only be attempted for frequencies where $\text{SNR} > \tau$ (τ is a threshold related to the fidelity of the measuring process). This example shows the inverse filter for all frequencies with $\text{SNR} > 10\text{dB}$.

—— $|H(\xi)|$ —— PSD_n —— $|H_{10\text{dB}}^{-1}(\xi)|$
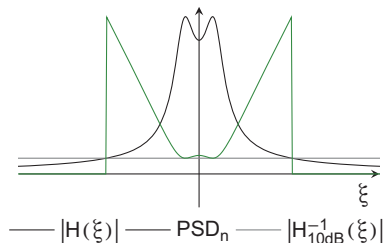
Wiener–Helstrom Filter

The **Wiener–Helstrom filter** is a filter that minimizes the square error between a signal and the estimate of the signal, provided that PSDs of the signal and the noise and system's transfer function are known.

Express the measurement $f(t)$ as the underlying signal $s(t)$ measured by a system characterized by its impulse response $h(t)$ and perturbed by additive noise $n(t)$:

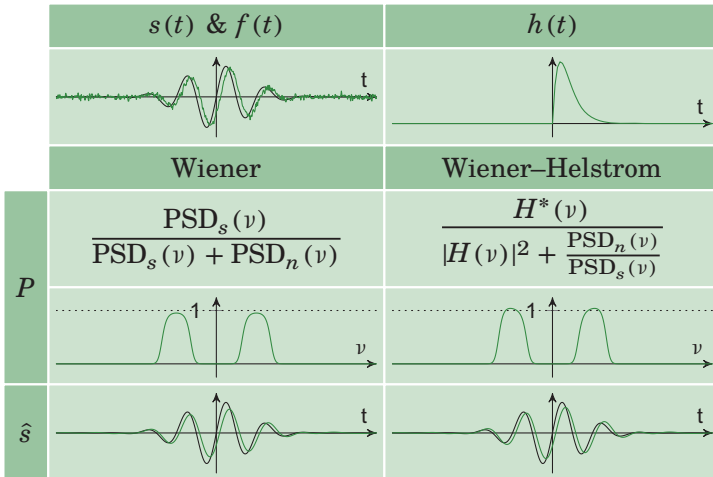$$f(t) = s(t) * h(t) + n(t) \qquad F(\nu) = S(\nu)H(\nu) + N(\nu)$$

The **Wiener filter** derivation is the same except for $h(t) = \delta(t)$. Apply a linear filter $p(t)$ to estimate $s(t)$:

$$\hat{s}(t) = p(t) * f(t) \qquad\qquad \hat{S}(\nu) = P(\nu)F(\nu)$$

Leveraging Parseval's theorem, the error between the signal and its estimate can be written as

$$e = \left\langle \int_{-\infty}^{\infty} |s(t) - \hat{s}(t)|^2 \, dt \right\rangle = \left\langle \int_{-\infty}^{\infty} \left|S(\nu) - \hat{S}(\nu)\right|^2 \, d\nu \right\rangle$$

The functional form of the filter follows:

$s(t)$ & $f(t)$		$h(t)$	
Wiener		**Wiener–Helstrom**	
P	$\dfrac{\mathrm{PSD}_s(\nu)}{\mathrm{PSD}_s(\nu) + \mathrm{PSD}_n(\nu)}$		$\dfrac{H^*(\nu)}{|H(\nu)|^2 + \dfrac{\mathrm{PSD}_n(\nu)}{\mathrm{PSD}_s(\nu)}}$
$\hat{s}$			

In many cases, $h(t) \approx \delta(t)$ is a good enough approximation, as can be seen with the scenario depicted in the example.

Modes

The concept of **basis functions** was introduced earlier to break up an arbitrary function into a superposition of basis functions that had particular properties of interest:

$$f(x) = \sum_{k=1}^{N} a_k \psi_k(x)$$

The number of basis functions N can be finite, countably infinite, or uncountably infinite, in which case the summation becomes an integral.

An important application of linear systems in optics is to treat propagation of optical fields through a system using basis functions that satisfy the **wave equation**

$$\nabla^2 \psi(\mathbf{r}) - \frac{1}{c^2} \frac{\partial^2 u}{\partial t^2} = 0$$

where ∇^2 is the **Laplacian**, and c is a constant. Solutions to this equation that satisfy particular boundary conditions are termed **modes**. Several important examples follow:

Mode Type	Uses
Optical Rays	Optical design
Uniform Plane Waves	Propagation in stratified structures and media
Hermite–Gauss Beams	Modes of an optical cavity and coherent optical design
Laguerre–Gauss Beams	Analysis of systems with azimuthal symmetry
Huygens Wavelets	Diffraction as superposition of spherical wavefronts
TE/TM Plane Waves	Modes of an optical waveguide

The scalar optical field is denoted as $u(x, y, z)$. The notation **E** is used to explicitly indicate the vector electric field where necessary.

Plane Wave Spectrum

The direction of propagation of a plane wave is
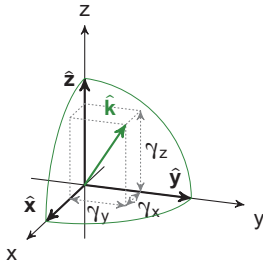
$$\hat{\mathbf{k}} = \gamma_x\hat{\mathbf{x}} + \gamma_y\hat{\mathbf{y}} + \gamma_z\hat{\mathbf{z}}$$

where γ_x, γ_y, and γ_z are the **direction cosines** given by the cosine of the angle between $\hat{\mathbf{k}}$ and the coordinate axes, and they satisfy

$$\gamma_x{}^2 + \gamma_y{}^2 + \gamma_z{}^2 = 1$$

A plane wave with **wavenumber** $k = 2\pi/\lambda$ propagating along $\hat{\mathbf{k}}$ can be expressed as

$$u(x,y,z) = u_0 e^{ik(\gamma_x x + \gamma_y y + \gamma_z z)}$$

Consider the fields in the $z = 0$ plane written in terms of their Fourier transform:

$$u(x,y,z=0) = \iint_{-\infty}^{\infty} U(\xi,\eta,z=0)e^{i2\pi(\xi x + \eta y)}d\xi d\eta$$

The direction cosines can be expressed in terms of the spatial frequencies as

$$\gamma_x = \lambda\xi \qquad \gamma_y = \lambda\eta \qquad \gamma_z = \sqrt{1 - \lambda^2\left(\xi^2 + \eta^2\right)}$$

The inverse Fourier transform is just a superposition of plane waves propagating in different directions, which gives rise to the term **plane wave spectrum** or **angular spectrum**. For spatial frequencies

$$\lambda^2\left(\xi^2 + \eta^2\right) \leqslant 1$$

the direction cosine γ_z is purely real, and the plane wave is **propagating**. For spatial frequencies

$$\lambda^2\left(\xi^2 + \eta^2\right) > 1$$

the direction cosine γ_z is imaginary, and the plane wave is **evanescent**. Evanescent waves decay exponentially and can only be observed in the near field.

Transfer Function/Impulse Response of Free Space

After the field's plane wave spectrum is determined, it can be propagated from $z = z_0$ to $z = z_1$ using the **transfer function of free space**:
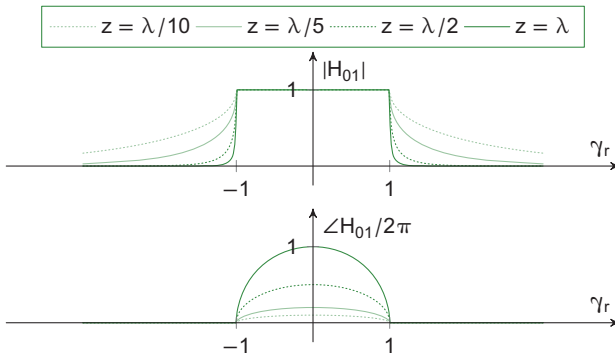
$$U(\xi, \eta, z = z_1) = U(\xi, \eta, z = z_0)H_{01}(\xi, \eta)$$

where H_{01} is defined as

$$H_{01}(\xi, \eta) = e^{ik\gamma_z(z_1 - z_0)}$$

and adds the proper amount of phase and decay to the propagating and the evanescent waves, respectively.

Defining $\gamma_r = \sqrt{\gamma_x{}^2 + \gamma_y{}^2}$ reveals the following structure:

Alternatively, starting with **Weyl's integral**

$$\frac{e^{ikR}}{ikR} = \frac{-1}{2\pi} \iint_{-\infty}^{\infty} e^{ik(\gamma_x x + \gamma_y y + \gamma_z z)} \, d\Omega, \text{ where } d\Omega = \frac{\lambda^2 d^2\rho}{\gamma_z}$$

it is possible to show that

$$\frac{\partial}{\partial z}\left[\frac{e^{ikR}}{2\pi R}\right] = \iint_{-\infty}^{\infty} e^{ik\gamma_z z} e^{i2\pi(\xi x + \eta y)} \, d^2\rho = \mathcal{F}_{xy}^{-1}\left\{e^{ik\gamma_z z}\right\}$$

The **transfer function** and **impulse response** are, in fact, Fourier pairs, even though they were derived from different physical arguments:

$$h_{01}(x, y) = \mathcal{F}^{-1}\left\{H_{01}(\xi, \eta)\right\}$$

This lays the groundwork for the general concept of diffraction as a convolution of the field with the impulse response.

Propagation of Optical Beams

An optical beam is an example of a field distribution that is well suited for analysis by its plane wave spectrum. Consider a 1D Gaussian beam with fields at the $z = 0$ plane, given as
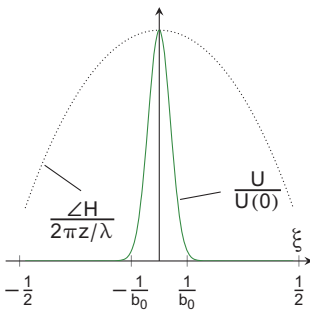
$$u_0(x, y) = \text{Gaus}\left(\frac{x}{b_0}\right)$$

The plane wave spectrum corresponding to these fields is

$$U_0(\xi, \eta) = b_0 \text{Gaus}(b_0 \xi)\,\delta(\eta)$$

The fields at an arbitrary plane z can be written as

$$u_z(x, y) = \mathcal{F}^{-1}\left\{U_0(\xi, \eta)H_{0z}(\xi, \eta)\right\}$$

At the plane $z = 0$ (the **beam waist**), all of the plane wave components are in phase, and the result is a Gaussian beam with width b_0. As the observation plane moves away from the waist, the phase relationship among the Fourier components is altered by the transfer function of free space $H(\xi; z)$. The beam widens spatially, and the phase front becomes curved.

Amplitude Profile   **Wavefront Shape**
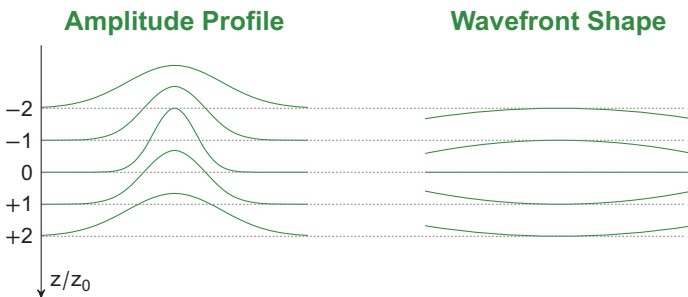
Spatial and Temporal Coherence

Coherence in an optical system refers to the ability of two different beams of light to constructively and destructively interfere. When coherent beams are coincident at a point, then their complex field amplitudes must be added:

$$|u_{\text{tot}}|^2 = |u_1 + u_2|^2 = |u_1|^2 + |u_2|^2 + 2\Re\left[u_1 u_2{}^*\right]$$

However, when incoherent beams are coincident, their irradiances are the quantity that adds linearly:

$$|u_{\text{tot}}|^2 = |u_1 + u_2|^2 = |u_1|^2 + |u_2|^2$$

Temporal coherence is the ability of a beam to interfere with a copy of itself that is shifted in time. Temporal coherence is measured in an amplitude division interferometer, such as a Michelson. This interferometer uses a beamsplitter to create two copies of the incident beam. One of these copies is delayed in time by an amount ℓ/c before both are brought together to interfere at the detector.
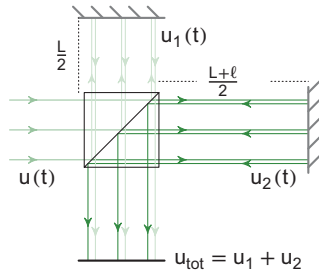
Spatial coherence is the ability of the light from two different locations on the wavefront to interfere. Spatial coherence is measured in a wavefront division interferometer, such as a Young's double-pinhole interferometer. The difference in path lengths ℓ_1 and ℓ_2 produces a variable time delay at the observation point.
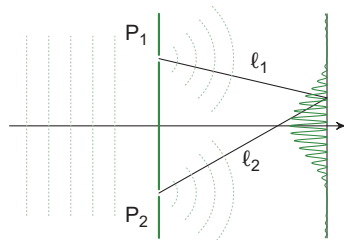
This *Field Guide* considers light that is either completely coherent or completely incoherent, but most physical sources of light are best modeled as **partially coherent** in space and/or time.

Diffraction

The scalar **Rayleigh–Sommerfeld diffraction** calculation is derived from Dirichlet boundary conditions and is correct for all z more than a few wavelengths beyond the aperture.
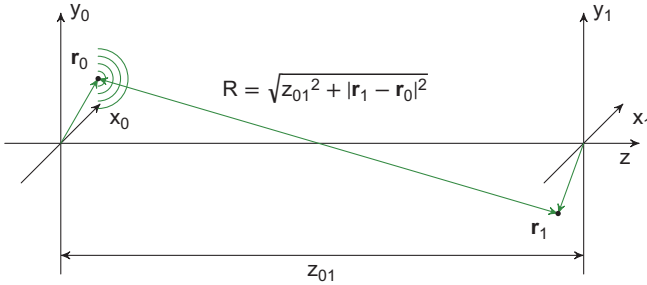
This calculation employs the **Huygens wavelet**, which is directly related to free space propagation and arises as a z-derivative of the spherical wave:

$$h_z^{\text{Huy}}(\mathbf{r}_1 - \mathbf{r}_0) = \frac{-1}{2\pi} \frac{\partial}{\partial z} \left[\frac{e^{ikR}}{R} \right] = \frac{-1}{2\pi} \left(ik - \frac{1}{R} \right) \frac{z_{01}}{R} \frac{e^{ikR}}{R}$$

The use of h for the wavelet is not coincidental — it represents the impulse response. The propagated field is then the convolution of the field at $z = z_0$ with the wavelet propagated to $z = z_1$:

$$u(\mathbf{r}_1) = \iint_{-\infty}^{\infty} u(\mathbf{r}_0) h_z^{\text{Huy}}(\mathbf{r}_1 - \mathbf{r}_0) \, d^2 r_0$$

Although the Huygens formulation is mathematically exact, scalar diffraction itself is not physically accurate near the aperture, and a more-useful result follows by requiring that $z \gg \lambda$. Aside for a phase shift, this restriction allows a Huygens wavelet to be approximated with a spherical wave producing the **Rayleigh–Sommerfeld diffraction integral**:

$$u(\mathbf{r}_1) \approx \frac{z_{01}}{i\lambda} \iint_{-\infty}^{\infty} u(\mathbf{r}_0) \frac{e^{ikR}}{R^2} \, d^2 r_0$$

Further approximations are often applied that lead to even simpler calculations.

Field Guide to Linear Systems in Optics

## **Paraxial Approximation and Scalar Diffraction**

The first sensible approximation within the **scalar diffraction** context is usually the **paraxial approximation**. It eases the calculation by setting trigonometric functions of small angles to only include the first order of their respective Taylor expansions:

$$\sin(\theta) \approx \theta \qquad \tan(\theta) \approx \theta \qquad \cos(\theta) \approx 1$$

More specifically, it is used to ignore γ_x and γ_y in amplitude calculations by setting $\gamma_z = 1$, which leads to $R \approx z_{01}$:

$$u(\mathbf{r}_1) \approx \frac{1}{i\lambda z_{01}} \iint_{-\infty}^{\infty} u(\mathbf{r}_0) e^{ikR} d^2 r_0$$

These angle approximations are only applied to the amplitude term of the calculation because the consequence of applying them to the phase will provide trivial and extremely inaccurate results. As can be seen on the next page, an extra Taylor term is kept for the phase in the Fresnel diffraction approximation.
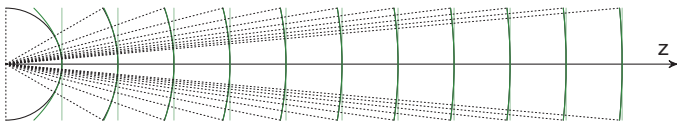
The above figure illustrates why the paraxial approximation is appropriate—as the propagation distance increases, the radius of the spherical wave increases, and its contribution looks increasingly like a plane wave.
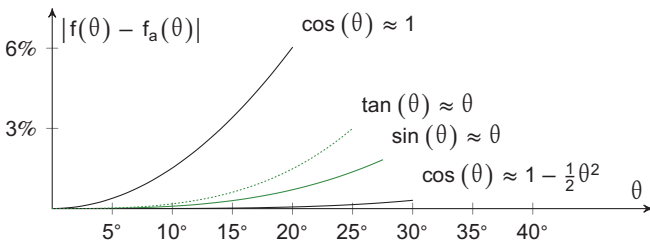
The addition of the θ^2 term to the cosine approximation provides a much-broader range of accuracy.

Fresnel Diffraction

Fresnel diffraction refers to a set of approximations that simplify the calculation, while maintaining accuracy over a broad range. The distance R from a point in the aperture to the observation point is expanded in a Taylor series:

$$R = \sqrt{z_{01}^2 + |\mathbf{r}_1 - \mathbf{r}_0|^2} = z_{01} + \frac{|\mathbf{r}_1 - \mathbf{r}_0|^2}{2z_{01}} - \frac{|\mathbf{r}_1 - \mathbf{r}_0|^4}{8z_{01}^3} + \cdots$$

Depending on where R appears, a different approximation is required to achieve sufficient accuracy.

Amplitude	$R \approx z_{01}$	$z_{01} \gg |\mathbf{r}_1 - \mathbf{r}_0|$
Phase	$R \approx z_{01} + \frac{|\mathbf{r}_1 - \mathbf{r}_0|^2}{2z_{01}}$	$z_{01}^3 \gg \frac{\pi|\mathbf{r}_1 - \mathbf{r}_0|^4}{4\lambda}$

The resultant impulse response in the Fresnel region is

$$h_z^{\text{Fre}}(\mathbf{r}_1 - \mathbf{r}_0) = \frac{e^{ikz_{01}}}{i\lambda z_{01}} \exp\left[\frac{i\pi|\mathbf{r}_1 - \mathbf{r}_0|^2}{\lambda z_{01}}\right]$$

$$|\mathbf{r}_1 - \mathbf{r}_0|^2 = r_1^2 - 2(\mathbf{r}_0 \cdot \mathbf{r}_1) + r_0^2$$

which contributes quadratic phase curvature to the initial and the propagated fields. The dot product term

$$\exp\left[\frac{-2\pi i(\mathbf{r}_0 \cdot \mathbf{r}_1)}{\lambda z_{01}}\right]$$

can be readily recognized as a Fourier transform kernel where $\mathbf{r}_0$ and $\mathbf{r}_1/(\lambda z_{01})$ are the conjugate variables. This results in

$$u(\mathbf{r}_1) = \frac{e^{ikz_{01}}}{i\lambda z_{01}} \iint_{-\infty}^{\infty} u(\mathbf{r}_0) \exp\left[\frac{i\pi|\mathbf{r}_1 - \mathbf{r}_0|^2}{\lambda z_{01}}\right] d^2r_0$$

$$= \frac{e^{ikz_{01}}}{i\lambda z_{01}} \exp\left[\frac{i\pi r_1^2}{\lambda z_{01}}\right] \mathcal{F}\left\{u(\mathbf{r}_0) \exp\left[\frac{i\pi r_0^2}{\lambda z_{01}}\right]\right\}\Bigg|_{\boldsymbol{\rho} = \frac{\mathbf{r}_1}{\lambda z_{01}}}$$

Even though the Fourier transform shifts the result into the frequency domain, the evaluation properly scales the frequencies into their propagated spatial equivalents.

Fraunhofer Diffraction

Fraunhofer diffraction is commonly referred to as far-field radiation. The calculation adds another approximation on top of the set of approximations used for Fresnel diffraction:

$$\exp\left[\frac{i\pi r_0{}^2}{\lambda z_{01}}\right] \approx 1 \;\rightarrow\; \frac{\pi r_0{}^2}{\lambda z_{01}} \ll 1 \;\rightarrow\; z_{01} \gg \frac{\pi r_0{}^2}{\lambda}$$

allowing the quadratic curvature to be ignored and further simplifying the impulse response:

$$h_z^{\text{Fra}}(\mathbf{r}_1 - \mathbf{r}_0) = \frac{e^{ikz_{01}}}{i\lambda z_{01}} \exp\left[\frac{i\pi r_1{}^2}{\lambda z_{01}}\right] \exp\left[\frac{-2\pi i(\mathbf{r}_0 \cdot \mathbf{r}_1)}{\lambda z_{01}}\right]$$

The radiated field can then be written as

$$u(\mathbf{r}_1) = \frac{e^{ikz_{01}}}{i\lambda z_{01}} \exp\left[\frac{i\pi r_1{}^2}{\lambda z_{01}}\right] \mathcal{F}\{u(\mathbf{r}_0)\}\big|_{\boldsymbol{\rho} = \frac{\mathbf{r}_1}{\lambda z_{01}}}$$

All of the diffraction integral calculations can be thought of as a progressively simplified contribution from each of the secondary source's wavelets. Due to the nature of free space propagation, each wavelet expands, and as the propagation distance increases, the importance of keeping extra terms to achieve accurate results decreases.
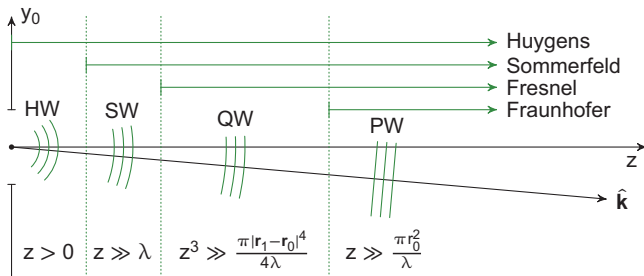

In all of the aforementioned cases, the assumed linear shift invariance enables relatively straightforward evaluation.

Fraunhofer/Fresnel Basis Functions

Fresnel and Fraunhofer diffraction integrals implicitly separate functions into basis functions. To help visualization, the complex exponentials in the Fresnel integral can be split into four combinations of purely real **basis functions**:

$$\exp\left(-2\pi i x\xi\right)\exp\left(i\pi x^2/\lambda z\right) \; \rightarrow \; {\textstyle \frac{\cos}{\sin}}\left(2\pi x\xi\right){\textstyle \frac{\cos}{\sin}}\left(\pi x^2/\lambda z\right)$$

Below are $(x\xi)$–$(\lambda z\xi^2)$ planes of the four basis functions.
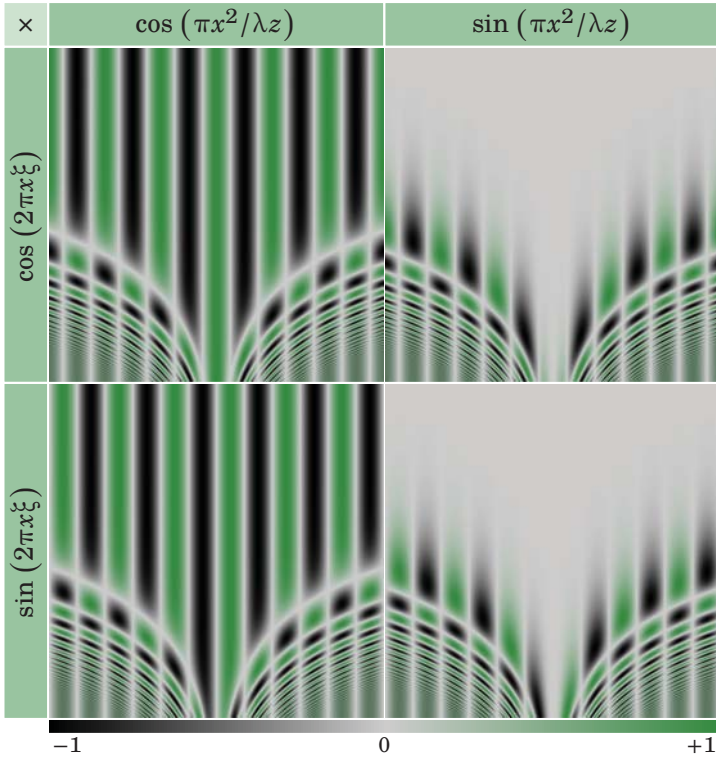
As $\lambda z\xi^2 \rightarrow \infty$, $\sin\left(\pi x^2/\lambda z\right) \rightarrow 0$ and $\cos\left(\pi x^2/\lambda z\right) \rightarrow 1$. This parallels the Fraunhofer limit and has the effect of simplifying the field contribution to that of a plane wave and reducing the basis functions to $\cos\left(2\pi x\xi\right)$ and $\sin\left(2\pi x\xi\right)$. The point at which this approximation is appropriate depends on the frequency content of the problem.

Fourier Transforming Properties of Lenses

A **lens** applies a transformation onto the propagating field:

$$t_{\text{lens}}(\mathbf{r}) = A \exp\left[-\frac{i\pi r^2}{\lambda f}\right] t_{\text{ap}}(\mathbf{r})$$

A is a complex constant. Setting $A = 1$ and using Fresnel diffraction to propagate the object field through the lens from $z = -f$ to $z = +f$ yields

$$u_2(\mathbf{r}_2) = \frac{e^{ik2f}}{i\lambda f} \mathcal{F}\{u_1(\mathbf{r}_1)\}\Big|_{\boldsymbol{\rho} = \frac{\mathbf{r}_f}{\lambda f}} = \frac{e^{ik2f}}{i\lambda f} U_1\left(\frac{\mathbf{r}_2}{\lambda f}\right)$$

which reveals that a lens takes a Fourier transform of the field. This property allows the construction of a **4f-system**, which enables optical frequency content manipulation.
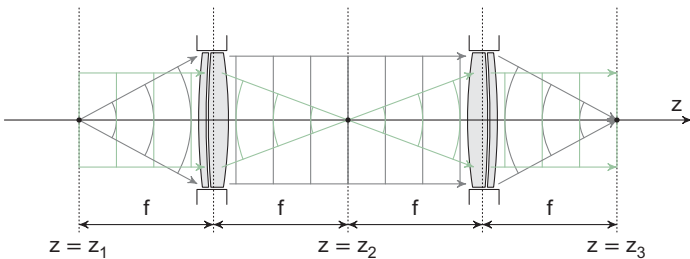
The object field's Fourier domain can be altered by placing a filter at the intermediate plane $z = z_2$:

$$u_2(\mathbf{r}_2) = \frac{e^{ik2f}}{i\lambda f} U_1\left(\frac{\mathbf{r}_2}{\lambda f}\right) t_{\text{filter}}(\mathbf{r}_2)$$

resulting in the following field at the image plane $z = z_3$:

$$u_3(\mathbf{r}_3) = \frac{e^{ik2f}}{i\lambda f} \mathcal{F}\left\{\frac{e^{ik2f}}{i\lambda f} U_1\left(\frac{\mathbf{r}_2}{\lambda f}\right) t_{\text{filter}}(\mathbf{r}_2)\right\}$$

$$= -e^{ik4f} u_1(\mathbf{r}_3) * T_{\text{filter}}\left(\frac{\mathbf{r}_3}{\lambda f}\right)$$

The result can be seen as a convolution of the object field with a blur impulse response. The additional phase can be associated with the total propagation distance, $4f$.

Fourier Description of Optical Cavity Modes

Consider a **confocal cavity** formed by two spherical mirrors separated by their common radius of curvature R. The paraxial focus of both mirrors is at the center of the cavity, one focal length $f = R/2$ from each of the two mirrors.

Stable **cavity modes** have two resonance conditions:

Transverse	$u^+(\mathbf{r}, z = 0) = \pm u^-(\mathbf{r}, z = 0)$, where u^+ and u^- are the forward and backward propagating waves in the cavity
Longitudinal	$n2\pi$ accumulated phase in one round-trip including phase due to propagation, reflections, and Gouy phase shift

The transverse resonance condition must satisfy

$$\frac{e^{i2kf}}{i\lambda f} \mathcal{F} \left\{ u^+(\mathbf{r}, z = 0) \right\} \big|_{\boldsymbol{\rho} = \frac{\mathbf{r}}{\lambda f}} = \pm u^+(\mathbf{r}, z = 0)$$

The Gaussian distribution with $b_0 = \sqrt{\lambda f}$ obeys this relationship and has phase fronts that match the curvature of the spherical mirrors. For any z, the beam has the form

$$u(\mathbf{r}, z) = \frac{b_0^2}{b_0^2 + i\lambda z} e^{ikz} \exp\left[\frac{-\pi r^2}{b(z)^2} \right] \exp\left[\frac{i\pi}{\lambda R(z)} \right]$$

$$b(z) = b_0 \sqrt{1 + \left(\frac{\lambda z}{b_0^2} \right)^2} \qquad R(z) = z \left[1 + \left(\frac{b_0^2}{\lambda z} \right)^2 \right]$$

The **Rayleigh range**, $z_0 = b_0^2 / \lambda$, denotes the location where the wavefront has the minimum radius of curvature.

Beam Wavefronts
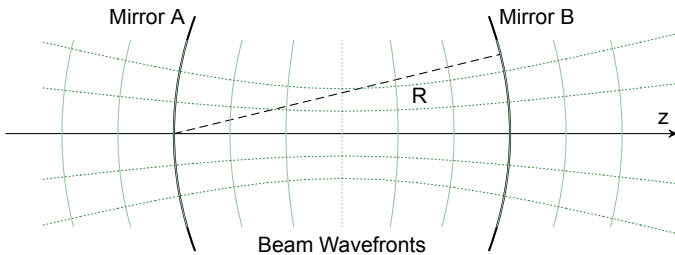
Higher-Order Cavity Modes

Hermite–Gaussian beams are eigenfunctions of the 1D Fourier transform operator and give separable solutions that satisfy the transverse resonance condition in terms of the **Hermite polynomials** $H_n(x)$:
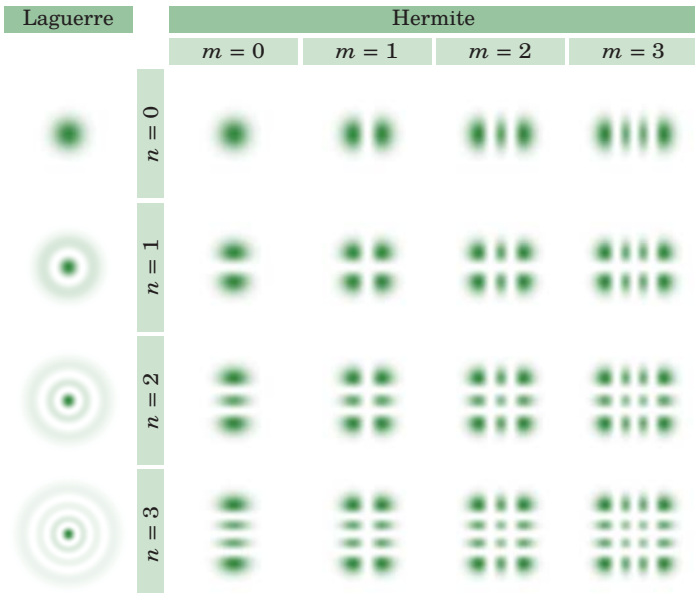
$$u_{mn}(x,y,z=0) = A_{mn}H_m\left(\frac{\sqrt{2\pi}x}{b_0}\right)H_n\left(\frac{\sqrt{2\pi}y}{b_0}\right)\exp\left[\frac{-\pi r^2}{b_0^2}\right]$$

$$H_n(x) = (-1)^n e^{x^2}\frac{d^n}{dx^n}e^{-x^2}$$

Laguerre–Gaussian beams are eigenfunctions of the Hankel transform operator and give azimuthally symmetric solutions that satisfy the transverse resonance condition in terms of the **Laguerre polynomials** $L_n(r)$:

$$u_n(\mathbf{r},z=0) = A_n L_n\left(\frac{2\pi r^2}{b_0^2}\right)\exp\left[\frac{-\pi r^2}{b_0^2}\right]$$

$$L_n(r) = \sum_{m=0}^{n}(-1)^m\binom{n}{m}\frac{x^m}{m!}$$

Laguerre		Hermite			
		$m=0$	$m=1$	$m=2$	$m=3$
$n=0$					
$n=1$					
$n=2$					
$n=3$					

Both solutions reduce to the Gaussian beam for $n=m=0$.

Slab Waveguides

A **slab waveguide** is an optical device with 1D confinement of the propagating wave. Only certain angles (α) or modes (m) can propagate. The range of propagating frequencies is limited by $kn_2 < \beta < kn_1$, where $\beta = k\cos(\alpha)$, the effective $\hat{z}$-wavenumber. Modes arise where the transverse propagation coefficient κ, the attenuation coefficient γ, and the boundary conditions match. Define

$$\left.\begin{array}{l} \kappa^2 = (n_1 k)^2 - \beta^2 \\ \gamma^2 = \beta^2 - (n_2 k)^2 \end{array}\right\} \quad \gamma^2 a^2 + \kappa^2 a^2 = (n_1{}^2 - n_2{}^2)k^2 a^2 = V^2$$
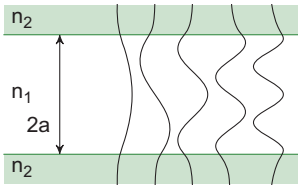
where V is the normalized frequency number, $\sqrt{n_1{}^2 - n_2{}^2}$ is the numerical aperture, and $2a$ is the thickness of the slab. The modes have the following structure:

$$u_m = \begin{cases} f_m(-\kappa_m a)e^{+\gamma_m(x+a)} & x < -a \\ f_m(\ \kappa_m x) & |x| < a \\ f_m(+\kappa_m a)e^{-\gamma_m(x-a)} & x > +a \end{cases}$$
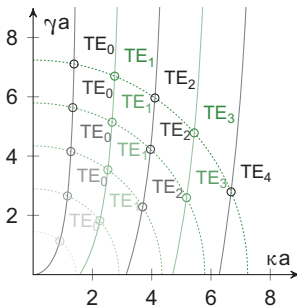
where f_m is cosine for even m and sine for odd m. The parameters κ_m and γ_m are intersections of $V^2 = \gamma^2 a^2 + \kappa^2 a^2$ and the lines denoted in the table below.

The plot only shows TE modes because for typical n_1 and n_2, TM modes are very similar.

	Even ($m = 0, 2, 4, ...$)	Odd ($m = 1, 3, 5, ...$)
TE	$\tan(\kappa a) = +\dfrac{\gamma}{\kappa}$	$\tan(\kappa a) = -\dfrac{\kappa}{\gamma}$
TM	$\tan(\kappa a) = +\dfrac{\gamma}{\kappa}\left(\dfrac{n_1}{n_2}\right)^2$	$\tan(\kappa a) = -\dfrac{\kappa}{\gamma}\left(\dfrac{n_2}{n_1}\right)^2$

Optical Fiber Waveguides

Optical fiber is now a ubiquitous optical element. It builds on the concepts of slab waveguides by adding azimuthal symmetry. The geometry bodes better for long-range transmission, but changes the modes of propagation. Define

$$\mathcal{J}_\nu = \frac{J_\nu'(\kappa a)}{\kappa J_\nu(\kappa a)} \qquad \mathcal{K}_\nu = \frac{K_\nu'(\gamma a)}{\gamma K_\nu(\gamma a)}$$

where J_ν and K_ν are Bessel functions, while κ, γ, and a hold the same meaning as before. Solving boundary conditions results in the following characteristic equation:

$$\frac{\beta^2 \nu^2}{a^2}\left(\frac{1}{\gamma^2} + \frac{1}{\kappa^2}\right) = \left(\mathcal{J}_\nu + \mathcal{K}_\nu\right)\left(k_1{}^2 \mathcal{J}_\nu + k_2{}^2 \mathcal{K}_\nu\right)$$

The modes' optical fields can be written as

$$u_z \propto \begin{cases} J_\nu(\kappa r)e^{i\nu\phi}e^{i(\omega t - \beta z)} & r < a \\ K_\nu(\gamma r)e^{i\nu\phi}e^{i(\omega t - \beta z)} & r > a \end{cases}$$

ν	Mode	Cutoff Condition
0	$\text{TE}_{0m}, \text{TM}_{0m}$	$J_0(\kappa a) = 0$
1	$\text{HE}_{1m}, \text{EH}_{1m}$	$J_1(\kappa a) = 0$
$\geqslant 2$	$\text{EH}_{\nu m}$	$J_\nu(\kappa a) = 0$
	$\text{HE}_{\nu m}$	$\left(\frac{n_1{}^2}{n_2{}^2} + 1\right)J_{\nu-1}(\kappa a) = \frac{\kappa a}{\nu - 1}J_\nu(\kappa a)$

For simplicity, Bessel functions are often approximated with simpler functions. The special case of J_1 is Gaussian-like

$$J_1(\kappa r) \approx \exp\left[\frac{-2r^2}{(ca)^2}\right], \; c = 0.65 + 1.619V^{-\frac{3}{2}} + 2.879V^{-6}$$

More generally:

$$J_\nu(\kappa r) \approx \sqrt{\frac{2}{\pi\kappa r}}\cos\left(\kappa r - \nu\frac{\pi}{2} - \frac{\pi}{4}\right) \quad K_\nu(\gamma r) \approx \frac{e^{-\gamma r}}{\sqrt{2\pi\gamma r}}$$

Diffraction-Limited Focal Imaging Systems
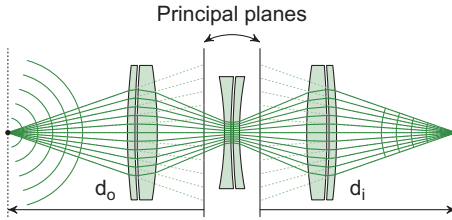
A **diffraction-limited** focal imaging system converts a perfect spherical wave at the **entrance pupil** into a perfect spherical wave at the **exit pupil**. The optical power of the system relates the radii of curvature of the spherical waves at the pupils in terms of distances from the **principal planes** to the object and the image planes:

$$\frac{1}{d_i} - \frac{1}{d_o} = \frac{1}{f}$$

Diffraction of this spherical wave from the exit pupil to the plane of focus produces the impulse response

$$\tilde{h}\,(x_i, y_i) = \mathcal{F}\{P(\lambda d_i \tilde{x}, \lambda d_i \tilde{y})\}$$

The pupil transmission function is $P(x,y)$, and the normalized pupil coordinates are

$$\tilde{x} = \frac{x}{\lambda d_i} \qquad\qquad \tilde{y} = \frac{y}{\lambda d_i}$$

Note that the normalized pupil coordinates carry units of spatial frequency, indicating that the pupil function is expected to be related to the transfer function of the system.

Using this impulse response allows the **physical optics** image fields to be written as a convolution with the geometrical optics prediction

$$u_i(x_i, y_i) = \tilde{h}\,(x_i, y_i) * u_g(x_i, y_i)$$

The geometrical optics field prediction for a system with magnification $M = d_i/d_o$ is

$$u_g(x_i, y_i) = \frac{1}{M} u_o\left(\frac{x_i}{M}, \frac{y_i}{M}\right)$$

Note that for systems forming a real image, $M < 0$.

Airy Disk

The class of optical systems with circular pupils is very important. These systems have the impulse response

$$\tilde{h}(x_i, y_i) = \mathcal{F}\left\{ \text{cyl}\left(\frac{r}{D}\right) \right\} = \frac{\pi D^2}{4} \text{somb}(D\rho)$$

For a point source, the observed intensity is the square of the impulse response, which is known as the **Airy disk**

$$\left| \tilde{h}(x_i, y_i) \right|^2 = \frac{\pi^2 D^4}{16} \text{somb}^2(D\rho)$$

This function has a first null that corresponds with the null of the Bessel function at $\pm 1.22 \lambda f/\#$. The first ring outside the main lobe has an amplitude that is 1.75% of the peak of the Airy disk. Lord

$\text{somb}^2(\rho)$

Rayleigh used this to define a resolution criterion, now known as the **Rayleigh criterion**, for the ability of a diffraction-limited system to resolve two point sources. He postulated that two incoherent point sources are barely resolvable when the center of one falls exactly on the first null of the second. With modern detectors, it is feasible to resolve two sources much closer than the Rayleigh limit, but the principle remains the same, and Rayleigh's limit need only be adjusted by a scaling factor related to system SNR. The following example shows two sources at the Rayleigh limit with no noise and with an SNR of 10 dB.

$-0.61\lambda f/\#$ $0.61\lambda f/\#$
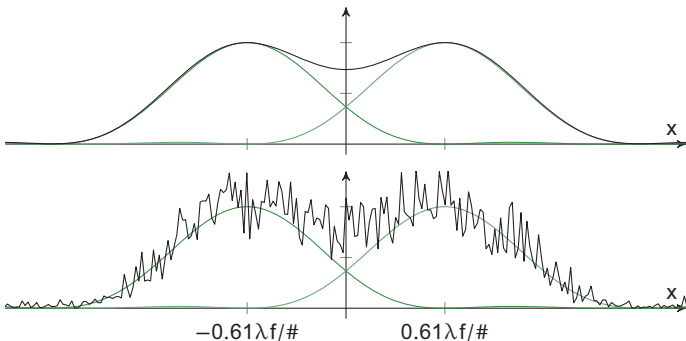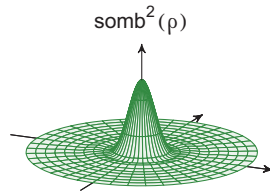
Coherent Transfer Function (CTF)

Because the system model is LSI, the image fields can also be determined in the spatial frequency domain:

$$U_i\left(\xi,\eta\right) = H\left(\xi,\eta\right)U_g\left(\xi,\eta\right)$$

The impulse response $\tilde{h}$ is already defined as the Fourier transform of the pupil function, leaving

$$H\left(\xi,\eta\right) = \mathcal{F}\{\mathcal{F}\{P\left(\lambda d_i x, \lambda d_i y\right)\}\} = P\left(-\lambda d_i\xi, -\lambda d_i\eta\right)$$

The function $H\left(\xi,\eta\right)$ is known as the **coherent transfer function** (CTF), and it is directly related to the properties of the exit pupil. Each point (x,y) in the pupil corresponds to a single spatial frequency:

$$\xi = \frac{x}{\lambda d_i} \qquad\qquad \eta = \frac{y}{\lambda d_i}$$

The pupil of the optical system can then be thought of as a filtering plane where the transfer function of the system can be modified by controlling the amplitude and/or phase of the transmission function within the pupil.

Because of physical limitations on pupil size, all optical systems are inherently LPFs. The highest spatial frequency that can pass a system is dictated by the optical power and the aperture size, and is

$$\rho_{\max} = \frac{D}{\lambda d_i} = \frac{1}{\lambda f/\#_w}$$

The working F-number of the system is denoted as $f/\#_w$, whereas D is twice the maximum radial extent of the pupil. The properties of the impulse response, such as the amount
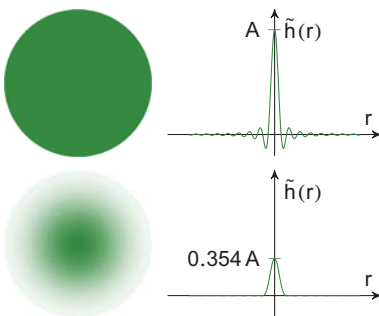
of ringing, can be controlled by shaping or **apodizing** the pupil. This is akin to both the windowing that was applied to sampled sequences to control ringing and to the Gibbs phenomenon seen with the DFT.

Optical Transfer Function (OTF)

When the illumination is coherent, the complex fields emanating from each point in the object interfere:

$$I_i = \left| \tilde{h} * u_g \right|^2$$

When the illumination is incoherent, the fields from two different locations on the object do not interfere, and instead the irradiances add linearly. In this case,

$$I_i = \left| \tilde{h} \right|^2 * I_g$$

with $I_g(x,y) = |u_g(x,y)|^2$. The quantity $|\tilde{h}(x,y)|^2$ is known as the **point spread function** (PSF).

Because the system is now LSI for the irradiance, a spatial frequency analysis is allowed. The normalized Fourier transform of the PSF is the **optical transfer function** (OTF):

$$\mathcal{H}(\xi, \eta) = \frac{H(\xi, \eta) * H^*(-\xi, -\eta)}{H(\xi, \eta) * H^*(-\xi, -\eta)\big|_{\substack{\xi=0 \\ \eta=0}}} = \frac{\gamma_H(\xi, \eta)}{\gamma_H(0, 0)}$$

- $\mathcal{H}(0,0) = 1$. The change in the average value of the scene irradiance is handled separately through the **étendue** of the system.
- For real systems with finite apertures, $\mathcal{H}(\xi, \eta) < 1$ for $\xi, \eta \neq 0$ from the Cauchy–Schwarz inequality.
- The **modulation transfer function** gives the change in sinusoidal modulation $\text{MTF}(\xi, \eta) = \left| \mathcal{H}(\xi, \eta) \right|$.
- For a diffraction-limited system with a circular pupil $\mathcal{H}(\rho) = \gamma_{\text{cyl}}\left(\rho \lambda f/\#, 1\right)$.

Input   OTF   Filtered
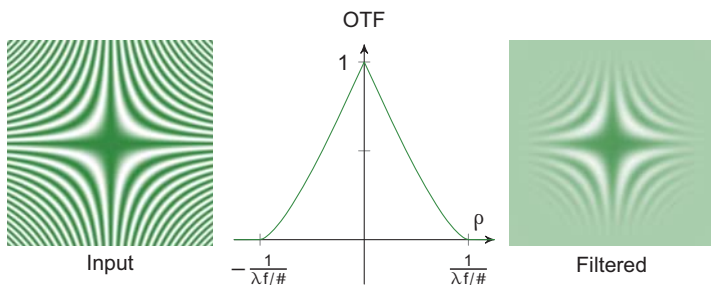
Aberrated Systems

A diffraction-limited system converts spherical waves incident on the entrance pupil into spherical waves leaving the exit pupil converging on (or diverging from) the plane of focus. Real optical systems have **aberrations**, which are deviations from these perfect spherical wavefronts. The deviation $W(x,y)$ is measured in wavelengths as a function of position in the exit pupil plane.
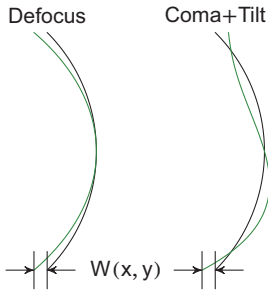
The aberrations can be treated as phase perturbations in the pupil plane, and the CTF becomes

$$H\left(\xi,\eta\right) = P\left(-\lambda d_i \xi, -\lambda d_i \eta\right) e^{i2\pi W\left(-\lambda d_i \xi, -\lambda d_i \eta\right)}$$

With this modification both the CTF and OTF become **complex**, leading to both amplitude and phase changes in the various spatial frequencies. The MTF of an aberrated system is always lower than the MTF of a diffraction-limited system as a consequence of the Cauchy–Schwarz inequality:

$$\mathrm{MTF}_W\left(\xi,\eta\right) \leqslant \mathrm{MTF}_{DL}\left(\xi,\eta\right)$$

The phase perturbations can lead to **contrast nulls**, where the MTF goes to zero and **contrast reversals**, where the OTF passes through a null and experiences a 180° phase change. An example of such location is highlighted in the following figure.

OTF

Diff. Limited $-\dfrac{1}{\lambda f/\#}$ $\dfrac{1}{\lambda f/\#}$ 1 λ Defocus
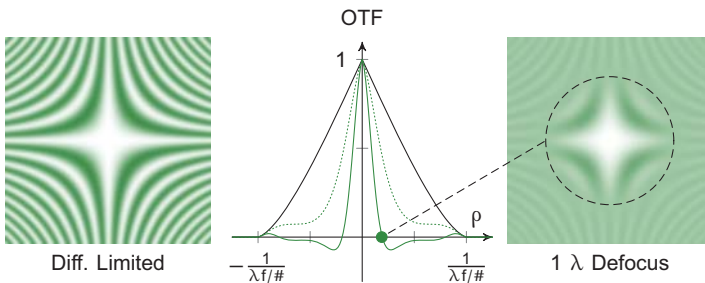
Comparisons of Coherent and Incoherent Output

An optical system has different responses under coherent and incoherent illumination, but these responses are related through the properties of the autocorrelation function. Consider the following two objects:

$$u_1(x,y) = \cos(2\pi\xi_0 x) \qquad u_2(x,y) = \left|\cos(2\pi\xi_0 x)\right|$$

Although the fields are different, the objects have identical ideal irradiance distributions:

$$\left|u_g\right|^2 = \cos^2(2\pi\xi_0 x) = \frac{1}{2}\left[1 + \cos(4\pi\xi_0 x)\right]$$

$$\mathcal{F}\left\{\left|u_g\right|^2\right\} = \frac{1}{2}\left\{\delta(\xi) + \frac{1}{2}\left[\delta(\xi - 2\xi_0) + \delta(\xi + 2\xi_0)\right]\right\}$$

A simple 1D CTF is assumed:

$$H(\xi) = \mathrm{rect}\left(\lambda f/\#\xi\right) \qquad \mathcal{H}(\xi) = \mathrm{tri}\left(\lambda f/\#\xi\right)$$

	Coherent	Incoherent
x	$\left|u_g * \tilde{h}\right|^2$	$\left|\tilde{h}\right|^2 * \left|u_g\right|^2$
ξ	$(U_g H) \star (U_g H)$	$(U_g \star U_g)(H \star H)$
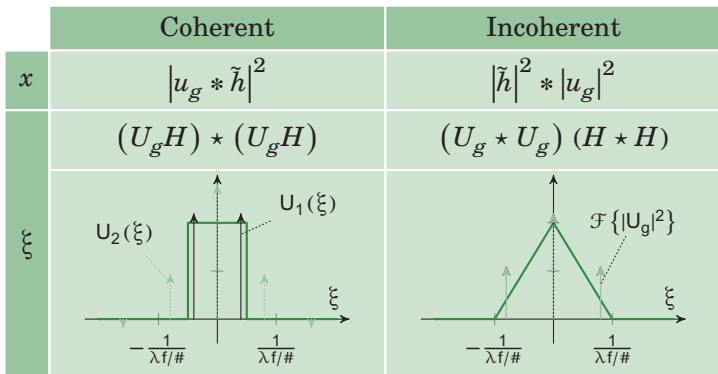
When the spatial frequency of the object is chosen, as shown in the figure, the images under coherent illumination are different. Object #1 is passed by the system unaltered. Object #2 has only the zero-frequency term that is passed by the optical filter. In contrast, both objects produce identical images under incoherent illumination, with some loss of modulation contrast, as indicated by the lower MTF for the oscillating terms in the Fourier series.

Two-Point Resolution with Coherent Light

When the light coming from two point sources is incoherent, the Rayleigh diffraction criterion can be used to establish a minimum separation at which the points are resolvable. However, when the light coming from the two points is coherent, greater care must be taken.

Assume that the geometric optics field image is

$$u_g(x,y) = \delta\left(x - \frac{x_o}{2}\right) + e^{i\phi_0}\delta\left(x + \frac{x_o}{2}\right)$$

The two points have equal amplitude but a relative phase difference of ϕ_0. The resulting irradiance distribution predicted by physical optics is

$$u_i(x,y) = \left| h\left(x - \frac{x_o}{2}\right) + e^{i\phi_0}h\left(x + \frac{x_o}{2}\right)\right|^2$$

$\phi_0 = 180°$

$\phi_0 = 90°$

$\phi_0 = 0°$
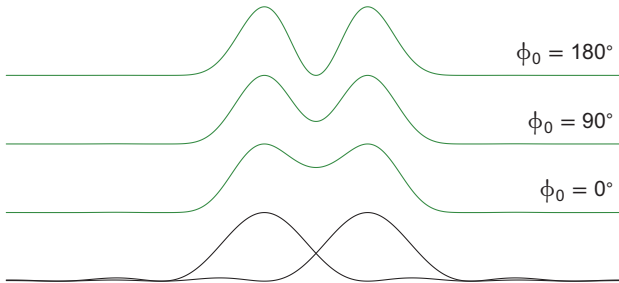
When the two sources are 90° out-of-phase, the resulting irradiance distribution is identical to the incoherent case. When $\phi_0 = 0°$, the constructive interference between the two images makes resolution more difficult. When $\phi_0 = 180°$, the destructive interference between the two images at the center point actually makes it easier to resolve them.

Two sources that are 90° out-of-phase are often termed **uncorrelated** or **orthogonal**. The net effect is that the result is the same as if the sources were incoherent. An identical result would also be obtained for two sources that are **orthogonally polarized** and thus cannot interfere.

Roughness and Scattered Light

Optical components are polished to a desired shape within a certain degree of tolerance. Deviations from the ideal figure due to **surface roughness** can be treated in much the same way as aberrations. Unlike aberrations, which generally have an analytic form, the phase variations across the wavefront due to surface roughness are best treated using the tools of random processes discussed earlier. The result of surface roughness is **scattered light**, which is a broadening of the system PSF.

The additional phase in the pupil of the system can be captured in a hypothetical **roughness plate** with transmission

$$T_r(x,y) = e^{ik\rho(x,y)}$$
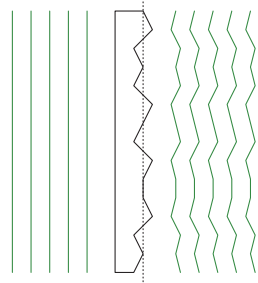
The function $\rho(x,y)$ gives the additional optical path due to roughness with a resulting impulse response

$$\tilde{h}'(x,y) = \tilde{h}(x,y) * t(x,y)$$

The function $t(x,y) = \mathcal{F}^{-1}\left\{T_r\left(-\lambda d_i \xi, -\lambda d_i \eta\right)\right\}$ plays the role of a random signal being filtered by the deterministic impulse response. When the roughness is small compared to a wavelength, the scatter plate can be modeled as a weak phase filter:

$$t(x,y) \approx \mathcal{F}^{-1}\left\{1 + i2\pi \frac{\rho\left(-\lambda d_i \xi, -\lambda d_i \eta\right)}{\lambda}\right\}$$

$$= \delta(x,y) + i\frac{2\pi}{d_i}\mathrm{P}\left(\frac{x}{\lambda d_i}, \frac{y}{\lambda d_i}\right)$$

The second term on the right side gives the scattering blur. Because $\rho(x,y)$ is a random function, the blurring effect is best described in terms of the PSD of the roughness:

$$E\left[|t_s(x,y)|^2\right] = \left(\frac{2\pi}{d_i}\right)^2 \mathrm{PSD}_\rho\left(\frac{x}{\lambda d_i}, \frac{y}{\lambda d_i}\right)$$

Fourier Transform Spectroscopy (FTS)

$u_1(t)$

$\frac{L}{2}$

$\frac{L+\ell}{2}$

$u(t)$

$u_2(t)$

$u_{tot} = u_1 + u_2$
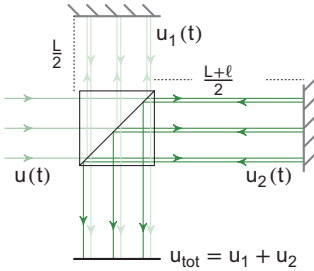
Fourier transform spectroscopy (FTS) exploits the Wiener–Khinchin theorem to measure the spectrum of optical radiation in an interferometer. The path length in arm 2 of the interferometer is ℓ longer than that in arm 1. The fields at the detector are

$$u_{\text{tot}} = \frac{1}{4}\left(u(t) + u(t + \ell/c)\right)$$

Because the underlying field is $u(z - ct)$ it can be equally well be written in terms of space. If ℓ is scanned through a range of values, and the irradiance is measured by averaging the signal in time, the resulting **interferogram** is

$$I(\ell) = \left\langle \frac{|u(t)|^2}{16} \right\rangle_t + \left\langle \frac{\left|u\left(t + \frac{\ell}{c}\right)\right|^2}{16} \right\rangle_t + \Re\left[\frac{\left\langle u(t)u^*\left(t + \frac{\ell}{c}\right)\right\rangle_t}{8}\right]$$

The time averages can be replaced by expectation values

$$I(\tau) = \frac{1}{2}\left(I_0 + \gamma_u(\tau)\right)$$

where $\tau = \ell/c$. Note that because $u(t)$ is purely real, $\gamma_u(\tau)$ is real and even. According to the Wiener–Khinchin theorem, the PSD is

$$\mathcal{F}\{I(\tau)\}_\tau = \frac{1}{2}\left[\delta(\nu) + I(\nu)\right]$$

Interferogram

ℓ

$-4\lambda_0$ $4\lambda_0$

Spectrum or PSD

ν

$\frac{c}{\lambda_0}$
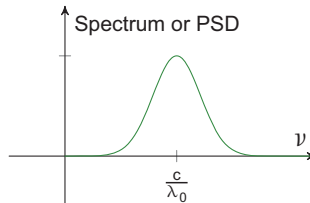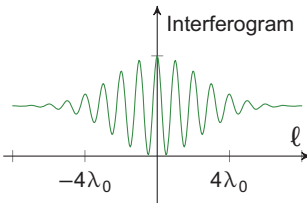
Multiplexing

Multiplexing gives the ability to carry multiple signals over the same physical communication medium. There are several common multiplexing protocols in wide use:

- **Time division multiplexing** (TDM)
  Users get access to the medium at different times.
- **Frequency division multiplexing** (FDM)
  The EM spectrum is divided into **bands** of **channels** that are orthogonal to each other in a Fourier sense.
- **Code division multiplexing** (CDM)
  Multiple pieces of information are sent over the same communications medium, but they are made orthogonal by modulating non-Fourier **codes**, which are often pseudo-random sequences.
- **Wavelength division multiplexing** (WDM)
  Used in fiber optics. Users are assigned to different wavelengths, all of which are sent through the same fiber. The wavelengths are separated optically at the receiver in order to demultiplex the signals.
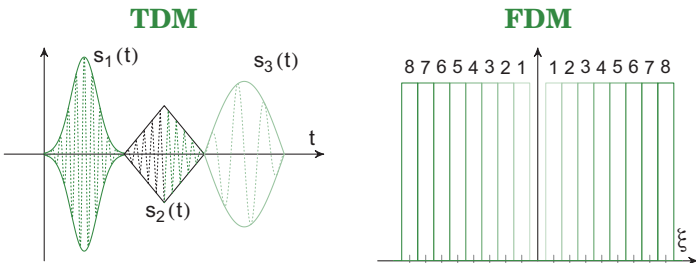
Consider a set of N signals $s_1(x), \ldots, s_N(x)$, each assigned to a carrier frequency ξ_n. **Amplitude-modulated** signals are all present simultaneously, but the information can be recovered so long as the channels do not overlap:

$$f(x) = \sum_{n=1}^{N} f_n(x), \quad f_n(x) = s_n(x) \cos(2\pi \xi_n x)$$

$$F(\xi) = \frac{1}{2} \sum_{n=1}^{N} \left[S_n(\xi - \xi_n) + S_n(\xi + \xi_n) \right]$$

Sampled Color Imaging Systems

The goal of imaging systems is to record information pertaining to the scene in order to reproduce or display a recognizable version of it. The scene can be denoted as an intensity distribution $I(x, y, \sigma, t)$.

Present state-of-the-art detectors cannot detect such information with infinite resolution. Instead of performing an ideal sampling operation with $\delta(x, y, \sigma, t)$, a non-ideal sampling operation is performed. It can be described with a pulse function

$$p\left(\frac{x}{x_p}, \frac{y}{y_p}, \frac{\sigma}{\sigma_p}, \frac{t}{t_p}\right)$$

where x_p, y_p, σ_p, and t_p represent the integration widths in each domain and set a lower limit on the corresponding sampling intervals in each domain. A sparse sampling rate will cause a variety of artifacts in the recorded representation. A long t_p will lead to motion blur, while larger x_p and y_p will lead to pixelation.

Temporal resolution is generally of less interest because t_p is independent and can potentially be reduced without infringing on other domains. Measuring (x, y) and σ concurrently, however, invariably introduces a trade-off. A typical 2D imaging configuration measures color information by placing a grid of alternating color filters on the detector. Thus, the dataset is inherently partial and requires careful treatment to achieve the goal of accurate representation.

Finally, in order to maximize the overall performance of the system, it is important to match the OTF of the optics with the pixel-pitch/color-sampling of the detector. The displayed image can be written as

$$I_D(\xi, \eta) = \left[O(\xi, \eta)\mathcal{H}(\xi, \eta)P(\xi, \eta) * \text{comb}(x_s\xi, y_s\eta)\right]D(\xi, \eta)$$

where O is the object's Fourier transform, $\mathcal{H}$ is the optical transfer function, P is the pixel transfer function, and D is the display transfer function. Focusing on improving only one of the components of the overall chain is unlikely to significantly increase system performance. Every aspect of the system needs balanced improvement in order to obtain optimal results.

RGB Detector and Display Arrays

Sensor Arrangements
The Bayer pattern is 2×2 super-pixel of color-selecting filters that contains two green pixels, one red pixel, and one blue pixel. There are more green-selecting pixels because that part of the spectrum has the dominant effect in human vision. To store a set of RGB values for each pixel, various demosaicing schemes were developed. Alternatively, this pattern can be described as a channeled system by leveraging the implied Fourier structure of the information splitting.

R	G	R	G
G	B	G	B
R	G	R	G
G	B	G	B
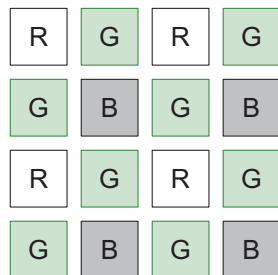
Display Arrangements
RGB stripe is a 1×3 arrangement of sub-pixels used in LCD displays to map the stored RGB information directly to each sub-pixel. This mapping is the main advantage of the arrangement because it places fewer requirements on the software developer to achieve proper display of information.

R	G	B	R	G	B
R	G	B	R	G	B
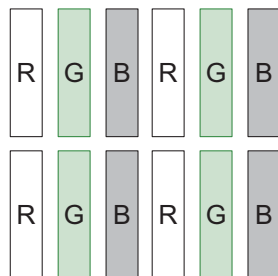
Pentile RGBG is a 1×4 arrangement of sub-pixels that is used primarily in some AMOLED displays. It uses the same objective of leveraging human color sensitivity preference as the Bayer pattern in sensors. There are many sub-pixel arrangements with an unequal distribution of color sensitivity.

R	G	B	G
B	G	R	G
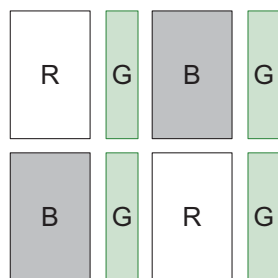
Depending on the pixel arrangement, pixel count might not convey the system performance fully. Instead, a more-elaborate analysis may be required.

Channeled Spectropolarimetry

Polarimetry involves the measurement of the vector nature of the optical polarization signature. Because optical detectors do not generally respond to polarization, polarimeters operate by modulating the intensity of the light in a polarization-dependent fashion. An important class of polarimeters are the **channeled polarimeters** that accomplish this by creating FDM channels in space, time, wavenumber, angle, or some other modulation dimension.

Consider a beam of light $\mathbf{E} = u_x\hat{x} + u_y\hat{y}$. The Stokes parameters describing its polarization state are

$$\mathbf{S}(t) = \begin{bmatrix} s_0(t) \\ s_1(t) \\ s_2(t) \\ s_3(t) \end{bmatrix} = \begin{bmatrix} \left\langle |u_x(t)|^2 \right\rangle + \left\langle |u_y(t)|^2 \right\rangle \\ \left\langle |u_x(t)|^2 \right\rangle - \left\langle |u_y(t)|^2 \right\rangle \\ 2\Re \left\langle u_x(t)u_y^*(t) \right\rangle \\ -2\Im \left\langle u_x(t)u_y^*(t) \right\rangle \end{bmatrix}$$

This beam is analyzed by a linear polarizer rotating at a constant angular velocity $\theta = 2\pi\nu_0 t$. This analyzer produces a time-varying irradiance

$$I(t) = \frac{1}{2} \begin{bmatrix} 1 & \cos 2\theta & \sin 2\theta & 0 \end{bmatrix}^T \cdot \mathbf{S}(t)$$

$$= \frac{1}{2} \left[s_0(t) + s_1(t)\cos(4\pi\nu_0 t) + s_2(t)\sin(4\pi\nu_0 t) \right]$$

The resulting irradiance is a **multiplexed signal** with the s_0 information carried in a channel centered at $\nu = 0$, the s_1 information in the real part of a channel centered at $\nu = 2\nu_0$, and the s_2 information in the imaginary part of a
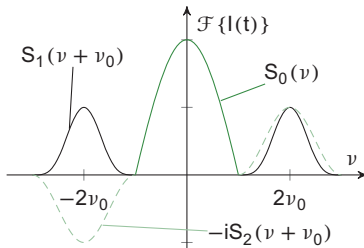
channel centered at $\nu = 2\nu_0$. More-complicated strategies have been developed that modulate in time, space, wavenumber, angle of incidence, and combinations of multiple independent variables.

Optical Signal Processing

Before the emergence of digital computing, analog computers were an essential part of science and engineering. One important example of an analog computer is the **optical signal processor**. This technique relies on the Fourier transforming properties of lenses that makes it possible to place a filter at the intermediate plane of a $4f$-system.

Optical signal processing has the advantage of nearly instantaneous performance because the information travels at the speed of light. It was used in pattern matching/matched filtering applications, such as **optical character recognition** and satellite photographic analysis. Optical signal processing also found wide use for processing radiographs and other types of medical images. Filters could be changed in the focal plane to test multiple templates. The maturation of liquid crystal technology makes it possible to dynamically change the properties of the filter in the intermediate plane.
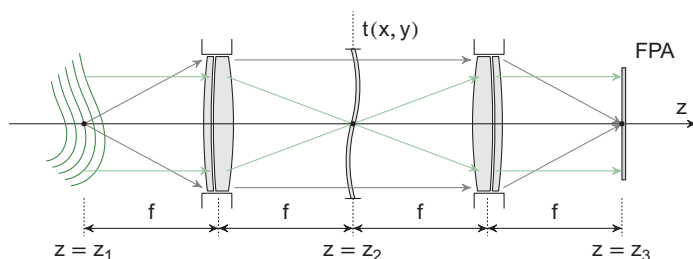
Optical signal processing has fallen out of favor for general-purpose applications because of the accessibility and extremely low cost of computational power. Furthermore, optical processing is limited in the types of filters and number of cases that can be implemented, at least relative to computer-based methods. However, for high-performance tasks, such as on-board processing for military applications, a specialized optical signal processor can outperform a digital solution for specific needs.

Green's Functions

A **Green's function** is an **impulse response** of a system described by a linear partial differential equation. Examples in electromagnetics are the free-space scalar Green's functions for the time-harmonic Helmholtz equation for the z-component of the magnetic vector potential:

$$\nabla^2 A_z + k^2 A_z = -\mu J_z$$

The 1D problem can be rewritten in a simpler form:

$$\frac{dG}{dx^2} + k^2 G = -\mu \delta(x - x_0)$$
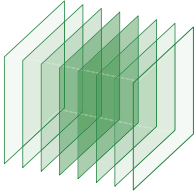
The solution is plane waves propagating away from the location of the source current at x_0:

$$G(x) = \begin{cases} A_+ e^{ikx} & x > x_0 \\ A_+ e^{-ikx} & x < x_0 \end{cases}$$

The 2D problem is best treated in cylindrical coordinates:

$$\frac{1}{r}\frac{\partial G}{\partial r}\left(r\frac{\partial G}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 G}{\partial \theta^2} + k^2 G = -\mu \delta(\mathbf{r} - \mathbf{r}_0)$$
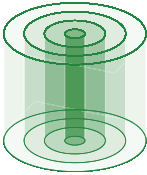
The solution is zeroth-order Hankel functions of the first kind propagating away from the source at $\mathbf{r}_0 = \{x_0, y_0\}$:

$$G(\mathbf{r}) = A H_0^{(1)}(\mathbf{r} - \mathbf{r}_0)$$

The 3D problem is best treated in spherical coordinates:

$$\frac{1}{r^2}\left[\frac{\partial}{\partial r}\left(r\frac{\partial G}{\partial r}\right) + \frac{\partial^2 G}{\sin^2\phi\,\partial\theta^2} + \frac{1}{\sin\phi}\frac{\partial}{\partial\phi}\left(\sin\phi\frac{\partial G}{\partial\phi}\right)\right] + k^2 G = -\mu J$$
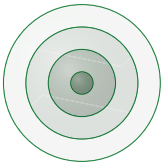
The solution is spherical waves outwardly propagating away from the source at $\mathbf{r}_0 = \{x_0, y_0, z_0\}$:

$$G(\mathbf{r}) = \frac{e^{ik|\mathbf{r}-\mathbf{r}_0|}}{|\mathbf{r}-\mathbf{r}_0|}$$

Moment Method

One of the tools of linear systems is the ability to decompose a complicated function into a weighted sum of fundamental basis functions. When linearity holds, a problem can be solved for each of the basis functions, and then the total answer is obtained by summing the individual solutions:

$$\mathscr{S}\left\{\sum_{n=1}^{N} a_n \psi_n\right\} = \sum_{n=1}^{N} a_n \mathscr{S}\left\{\psi_n\right\}$$

The **moment method** uses this principle to solve integral equations. A common example involves electromagnetic scattering problems. The scattered electric field can be written in terms of the unknown currents $\mathbf{J}$:

$$\mathbf{E}_s(\mathbf{r}) = \int_S \overleftrightarrow{\mathbf{G}}(\mathbf{r};\mathbf{r}')\,\mathbf{J}(\mathbf{r}')d\mathbf{r}'$$

The domain of integration S includes the surface of the scatterer, and the Green's function $\overleftrightarrow{\mathbf{G}}$ is a dyadic function of the source location $\mathbf{r}'$ and the observation location $\mathbf{r}$. The estimate of the total field is the sum of the known incident and the unknown scattered fields replacing $\mathbf{J}$ with a superposition of basis functions:

$$\hat{\mathbf{E}}_{\text{tot}} = \mathbf{E}_{\text{inc}} + \hat{\mathbf{E}}_s = \mathbf{E}_{\text{inc}}(r) + \sum_{n=1}^{N} a_n \int_{S_n} \overleftrightarrow{\mathbf{G}}(\mathbf{r};\mathbf{r}')\,\boldsymbol{\psi}_n(\mathbf{r}')d\mathbf{r}'$$

Boundary conditions are used to convert the integral equation into a system of linear equations for the coefficients $\{a_n\}$. As an example, when the scatterer is a perfect conductor, the total tangential field $\hat{\mathbf{n}} \times \mathbf{E}_{\text{tot}} = 0$, and the equation can be evaluated as a set of M observation points:

$$\begin{bmatrix} Z_{11} & \cdots & Z_{1N} \\ \vdots & \ddots & \vdots \\ Z_{M1} & \cdots & Z_{MN} \end{bmatrix} \begin{bmatrix} a_n \\ \vdots \\ a_N \end{bmatrix} = -\begin{bmatrix} \hat{\mathbf{n}} \times \mathbf{E}_{\text{inc}}(\mathbf{r}_1) \\ \vdots \\ \hat{\mathbf{n}} \times \mathbf{E}_{\text{inc}}(\mathbf{r}_M) \end{bmatrix}$$

The computational burden lies in developing the Green's function $\overleftrightarrow{\mathbf{G}}(\mathbf{r};\mathbf{r}')$ and computing the matrix

$$Z_{mn} = \hat{\mathbf{n}} \times \int_{S_n} \overleftrightarrow{\mathbf{G}}(\mathbf{r}_m;\mathbf{r}')\,\boldsymbol{\psi}_n(\mathbf{r}')d\mathbf{r}'$$

The basis functions can be pulse functions, polynomials, Fourier functions, or any other appropriate function set.

Array Apertures

Arrays are important in applications involving radio frequencies and microwaves, and are becoming increasingly important in optical applications. An array can have a real aperture that is sampled by a number of array elements, or it can have a **synthetic aperture**. In such an array, the sampling is carried out by a single element that is scanned across the aperture.
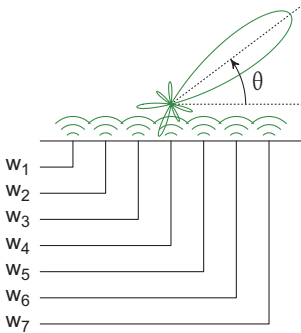
Consider the fields in the Fraunhofer zone of a 1D array of identical point-source elements. The elements are assumed to be periodically spaced every d units over the aperture. Non-periodic sampling is also acceptable, but is more complicated. The amplitude and phase of the nth element are given by the complex weight w_n. The fields observed at angle θ are given by the DSFT of the weights with $\xi = (\cos\theta)/\lambda$:

$$u(\theta) = \sum_{n=1}^{N} w_n e^{i2\pi n d \frac{\cos\theta}{\lambda}}$$

The choice of the weighting profile can produce beams of different shapes, and the ability to dynamically control the weights allows the array to be steered in different directions. **Reciprocity** dictates that the array operates in the same manner on transmission and reception, so this is equally applicable to a receiver.

Synthetic arrays do not fill the entire array with elements. Instead, they use a smaller set of elements – perhaps just one in some applications – and scan those elements across the aperture in time. Synthetic aperture arrays have become important in astronomy with the emergence of instruments such as the Large Binocular Telescope in optics and the **Very Large Array** (VLA) in radio.

Array Apertures (cont.)

In a synthetic aperture system, the entire array aperture is not filled simultaneously. Such synthetic aperture systems are generally used in the receive mode, and a good example is an interferometric astronomical telescope.

Consider two elements with identical apertures whose centers are separated by $2\mathbf{r}_0 = 2(x_0\hat{\mathbf{x}} + y_0\hat{\mathbf{y}})$:

$$P(x,y) = \text{cyl}\left(\frac{|\mathbf{r} - \mathbf{r}_0|}{D}\right) + \text{cyl}\left(\frac{|\mathbf{r} + \mathbf{r}_0|}{D}\right)$$

The resulting OTF for these two elements is

$$\mathcal{H}_{\mathbf{r}_0}(\xi, \eta) = \gamma_{\text{cyl}}\left(\lambda f/\#_w \left|\boldsymbol{\rho}\right|; 1\right) + \frac{1}{2}\left[\gamma_{\text{cyl}}\left(\lambda f/\#_w \left|\boldsymbol{\rho} - \boldsymbol{\rho}_0\right|; 1\right)\right.$$
$$\left. + \gamma_{\text{cyl}}\left(\lambda f/\#_w \left|\boldsymbol{\rho} + \boldsymbol{\rho}_0\right|; 1\right)\right]$$

The offset vector in the (ξ, η)-plane is

$$\boldsymbol{\rho}_0 = \frac{x_0}{\lambda f/\#_w}\hat{\boldsymbol{\xi}} + \frac{y_0}{\lambda f/\#_w}\hat{\boldsymbol{\eta}}$$
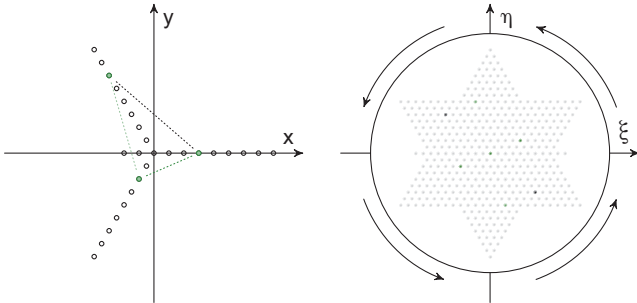
A single pair of elements samples the Fourier transform of the object at the spatial frequencies given by $\pm\boldsymbol{\rho}_0$ with sampling function $\gamma_{\text{cyl}}\left(\lambda f/\#_w \left|\boldsymbol{\rho}\right|; 1\right)$. If the spacing and orientation between the two apertures is scanned through a set of values in the spatial frequency plane, then the Fourier transform of the object can be built point-by-point, and the resulting image can be estimated by the inverse Fourier transform. In an array like the VLA, the distances between the antennas can be altered, and the Earth's motion in its orbit helps scan the spatial frequency plane.

Crystal Lattices and Reciprocal Lattices

A crystalline structure is described by a unit cell and a **Bravais lattice**, which is a collection of vectors that gives the periodicity of the crystal

$$\mathbf{R} = n\mathbf{a} + m\mathbf{b} + p\mathbf{c}$$

The vectors **a**, **b**, and **c** are the **primitive vectors** that span the lattice, and n, m, and p are integers. The Bravais lattice describes the spatial properties of the periodic structure.
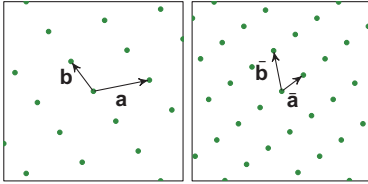
The position **r** in space and the wavevector **σ** of plane waves propagating in that space are Fourier-dual variables. The plane wave spectrum gives an example of how the Fourier transform of a spatial field distribution produces a spectrum of plane waves propagating in different directions.

Consider a 2D example for consistency with earlier notation. However, the concept easily generalizes to 3D. A lattice of point scatterers is modeled as

$$f(x,y) = \text{comb}\left(a_x x + a_y y\right) \text{comb}\left(b_x x + b_y y\right)$$

The corresponding primitive vectors of the lattice are

$$\mathbf{a} = a_x\hat{\mathbf{x}} + a_y\hat{\mathbf{y}} \qquad\qquad \mathbf{b} = b_x\hat{\mathbf{x}} + b_y\hat{\mathbf{y}}$$

The Fourier transform of $f(x,y)$ describes the **reciprocal lattice**, which is the collection of all points that satisfy

$$e^{i2\pi\boldsymbol{\sigma}\cdot\mathbf{R}} = 1$$

for all values of n, m, and p. In the 2D example above:

$$F(\xi,\eta) = \frac{1}{|D|}\text{comb}\left(\frac{b_y}{D}\xi - \frac{b_x}{D}\eta, -\frac{a_y}{D}\xi + \frac{a_x}{D}\eta\right)$$

The primitive vectors of the reciprocal lattice are

$$\bar{\mathbf{a}} = \frac{b_y\hat{\boldsymbol{\xi}} - b_x\hat{\boldsymbol{\eta}}}{D} \qquad\qquad \bar{\mathbf{b}} = \frac{-a_y\hat{\boldsymbol{\xi}} + a_x\hat{\boldsymbol{\eta}}}{D}$$

with $D = a_x b_y - b_x a_y$, $\mathbf{a} \cdot \bar{\mathbf{b}} = \bar{\mathbf{a}} \cdot \mathbf{b} = 0$, and $\mathbf{a} \cdot \bar{\mathbf{a}} = \bar{\mathbf{b}} \cdot \mathbf{b} = 1$.

Fourier Transform Tables

$$e^{\pm i 2\pi \xi_o x} \longleftrightarrow \delta(\xi \mp \xi_o)$$

$$\delta(x \pm x_o) \longleftrightarrow e^{\pm i 2\pi x_o \xi}$$

$$\cos(2\pi \xi_o x) \longleftrightarrow \tfrac{1}{2}\left[\delta(\xi - \xi_o) + \delta(\xi + \xi_o)\right]$$

$$\tfrac{1}{2}\left[\delta(x - x_o) + \delta(x + x_o)\right] \longleftrightarrow \cos(2\pi x_o \xi)$$

$$\sin(2\pi \xi_o x) \longleftrightarrow \tfrac{1}{2i}\left[\delta(\xi - \xi_o) - \delta(\xi + \xi_o)\right]$$

$$\tfrac{1}{2i}\left[\delta(x - x_o) - \delta(x + x_o)\right] \longleftrightarrow \sin(2\pi x_o \xi)$$

$$\mathrm{rect}(x) \longleftrightarrow \mathrm{sinc}(\xi)$$

$$\mathrm{sinc}(x) \longleftrightarrow \mathrm{rect}(\xi)$$

$$\mathrm{tri}(x) \longleftrightarrow \mathrm{sinc}^2(\xi)$$

$$\mathrm{sinc}^2(x) \longleftrightarrow \mathrm{tri}(\xi)$$

$$\mathrm{sgn}(x) \longleftrightarrow \frac{1}{i\pi\xi}$$

$$\frac{1}{i\pi x} \longleftrightarrow -\mathrm{sgn}(\xi)$$

$$\mathrm{step}(x) \longleftrightarrow \tfrac{1}{2}\delta(\xi) + \frac{1}{i2\pi\xi}$$

$$\tfrac{1}{2}\delta(x) - \frac{1}{i2\pi x} \longleftrightarrow \mathrm{step}(\xi)$$

$$\mathrm{ramp}(x) \longleftrightarrow \frac{1}{4\pi^2}\left[i\pi\delta^{(1)}(\xi) - \frac{1}{\xi^2}\right]$$

$$\frac{1}{4\pi^2}\left[\frac{1}{x^2} + i\pi\delta^{(1)}(x)\right] \longleftrightarrow \mathrm{ramp}(\xi)$$

$$e^{-|x|} \longleftrightarrow \frac{2}{1+(2\pi\xi)^2}$$

$$\frac{2}{1+(2\pi x)^2} \longleftrightarrow e^{-|\xi|}$$

$$e^{-x}\mathrm{step}(x) \longleftrightarrow \frac{1}{1+i2\pi\xi}$$

$$\frac{1}{1-i2\pi x} \longleftrightarrow e^{-\xi}\mathrm{step}(\xi)$$

$$x^k \longleftrightarrow \left(\frac{-1}{i2\pi}\right)^k \delta^{(k)}(\xi)$$

$$\left(\frac{1}{i2\pi}\right)^k \delta^{(k)}(x) \longleftrightarrow \xi^k$$

$$\mathrm{comb}(x) \longleftrightarrow \mathrm{comb}(\xi)$$

$$\mathrm{Gaus}(x) \longleftrightarrow \mathrm{Gaus}(\xi)$$

$$\mathrm{sech}(\pi x) \longleftrightarrow \mathrm{sech}(\pi\xi)$$

Field Guide to Linear Systems in Optics

Fourier Transform Tables (cont.)

$$|x|^{-1/2} \longleftrightarrow |\xi|^{-1/2}$$

$$\cos\left(\pi(x^2 - \tfrac{1}{8})\right) \longleftrightarrow \cos\left(\pi(\xi^2 - \tfrac{1}{8})\right)$$

$$\sin\left(\pi(x^2 - \tfrac{1}{8})\right) \longleftrightarrow -\sin\left(\pi(\xi^2 - \tfrac{1}{8})\right)$$

$$\cos(\pi x^2) \longleftrightarrow \cos\left(\pi(\xi^2 - \tfrac{1}{4})\right)$$

$$\sin(\pi x^2) \longleftrightarrow -\sin\left(\pi(\xi^2 - \tfrac{1}{4})\right)$$

$$e^{\pm i\pi(x^2 - 1/8)} \longleftrightarrow e^{\mp i\pi(\xi^2 - 1/8)}$$

$$e^{\pm i\pi x^2} \longleftrightarrow e^{\pm i\pi/4} e^{\mp i\pi\xi^2}$$

$$\exp\left[\frac{-\pi x^2}{a + ic}\right], \genfrac{}{}{0pt}{}{a \geq 0}{a^2 + c^2 < \infty} \longleftrightarrow \sqrt{a + ic}\, \exp\left[-\pi(a + ic)\xi^2\right]$$

$$f(\pm x) \longleftrightarrow F(\pm\xi)$$

$$f^*(\pm x) \longleftrightarrow F^*(\mp\xi)$$

$$F(\pm x) \longleftrightarrow f(\mp\xi)$$

$$F^*(\pm x) \longleftrightarrow f^*(\pm\xi)$$

$$f(\tfrac{x}{b}) \longleftrightarrow |b| F(b\xi)$$

$$|d| f(dx) \longleftrightarrow F(\tfrac{\xi}{d})$$

$$f(x \pm x_o) \longleftrightarrow e^{\pm i2\pi\xi x_o} F(\xi)$$

$$e^{\pm i2\pi\xi_o x} f(x) \longleftrightarrow F(\xi \mp \xi_o)$$

$$f^{(k)}(x) \longleftrightarrow (i2\pi\xi)^k F(\xi)$$

$$(-i2\pi x)^k f(x) \longleftrightarrow F^{(k)}(\xi) \;\rightarrow\; m_k = \frac{F^{(k)}(0)}{(-i2\pi)^k}$$

$$\int_{-\infty}^{x} f(x')\, dx' \longleftrightarrow \frac{1}{i2\pi\xi} F(\xi) + \frac{F(0)}{2}\delta(\xi)$$

$$\frac{-1}{i2\pi x} f(x) + \frac{f(0)}{2}\delta(x) \longleftrightarrow \int_{-\infty}^{\xi} f(\xi')\, d\xi'$$

$$A_1 f(x) + A_2 h(x) \longleftrightarrow A_1 F(\xi) + A_2 H(\xi)$$

$$f(x) * h(x) \longleftrightarrow F(\xi) H(\xi)$$

$$f(x) h(x) \longleftrightarrow F(\xi) * H(\xi)$$

$$\int_{-\infty}^{\infty} |f(x)|^2\, dx \Longleftrightarrow \int_{-\infty}^{\infty} |F(\xi)|^2\, d\xi$$

Fourier Transform Tables (cont.)

$$f(x \pm x_0, y \pm y_0) \longleftrightarrow e^{i2\pi(\pm x_0\xi \pm y_0\eta)} F(\xi, \eta)$$

$$e^{i2\pi(\pm \xi_0 x \pm \eta_0 y)} f(x,y) \longleftrightarrow F\left(\xi \mp \xi_0, \eta \mp \eta_0\right)$$

$$f\left(\frac{x}{b}, \frac{y}{d}\right) \longleftrightarrow |bd| F\left(b\xi, d\eta\right)$$

$$f(ax, cy) \longleftrightarrow \frac{1}{|ac|} F\left(\frac{\xi}{a}, \frac{\eta}{c}\right)$$

$$\frac{d^{(n+m)}}{dx^n dy^m} f(x,y) \longleftrightarrow (2\pi i\xi)^n (2\pi i\eta)^m F(\xi, \eta)$$

$$(2\pi ix)^n (2\pi iy)^m f(x,y) \longleftrightarrow (-1)^{n+m} \frac{d^{(n+m)}}{d\xi^n d\eta^m} F(\xi, \eta)$$

$$\int_{-\infty}^{x} \int_{-\infty}^{\infty} f(x', y') dx' dy' \longleftrightarrow \left[\frac{F(0,0)\delta(\xi)}{2} + \frac{F(\xi, 0)}{i2\pi\xi}\right]\delta(\eta)$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{y} f(x', y') dx' dy' \longleftrightarrow \left[\frac{F(0,0)\delta(\eta)}{2} + \frac{F(0, \eta)}{i2\pi\eta}\right]\delta(\xi)$$

$$\int_{-\infty}^{\infty} f(x, y') dy' \longleftrightarrow F(\xi, 0)\delta(\eta)$$

$$\int_{-\infty}^{\infty} f(x', y) dx' \longleftrightarrow F(0, \eta)\delta(\xi)$$

$$\exp\left[\pm i\frac{r^2}{\lambda z}\right] \longleftrightarrow \pm i\lambda z e^{\mp i\lambda z \rho^2}$$

$$\mathrm{cyl}(r) \longleftrightarrow \frac{\pi}{4}\mathrm{somb}(\rho)$$

$$\gamma_{\mathrm{cyl}}\left(\frac{r}{d}, a\right) \longleftrightarrow \frac{\pi a^2 d^2}{4}\mathrm{somb}(a\rho)\mathrm{somb}(ad\rho)$$

$$e^{-r} \longleftrightarrow 2\pi \left(4\pi\rho^2 + 1\right)^{2/3}$$

$$\cos(\pi r^2) \longleftrightarrow \sin(\pi\rho^2)$$

$$r^{-1} \longleftrightarrow \rho^{-1}$$

Equation Summary

Cartesian-Polar Notation for Complex Numbers:

$$u = x + iy = re^{i\theta} = r\left(\cos\theta + i\sin\theta\right)$$

$$r = \sqrt{x^2 + y^2} \qquad \theta = \arctan\left(y/x\right)$$

Complex Conjugate:

$$u = re^{i\theta} \qquad\qquad u^* = re^{-i\theta}$$

$$u = x + iy \qquad\qquad u^* = x - iy$$

Special Functions:

$$\text{step}(x) = \begin{cases} 1 & x \geqslant 0 \\ 0 & x < 0 \end{cases}$$

$$\text{ramp}(x) = \int_{-\infty}^{x} \text{step}(x')\,dx' = \begin{cases} x & x \geqslant 0 \\ 0 & x < 0 \end{cases}$$

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

$$\text{rect}(x) = \text{step}\left(x + \frac{1}{2}\right) - \text{step}\left(x - \frac{1}{2}\right)$$

$$\text{tri}(x) = \begin{cases} 1 - |x| & |x| < 1 \\ 0 & |x| \geqslant 1 \end{cases}$$

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

$$\text{Gaus}(x) = e^{-\pi x^2}$$

$$\text{cyl}(r) = \text{rect}(r)$$

$$\text{somb}(r) = \frac{2J_1(\pi r)}{\pi r}$$

Impulse Function Properties:

$$\delta\left(\frac{x - x_0}{b}\right) = |b|\delta(x - x_0)$$

$$\int_a^b \delta(x - x_0)\,dx = \begin{cases} 1 & a < x_0 < b \\ 0 & \text{else} \end{cases}$$

$$\int_{-\infty}^{\infty} f(x')\,\delta(x' - x_0)\,dx' = f(x_0)$$

Equation Summary

Impulse Function Properties (cont.):

$$\int_{-\infty}^{\infty} f(x')\delta(x'-x_0)dx' = f(x_0)$$

$$\int_{-\infty}^{\infty} f(x')\delta^{(n)}(x'-x_0)dx' = (-1)^n \left.\frac{d^n}{dx^n}f(x)\right|_{x=x_0}$$

$$\int_{-\infty}^{x} \delta(x'-x_0)\,dx' = \text{step}(x-x_0)$$

$$\delta(x-x_0) = \frac{d}{dx}\text{step}(x-x_0)$$

$$\text{comb}(x) = \sum_{n=-\infty}^{\infty} \delta(x-n) = \sum_{n=-\infty}^{\infty} e^{i2\pi nx}$$

Fourier Series:

$$f_X(x) = \sum_{n=-\infty}^{\infty} c_n e^{i2\pi\xi_0 x}$$

$$\xi_0 = \frac{1}{X}$$

$$c_n = \frac{1}{X}\int_{x=x_0}^{x_0+X} f_X(x')e^{-i2\pi\xi_0 x'}dx'$$

Fourier Transform:

$$\mathcal{F}\{f(x)\} = F(\xi) = \int_{-\infty}^{\infty} f(x)e^{-i2\pi\xi x}dx$$

$$\mathcal{F}^{-1}\{F(\xi)\} = f(x) = \int_{-\infty}^{\infty} F(\xi)e^{+i2\pi\xi x}d\xi$$

Laplace Transform:

$$F(s) = \int_{-\infty}^{\infty} f(x)e^{-sx}dx$$

$$s = \sigma + i\omega$$

Hankel Transform:

$$\mathcal{F}\{f(r)\} = \int_{r=0}^{\infty}\int_{\theta=0}^{2\pi} f(r)e^{-i2\pi r(\xi\cos\theta+\eta\sin\theta)}rdrd\theta$$

Discrete-Space Fourier Transform (DSFT):

$$F\left(e^{i2\pi\upsilon}\right) = \sum_{n=-\infty}^{\infty} f_n e^{i2\pi\upsilon n}$$

$$f_n = f(nx_s)$$

$$\upsilon = x_s\xi = \xi/\xi_s$$

Equation Summary

z-Transform:

$$F(z) = \mathcal{Z}\{f_n\} = \sum_{n=-\infty}^{\infty} f_n z^{-n}$$

$$z = \rho e^{i2\pi \upsilon}$$

Discrete Fourier Transform (DFT):

$$\mathcal{D}\{f_k\} = F_n = \sum_{k=0}^{N-1} f_k e^{-i2\pi \frac{nk}{N}}$$

$$\mathcal{D}^{-1}\{F_n\} = f_k = \frac{1}{N}\sum_{n=0}^{N-1} F_n e^{+i2\pi \frac{nk}{N}}$$

Parseval Theorem:

$$\int_{-\infty}^{\infty} |f(x)|^2 \, dx = \int_{-\infty}^{\infty} |F(\xi)|^2 \, d\xi$$

Moment Theorem:

$$m_n[f(x)] = \int_{-\infty}^{\infty} x^n f(x) dx = \frac{1}{(-i2\pi)^n} \left. \frac{d^n F(\xi)}{d\xi^n} \right|_{\xi=0}$$

Convolution:

$$f(x) * h(x) = \int_{-\infty}^{\infty} f(x') h(x-x') dx'$$

$$f(x) * \delta^{(n)}(x-x_0) = f^{(n)}(x-x_0)$$

Correlation:

$$\gamma_{fg}(x) = f(x) \star g(x) = \int_{-\infty}^{\infty} f(x') g^*(x'-x) dx'$$

Wiener Filter:

$$\frac{\mathrm{PSD}_s(\nu)}{\mathrm{PSD}_s(\nu) + \mathrm{PSD}_n(\nu)}$$

Wiener–Helstrom Filter:

$$\frac{H^*(\nu)}{|H(\nu)|^2 + \frac{\mathrm{PSD}_n(\nu)}{\mathrm{PSD}_s(\nu)}}$$

Equation Summary

Rayleigh–Sommerfeld Diffraction Integral:

$$u(\mathbf{r}_1) \approx \frac{z_{01}}{i\lambda} \iint_{-\infty}^{\infty} u(\mathbf{r}_0) \frac{e^{ikR}}{R^2} d^2 r_0$$

Fresnel Diffraction Integral:

$$u(\mathbf{r}_1) = \frac{e^{ikz_{01}}}{i\lambda z_{01}} \iint_{-\infty}^{\infty} u(\mathbf{r}_0) \exp\left[\frac{i\pi|\mathbf{r_1} - \mathbf{r_0}|^2}{\lambda z_{01}}\right] d^2 r_0$$

$$= \frac{e^{ikz_{01}}}{i\lambda z_{01}} \exp\left[\frac{i\pi r_1^2}{\lambda z_{01}}\right] \mathcal{F}\left\{u(\mathbf{r}_0) \exp\left[\frac{i\pi r_0^2}{\lambda z_{01}}\right]\right\}\Big|_{\boldsymbol{\rho} = \frac{\mathbf{r}_1}{\lambda z_{01}}}$$

Fraunhofer Diffraction Integral:

$$u(\mathbf{r}_1) = \frac{e^{ikz_{01}}}{i\lambda z_{01}} \exp\left[\frac{i\pi r_1^2}{\lambda z_{01}}\right] \mathcal{F}\{u(\mathbf{r}_0)\}\big|_{\boldsymbol{\rho} = \frac{\mathbf{r}_1}{\lambda z_{01}}}$$

Gaussian Beams:

$$u(\mathbf{r}, z) = \frac{b_0^2}{b_0^2 + i\lambda z} e^{ikz} \exp\left[\frac{-\pi r^2}{b(z)^2}\right] \exp\left[\frac{i\pi}{\lambda R(z)}\right]$$

$$b(z) = b_0 \sqrt{1 + \left(\frac{\lambda z}{b_0^2}\right)^2}$$

$$R(z) = z\left[1 + \left(\frac{b_0^2}{\lambda z}\right)^2\right]$$

CTF and Impulse Response:

$$H(\xi, \eta) = P(-\lambda d_i \xi, -\lambda d_i \eta)$$

$$h(x) = \mathcal{F}^{-1}\{H(\xi, \eta)\}$$

OTF and PSF:

$$\text{PSF}(x, y) = \frac{|h(x, y)|^2}{\iint |h(x, y)|^2 dx dy}$$

$$\mathcal{H}(\xi, \eta) = \mathcal{F}\{\text{PSF}(x, y)\}$$

$$\mathcal{H}(\xi, \eta) = \frac{\iint H(\xi', \eta') H^*(\xi' - \xi, \eta' - \eta) d\xi' d\eta'}{\iint |H(\xi', \eta')|^2 d\xi' d\eta'}$$

Bibliography

M. J. Ablowitz and A. S. Fokas, *Complex Variables*, Cambridge University Press, New York (1997).

N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, Saunders College Press, Philadephia (1976).

H. H. Barrett and K. J. Myers, *Foundations of Image Science*, John Wiley & Sons, New York (2004).

R. L. Easton, Jr., *Fourier Methods in Imaging*, John Wiley & Sons, New York (2010).

J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics*, John Wiley & Sons, New York (1978).

J. W. Goodman, *Fourier Optics*, Third Ed., Roberts & Company, Englewood, CO (2005).

J. E. Greivenkamp, *Field Guide to Geometrical Optics*, SPIE Press, Bellingham, WA (2004) [doi: 10.1117/3.547461].

E. Hecht, *Optics*, Fourth Ed., Pearson Addison-Wesley, San Francisco (2002).

M. V. Klein and T. E. Furtak, *Optics*, Second Ed., John Wiley & Sons, New York (1986).

A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, Englewood Cliffs, NJ (1989).

D. G. Smith, *Field Guide to Physical Optics*, SPIE Press, Bellingham, WA (2013) [doi: 10.1117/3.883971].

Index

Index

Index

Index

J. Scott Tyo has been a professor in the College of Optical Sciences at the University of Arizona (UA) since 2006 where he directs the Advanced Sensing Laboratory. After receiving the PhD degree in electrical engineering in 1997 from the Univeristy of Pennsylvania, he was a professor in the Electrical and Computer Engineering Departments at the University of New Mexico and the US Naval Postgraduate School. He served as an officer in the US Air Force from 1994—2001. He is a fellow of SPIE, the Optical Society of America, and the IEEE. Prof. Tyo's research interests cover a wide range of optical and electromagnetic sensing, including imaging polarimetry, imaging spectrometry, infrared imaging, computational imaging, and active sensing systems.

Andrey S. Alenin has been with J. Scott Tyo's Advanced Sensing Laboratory at the University of Arizona (UA) since 2010 where he is currently finishing his PhD titled, "Matrix structure for information-driven polarimeter design." Prior to joining ASL, he received a BS degree in optical sciences, and a BS degree in electrical engineering from the University of Arizona. Topics of his research include conventional, channeled and partial Mueller matrix polarimetry, the mathematics of which are highly dependent on the concepts of linear systems. Besides his work at the UA, he has worked on various software engineering problems as a consultant.

Linear Systems in Optics

J. Scott Tyo
Andrey S. Alenin

"Linear systems" is a broad and important area in many scientific and engineering disciplines, and it is especially important in optics because it forms the basis for Fourier optics, diffraction theory, image-quality assessment, and many other areas. This Field Guide provides the practicing optical engineer with a reference for the basic concepts and techniques of linear systems, including Fourier series, continuous and discrete Fourier transforms, convolution, sampling and aliasing, and MTF/ PSF using the language, notation, and applications from optics, imaging, and diffraction.

SPIE Field Guides

The aim of each *SPIE Field Guide* is to distill a major field of optical science or technology into a handy desk or briefcase reference that provides basic, essential information about optical principles, techniques, or phenomena.

Written for you—the practicing engineer or scientist— each field guide includes the key definitions, equations, illustrations, application examples, design considerations, methods, and tips that you need in the lab and in the field.

John E. Greivenkamp
Series Editor

www.spie.org/press/fieldguides