Chapter 7
Sampling Requirements for Fresnel Diffraction

The primary reason to use simulations is to tackle problems that are analytically intractable. As a result, any computer code that simulates optical-wave propagation needs to handle almost any type of source field. Wave-optics simulations are based on DFTs, and we saw in Ch. 2 that aliasing poses a challenge to DFTs. When the waveform to be transformed is bandlimited, we just need to sample it finely enough to avoid aliasing altogether (satisfying the Nyquist criterion). However, most optical sources are not spatially bandlimited, and the quadratic phase term inside the Fresnel diffraction integral certainly is not bandlimited. These issues have been explored by many authors.[30,31,35,37,42,54,55]

Because an optical field's spatial-frequency spectrum maps directly to its plane-wave spectrum,[5] propagation geometry places a limit on how much spatial frequency content from the source can be seen within the observing aperture. Note that this is physical; it is not caused by sampling. This principle is the foundation of Coy's approach to sampling, and guides most of our discussion on sampling needs in this chapter.

7.1 Imposing a Band Limit

The optical field at each point in the source plane emits a bundle of rays that propagate toward the observation plane. Each ray represents a plane wave propagating in that direction. Let us start by examining the propagation geometry to determine the maximum plane-wave direction relative to the reference normal from the source that is incident upon the region of interest in the observation plane.

Clearly, it is critical to pick the grid spacing and number of grid points to ensure an accurate simulation. The following development uses the propagation geometry to place limits on the necessary spatial-frequency bandwidth, and consequently, the number of sample points and grid spacing. This determines the size and spacing of the source-plane grid and the size and spacing of the observation-plane grid.

At this point, we need to recall the Nyquist criterion to place a constraint on the

grid spacing such that

$$\delta \leq \frac{1}{2f_{max}}, \tag{7.1}$$

where f_{max} is the maximum spatial frequency of interest. To build a link between ray angles and spatial bandwidth, we can rewrite Eq. (6.5) in operator notation (just for the FT) as

$$U(x_2, y_2) = \frac{e^{ik\Delta z}}{i\lambda\Delta z} e^{i\frac{k}{2\Delta z}(x_2^2+y_2^2)} \mathcal{F}\left[\mathbf{r}_1, \mathbf{f}_1 = \frac{\mathbf{r}_2}{\lambda\Delta z}\right] \left\{U(x_1, y_1) e^{i\frac{k}{2\Delta z}(x_1^2+y_1^2)}\right\}. \tag{7.2}$$

The quadratic phase factor inside the FT is interesting; it represents a virtual spherical wave that is focused onto the observation plane. It appears as if the source field's phase is being measured with respect to this spherical surface. After "re-measuring" the phase in this way, the source field is transformed so that each spatial-frequency vector $\mathbf{f}_1$ corresponds to a specific coordinate in the observation plane. Below, we exploit this link between geometry and spatial frequency to levy constraints on the sampling grids.

In the angular-spectrum formulation of diffraction, the concept is that an optical field $U(x, y)$ may be decomposed into a sum of plane waves with varying amplitudes and directions. A plane wave $U_p(x, y, z, t)$ with arbitrary direction is given in phasor notation by

$$U_p(x, y, z, t) = e^{i(\mathbf{k}\cdot\mathbf{r}-2\pi\nu t)}, \tag{7.3}$$

where $\mathbf{r} = x\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}}$ is a three-dimensional position vector, $\mathbf{k} = (2\pi/\lambda)\left(\alpha\hat{\mathbf{i}} + \beta\hat{\mathbf{j}} + \gamma\hat{\mathbf{k}}\right)$ is the optical wavevector, and ν is the temporal frequency of the optical wave. These direction cosines are depicted in Fig. 7.1. Using phasor notation, a plane wave is given by

$$U_p(x, y, z) = e^{i\mathbf{k}\cdot\mathbf{r}} = e^{i\frac{2\pi}{\lambda}(\alpha x+\beta y)} e^{i\frac{2\pi}{\lambda}\gamma z}. \tag{7.4}$$

In the $z = 0$ plane, a complex-exponential source in the form $\exp[i2\pi(f_x x + f_y y)]$ may be regarded as a plane wave propagating with direction cosines

$$\alpha = \lambda f_x, \qquad \beta = \lambda f_y, \qquad \gamma = \sqrt{1 - (\lambda f_x)^2 - (\lambda f_y)^2}. \tag{7.5}$$

Therefore, the spatial-frequency spectrum of an optical source is also its plane-wave spectrum with the spatial frequencies mapped to direction cosines (α, β), where the mapping is given in Eqs. (7.5). Figure 7.1 illustrates the geometry of these direction cosines. From this, the angular spectrum's cutoff angle is defined as $\alpha_{max} = \lambda f_{max}$, where α_{max} is the maximum angle in the angular spectrum that can affect the observed field. Now, Eq. (7.1) may be rewritten to relate an optical field's maximum angular content to the grid spacing so that

$$\delta_1 \leq \frac{\lambda}{2\alpha_{max}}. \tag{7.6}$$

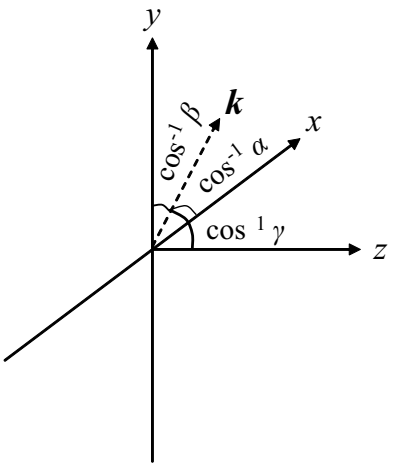

Figure 7.1 Depiction of direction cosines α, β, and γ.

Conversely, if the grid spacing is given, then the maximum angular content represented by the sampled version of the optical field is

$$\alpha_{max} = \frac{\lambda}{2\delta_1}. \tag{7.7}$$

This allows us to tie grid parameters to the propagation geometry.

7.2 Propagation Geometry

Now, the task is to use the sizes of the source and receiver to determine α_{max}. This section follows the developments of Coy, Praus, and Mansell.[35,42,54] The discussion is restricted to one spatial dimension, but it may easily be generalized to two dimensions. Additionally, the propagating wavefront is assumed to be spherical for generality.

As shown in Fig. 7.2, the source field has a maximum spatial extent D_1. In the observation plane, the region of interest has a maximum spatial extent D_2. Perhaps the optical field is propagating to a sensor, and D_2 is the diameter of the sensor. Additionally, let the grid spacing in the source plane be δ_1 and the grid spacing in the observation plane be δ_2.

While the source field can be considered a sum of plane waves as discussed above, it can alternately be considered a sum of point sources. This is precisely Huygens' principle. We take this view so that we ensure the grids are sampled finely enough that each point in the source field fully illuminates the observation-plane region of interest. The maximum ray angle α_{max} corresponds to the divergence angle of source-plane field points.

Consider a point at the lower edge of the source, at point $(x_1 = -D_1/2, z = z_1)$. The angle α_{max} can be written as the sum of two angles α_k and α_{edges}, as shown in Fig. 7.2. The angle between the bottom edge of the source and the top edge of

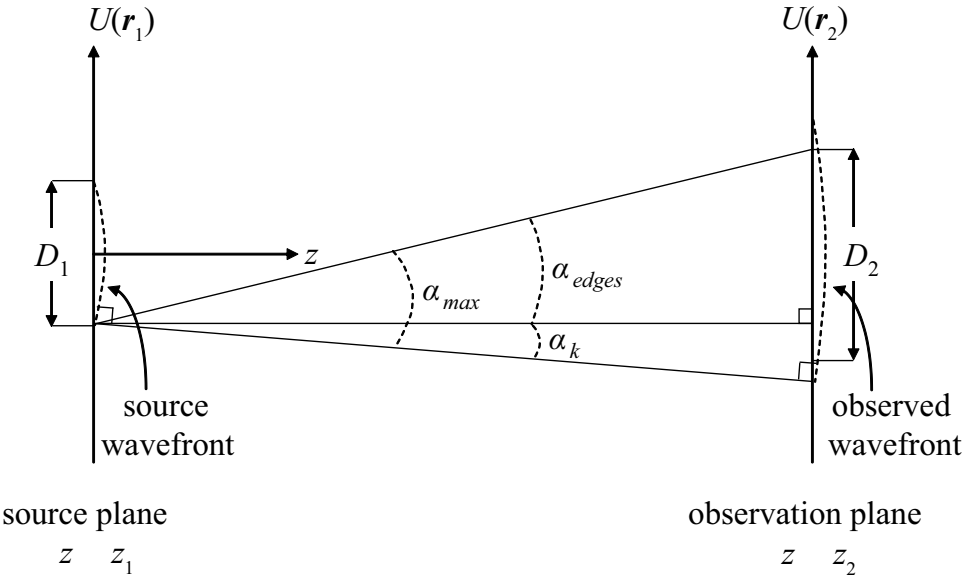

Figure 7.2 Definition of angles α_{max}, α_{edges}, and α_k.

the observing aperture, at point $(x_2 = D_2/2, z = z_2)$, is (in the paraxial approximation)

$$\alpha_{edges} = \frac{D_1 + D_2}{2\Delta z}. \tag{7.8}$$

At the lower edge of the source, the optical wavevector $\mathbf{k}$ of the virtual spherical wave apparent in Eq. (7.2) makes an angle with the z axis. Because there is a fixed number of grid points, spaced by a distance δ_1 in the source plane and δ_2 in the observation plane, the ratio of the grid sizes (observation/source) is δ_2/δ_1. Thus, $\mathbf{k}$ intersects the observation plane at $x_2 = -D_1\delta_2/\left(2\delta_1\right)$. Consequently, the (paraxial) angle α_k is given by

$$\alpha_k = \frac{D_1\delta_2}{2\delta_1\Delta z} - \frac{D_1}{2\Delta z} = \frac{D_1}{2\Delta z}\left(\frac{\delta_2}{\delta_1} - 1\right). \tag{7.9}$$

Then, α_{max} is given by

$$\alpha_{max} = \alpha_{edges} + \alpha_k \tag{7.10}$$

$$= \frac{D_1 + D_2}{2\Delta z} + \frac{D_1}{2\Delta z}\left(\frac{\delta_2}{\delta_1} - 1\right) \tag{7.11}$$

$$= \frac{D_1\,\delta_2/\delta_1 + D_2}{2\Delta z}. \tag{7.12}$$

When this is combined with the sampling requirement in Eq. (7.7), the result is

$$\frac{D_1\,\delta_2/\delta_1 + D_2}{2\Delta z} \leq \frac{\lambda}{2\delta_1} \tag{7.13}$$

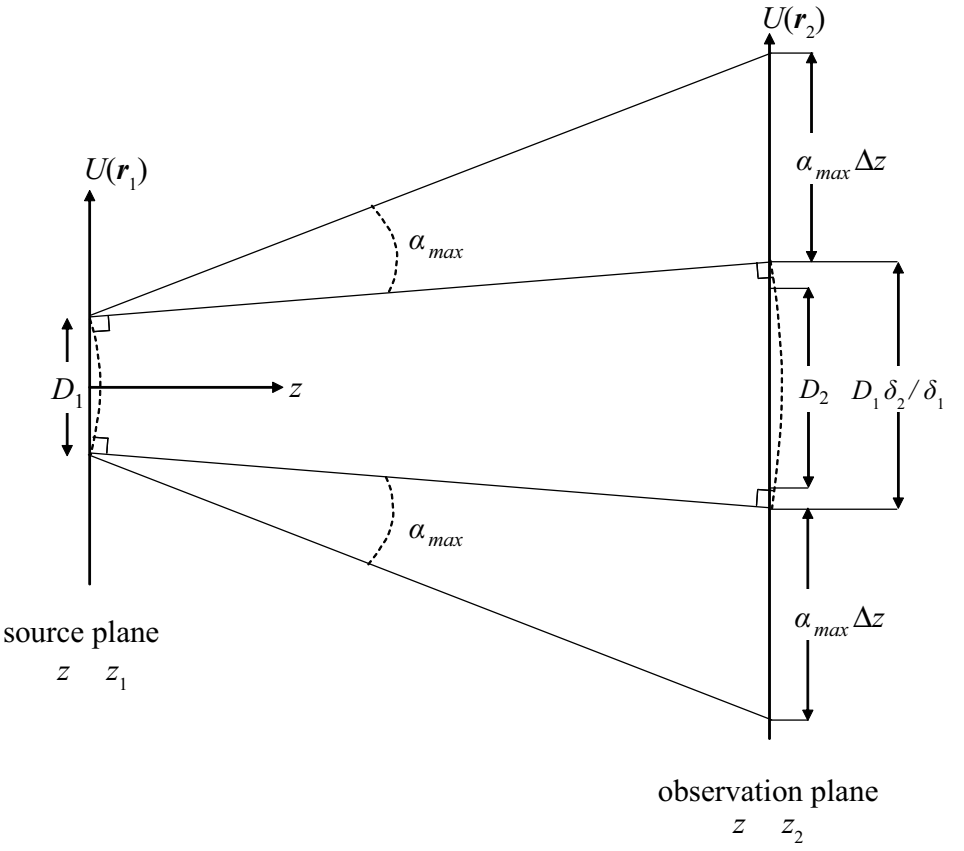

Figure 7.3 Portion of the observation plane affected by the maximum angular content.

$$\delta_2 \leq -\frac{D_2}{D_1}\delta_1 + \frac{\lambda \Delta z}{D_1}. \tag{7.14}$$

Satisfying Eq. (7.14) means that the selected grid spacings adequately sample the spatial bandwidth that affects the observation-plane region of interest.

Now, it is useful to determine the necessary spatial extent of the observation-plane grid. Figure 7.3 shows that the diameter D_{illum} of illuminated area (by a source with maximum angular content α_{max}) in the observation plane is

$$D_{illum} = D_1\delta_2/\delta_1 + 2\alpha_{max}\Delta z \tag{7.15}$$

$$= D_1\delta_2/\delta_1 + \frac{\lambda \Delta z}{\delta_1}. \tag{7.16}$$

Aliasing in the observation plane is allowable as long as it does not invade the area of the observing aperture. If the grid has a smaller spatial extent than the illuminated area, we can imagine the edges of the illuminated area wrapping around to the other side of the grid. Recall that this is apparent in Figs. 2.6(d) and 2.7(d). For the wrapping to get just to the edge of the observing aperture, the grid extent must

be at least as large as the mean of the illuminated area and the observing aperture diameter so that it wraps only half-way around, yielding

$$D_{grid} \geq \frac{D_{illum} + D_2}{2} \tag{7.17}$$

$$= \frac{D_1\delta_2/\delta_1 + \lambda\Delta z/\delta_1 + D_2}{2}. \tag{7.18}$$

Finally, the number of grid points required in the observation plane is

$$N = \frac{D_{grid}}{\delta_2} \tag{7.19}$$

$$\geq \frac{D_1}{2\delta_1} + \frac{D_2}{2\delta_2} + \frac{\lambda\Delta z}{2\delta_1\delta_2}. \tag{7.20}$$

Satisfying Eq. (7.20) means that the spatial extent of the observation plane is large enough to ensure that the light that wraps around does not creep into the observation-plane region of interest.

7.3 Validity of Propagation Methods

Unfortunately, satisfying the geometric constraints to avoid aliasing in the observation-plane region of interest does not guarantee satisfactory results. One must also consider which method of propagation can be used. The Fresnel-integral method and the angular-spectrum method have different constraints. One must avoid aliasing the quadratic phase factor inside the FTs that are used, and the two propagation methods have different two quadratic phase factors. With these different constraints, it turns out that the Fresnel-integral approach from Sec. 6.3 is valid for long propagations, while the angular-spectrum approach from Sec. 6.4 is valid for short propagations.[30,31,37]

7.3.1 Fresnel-integral propagation

This subsection begins by applying the geometric constraints with consideration of the particular grid spacing allowed by Fresnel-integral propagation. Then, it goes on to examine how to avoid aliasing of the quadratic phase factor in the source plane. These analyses result in a set of inequalities that must be satisfied when choosing the grid parameters.

7.3.1.1 One step, fixed observation-plane grid spacing

As discussed in the previous chapter, the observation-plane grid spacing δ_2 is fixed when one executes a single step of Fresnel-integral propagation. This fixed value is

$$\delta_2 = \frac{\lambda\Delta z}{N\delta_1}. \tag{7.21}$$

Relating this to the propagation geometry, we substitute this into Eq. (7.14), which yields

$$D_1 \frac{\lambda \Delta z}{N \delta_1} + D_2 \delta_1 \leq \lambda \Delta z \tag{7.22}$$

$$D_1 \frac{\lambda \Delta z}{\delta_1} + D_2 \delta_1 N \leq N \lambda \Delta z \tag{7.23}$$

$$D_1 \frac{\lambda \Delta z}{\delta_1} \leq N \left(\lambda \Delta z - D_2 \delta_1 \right) \tag{7.24}$$

$$N \geq \frac{D_1 \lambda \Delta z}{\delta_1 \left(\lambda \Delta z - D_2 \delta_1 \right)}. \tag{7.25}$$

Substituting for δ_2 in Eq. (7.20) yields

$$N \geq \frac{D_1}{2\delta_1} + \frac{D_2 \delta_1}{2 \lambda \Delta z} N + \frac{\lambda \Delta z}{2 \delta_1} \frac{N \delta_1}{\lambda \Delta z} \tag{7.26}$$

$$N \geq \frac{D_1}{2\delta_1} + \frac{D_2 \delta_1}{2 \lambda \Delta z} N + \frac{N}{2} \tag{7.27}$$

$$\frac{N}{2} - \frac{D_2 \delta_1}{2 \lambda \Delta z} N \geq \frac{D_1}{2 \delta_1} \tag{7.28}$$

$$N \left(1 - \frac{D_2 \delta_1}{\lambda \Delta z} \right) \geq \frac{D_1}{\delta_1} \tag{7.29}$$

$$N \geq \frac{D_1}{\delta_1 \left(1 - \frac{D_2 \delta_1}{\lambda \Delta z} \right)} \tag{7.30}$$

$$N \geq \frac{D_1 \lambda \Delta z}{\delta_1 \left(\lambda \Delta z - D_2 \delta_1 \right)}. \tag{7.31}$$

This is identical to Eq. (7.25)! Also notice two properties of this inequality: we must have $\lambda \Delta z > D_2 \delta_1$ because N can only be positive, and as $\lambda \Delta z \to D_2 \delta_1$ the minimum necessary N approaches ∞.

7.3.1.2 Avoiding aliasing

The free-space amplitude spread function has a very large bandwidth. In fact, the cutoff frequency is λ^{-1}, which is impractically high to represent on a grid of finite size.[5] If we tried to use a source-plane grid spacing of $\delta_1 = \lambda/2 \approx 500$ nm, the largest grid extent that could be used is $L = N\delta_1 \approx 500 \text{ nm} \times 1024 = 0.512$ mm (grid sizes up to 2048 or 4096 might be possible, depending on the computer being used). Of course, very few practical problems can be simulated on such a small grid.

In practice, the best one can do is to ensure that all of the frequencies present on the grid are represented correctly. We cannot plan for all possible kinds of source-plane fields, so we derive a sampling guideline by modeling the source as

an apodized beam with maximum spatial extent D_1 and a parabolic wavefront with radius R. This source field $U(\mathbf{r}_1)$ can be written as

$$U(\mathbf{r}_1) = A(\mathbf{r}_1)\, e^{i\frac{k}{2R}r_1^2}, \tag{7.32}$$

where $A(\mathbf{r}_1)$ describes the amplitude transmittance of the source aperture. The maximum spatial extent of the nonzero portions of $A(\mathbf{r}_1)$ is D_1. A diverging beam is indicated by $R < 0$, while a converging beam is indicated by $R > 0$. With this type of source, the Fresnel diffraction integral becomes

$$U(\mathbf{r}_2) = \mathcal{Q}\left[\frac{1}{\Delta z}, \mathbf{r}_2\right] \mathcal{V}\left[\frac{1}{\lambda \Delta z}, \mathbf{r}_1\right] \mathcal{F}[\mathbf{r}_1, \mathbf{f}_1]\, \mathcal{Q}\left[\frac{1}{\Delta z}, \mathbf{r}_1\right] \{U(\mathbf{r}_1)\} \tag{7.33}$$

$$= \mathcal{Q}\left[\frac{1}{\Delta z}, \mathbf{r}_2\right] \mathcal{V}\left[\frac{1}{\lambda \Delta z}, \mathbf{r}_1\right] \mathcal{F}[\mathbf{r}_1, \mathbf{f}_1]\, \mathcal{Q}\left[\frac{1}{\Delta z}, \mathbf{r}_1\right] \left\{A(\mathbf{r}_1)\, e^{i\frac{k}{2R}r_1^2}\right\} \tag{7.34}$$

$$= \mathcal{Q}\left[\frac{1}{\Delta z}, \mathbf{r}_2\right] \mathcal{V}\left[\frac{1}{\lambda \Delta z}, \mathbf{r}_1\right] \mathcal{F}[\mathbf{r}_1, \mathbf{f}_1]\, \mathcal{Q}\left[\frac{1}{\Delta z}, \mathbf{r}_1\right] \mathcal{Q}\left[\frac{1}{R}, \mathbf{r}_1\right] \{A(\mathbf{r}_1)\} \tag{7.35}$$

$$= \mathcal{Q}\left[\frac{1}{\Delta z}, \mathbf{r}_2\right] \mathcal{V}\left[\frac{1}{\lambda \Delta z}, \mathbf{r}_1\right] \mathcal{F}[\mathbf{r}_1, \mathbf{f}_1]\, \mathcal{Q}\left[\frac{1}{\Delta z} + \frac{1}{R}, \mathbf{r}_1\right] \{A(\mathbf{r}_1)\}. \tag{7.36}$$

The key to achieving an accurate result is to sample the quadratic phase factor inside the FT at a high enough rate to satisfy the Nyquist criterion. If it is not sampled finely enough, the intended high-frequency content would show up in the lower frequencies. Again, this effect is visible in Figs. 2.6(d) and 2.7(d). Lower frequencies map to lower ray angles that may erroneously impinge on the observation-plane region of interest.

To avoid or at least minimize aliasing, we need to determine the bandwidth of the product $\mathcal{Q}A$ from Eq. (7.36). Lambert and Fraser demonstrated that for very small apertures, the bandwidth is set by A, while for larger apertures, it is set by the phase of $\mathcal{Q}$ at the edge of the aperture.[47] Typically, the latter is the case, so we focus on the phase of $\mathcal{Q}$. Local spatial frequency $\mathbf{f}_{loc}$ is basically the local rate of change of a waveform given by[5]

$$\mathbf{f}_{loc} = \frac{1}{2\pi}\nabla\phi, \tag{7.37}$$

where ϕ is the optical phase measured in radians, and the Cartesian components of $\mathbf{f}_{loc}$ are measured in m^{-1}. Conceptually, a waveform with rapid variations (regions of large gradients) has high-frequency content. We want to find the maximum local spatial frequency of the quadratic phase factor inside the integral and sample at least twice this rate. Since the quadratic phase has the same variations in the both

Cartesian directions, we just analyze the x_1 direction, which yields

$$f_{locx} = \frac{1}{2\pi}\frac{\partial}{\partial x_1}\frac{k}{2}\left(\frac{1}{\Delta z}+\frac{1}{R}\right)r_1^2 \tag{7.38}$$

$$= \left(\frac{1}{\Delta z}+\frac{1}{R}\right)\frac{x_1}{\lambda}. \tag{7.39}$$

This takes on its maximum value at the edge of the grid where $x_1 = N\delta_1/2$. However, if the source is apodized, and the field is nonzero only within a centered aperture of maximum extent D_1, then that includes the phase. Thus, the product of the source field and the quadratic phase factor has its maximum local spatial frequency value at $x_1 = \pm D_1/2$. Then, applying the Nyquist criterion yields

$$\left(\frac{1}{\Delta z}+\frac{1}{R}\right)\frac{D_1}{2\lambda} \leq \frac{1}{2\delta_1}. \tag{7.40}$$

After some algebra, we obtain

$$\Delta z \geq \frac{D_1\delta_1 R}{\lambda R - D_1\delta_1} \qquad \text{for finite } R \tag{7.41}$$

$$\Delta z \geq \frac{D_1\delta_1}{\lambda} \qquad \text{for infinite } R. \tag{7.42}$$

Note that this is just a guideline. When Δz is close to its minimum required value, the simulation results may not match analytic results perfectly.

The following example illustrates the process of using a sound analysis of sampling to obtain accurate simulation results. Listing 7.1 gives an example of subsequent usage of `one_step_prop` for a square aperture with due consideration of sampling constraints. It goes on to plot the results along with the analytic result. In line 10, the minimum number of grid points is computed using Eq. (7.25). In this example, 66 grid points are required. Then, in line 11, the number of grid points to actually use is determined by using the next power of two, which is 128. This is done to take advantage of the FFT algorithm. After line 11 executes, the sampling-related parameters for this simulation are

$$\begin{aligned} D_1 &= 2\,\text{mm} \\ D_2 &= 3\,\text{mm} \\ \lambda &= 1\,\mu\text{m} \\ \Delta z &= 0.5\,\text{m} \\ \delta_1 &= 40\,\mu\text{m} \\ \delta_2 &= 97.7\,\mu\text{m} \\ N &= 128. \end{aligned} \tag{7.43}$$

Applying Eq. (7.42), we find that the minimum distance to use one step of Fresnel-integral propagation is 8 cm. Clearly, we can expect results that match theory

Listing 7.1 Example of evaluating the Fresnel diffraction integral in MATLAB using a single step.

```
% example_square_one_step_prop_samp.m

D1 = 2e-3;  % diam of the source aperture [m]
D2 = 3e-3;  % diam of the obs-plane region of interest [m]
delta1 = D1 / 50;   % want at least 50 grid pts across ap
wvl = 1e-6;  % optical wavelength [m]
k = 2*pi / wvl;
Dz = 0.5;         % propagation distance [m]
% minimum number of grid points
Nmin = D1 * wvl*Dz / (delta1 * (wvl*Dz - D2*delta1));
N = 2^ceil(log2(Nmin));     % number of grid pts per side
% source plane
[x1 y1] = meshgrid((-N/2 : N/2-1) * delta1);
ap = rect(x1/D1) .* rect(y1/D1);
% simulate the propagation
[x2 y2 Uout] = one_step_prop(ap, wvl, delta1, Dz);

% analytic result for y2=0 slice
Uout_an ...
    = fresnel_prop_square_ap(x2(N/2+1,:), 0, D1, wvl, Dz);
```

closely because there are more than enough grid points (by nearly a factor of two), and the propagation is much farther than the limit required by this simulation method. Figure 7.4 shows the resulting amplitude and phase. The simulation does, in fact, match the analytic results closely.

7.3.2 Angular-spectrum propagation

For the angular-spectrum method, the observation-plane grid spacing is not fixed like in the previous section. The grid spacings δ_1 and δ_2 can be chosen independently so, there are no simplifications to Eqs. (7.14) and (7.20) like with the Fresnel-integral method. Instead, there are two additional inequalities that must be satisfied to keep high-frequency content from corrupting the observation-plane region of interest. This is because the angular-spectrum method from Eq. (6.67) has its own requirements to avoid aliasing of a quadratic phase factor. As in the previous section, we restrict the source-plane field $U(\mathbf{r}_1)$ to the form in Eq. (7.32). With this form, the angular-spectrum method can be written as

$$U(\mathbf{r}_2) = \mathcal{Q}\left[\frac{m-1}{m\Delta z}, \mathbf{r}_2\right] \mathcal{F}^{-1}\left[\mathbf{f}_1, \frac{\mathbf{r}_2}{m}\right] \mathcal{Q}_2\left[-\frac{\Delta z}{m}, \mathbf{f}_1\right] \times \mathcal{F}[\mathbf{r}_1, \mathbf{f}_1] \mathcal{Q}\left[\frac{1-m}{\Delta z}, \mathbf{r}_1\right] \frac{1}{m}\{U(\mathbf{r}_1)\} \tag{7.44}$$

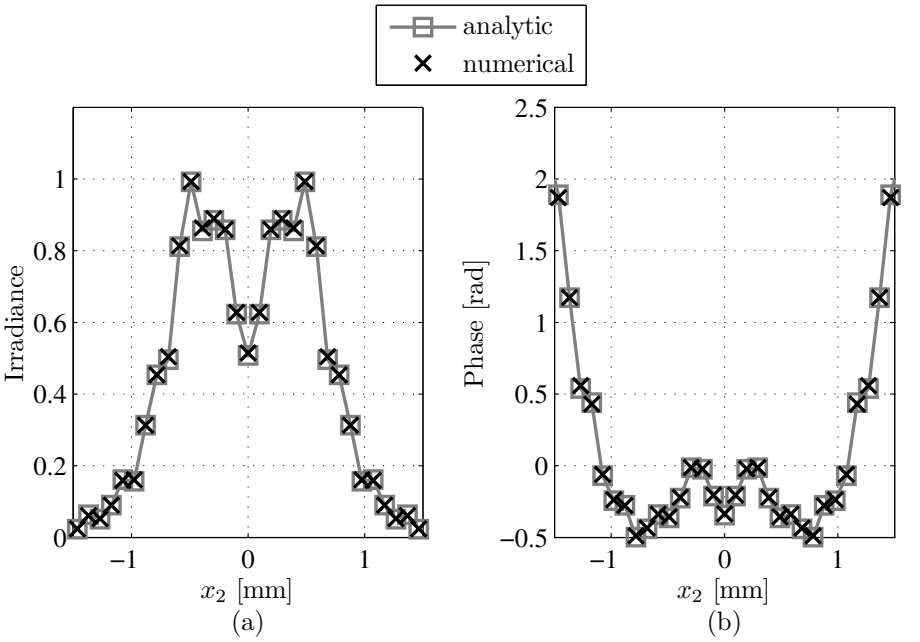

Figure 7.4 Fresnel diffraction from a square aperture, simulation and analytic: (a) observation-plane irradiance and (b) observation-plane phase.

$$
\begin{aligned}
&= \mathcal{Q}\left[\frac{m-1}{m\Delta z}, \mathbf{r}_2\right] \mathcal{F}^{-1}\left[\mathbf{f}_1, \frac{\mathbf{r}_2}{m}\right] \mathcal{Q}_2\left[-\frac{\Delta z}{m}, \mathbf{f}_1\right] \\
&\quad \times \mathcal{F}\left[\mathbf{r}_1, \mathbf{f}_1\right] \mathcal{Q}\left[\frac{1-m}{\Delta z}, \mathbf{r}_1\right] \frac{1}{m}\left\{A\left(\mathbf{r}_1\right) e^{i\frac{k}{2R}r_1^2}\right\} \\
&= \mathcal{Q}\left[\frac{m-1}{m\Delta z}, \mathbf{r}_2\right] \mathcal{F}^{-1}\left[\mathbf{f}_1, \frac{\mathbf{r}_2}{m}\right] \mathcal{Q}_2\left[-\frac{\Delta z}{m}, \mathbf{f}_1\right] \\
&\quad \times \mathcal{F}\left[\mathbf{r}_1, \mathbf{f}_1\right] \mathcal{Q}\left[\frac{1-m}{\Delta z}, \mathbf{r}_1\right] \frac{1}{m} \mathcal{Q}\left[\frac{1}{R}, \mathbf{r}_1\right] \left\{A\left(\mathbf{r}_1\right)\right\} \\
&= \mathcal{Q}\left[\frac{m-1}{m\Delta z}, \mathbf{r}_2\right] \mathcal{F}^{-1}\left[\mathbf{f}_1, \frac{\mathbf{r}_2}{m}\right] \mathcal{Q}_2\left[-\frac{\Delta z}{m}, \mathbf{f}_1\right] \\
&\quad \times \mathcal{F}\left[\mathbf{r}_1, \mathbf{f}_1\right] \frac{1}{m} \mathcal{Q}\left[\frac{1-m}{\Delta z} + \frac{1}{R}, \mathbf{r}_1\right] \left\{A\left(\mathbf{r}_1\right)\right\}. \qquad (7.45)
\end{aligned}
$$

There are two quadratic phase factors inside the FT (and IFT) operations to consider:

$$
\mathcal{Q}\left[\frac{1-m}{\Delta z} + \frac{1}{R}, \mathbf{r}_1\right] = \exp\left[-i\frac{k}{2}\left(\frac{1-m}{\Delta z} + \frac{1}{R}\right)\left|\mathbf{r}_1\right|^2\right] \qquad (7.46)
$$

$$
\mathcal{Q}_2\left[-\frac{\Delta z}{m}, \mathbf{f}_1\right] = \exp\left(i\pi^2 \frac{2\Delta z}{mk}\left|\mathbf{f}_1\right|^2\right). \qquad (7.47)
$$

Like in the previous section, we need to compute the maximum local spatial frequency in each factor and apply the Nyquist sampling criterion. This ensures that all

of the present spatial frequencies are not aliased, thus preserving the observation-plane field within the region of interest.

In the first phase factor, the phase ϕ is

$$\phi = \frac{k}{2}\left(\frac{1-m}{\Delta z} + \frac{1}{R}\right)|\mathbf{r}_1|^2 \tag{7.48}$$

$$= \frac{k}{2}\left(\frac{1-\delta_2/\delta_1}{\Delta z} + \frac{1}{R}\right)|\mathbf{r}_1|^2. \tag{7.49}$$

The local spatial frequency f_{lx} is

$$f_{lx} = \frac{1}{2\pi}\frac{\partial}{\partial x_1}\phi \tag{7.50}$$

$$= \frac{1}{\lambda}\left(\frac{1-\delta_2/\delta_1}{\Delta z} + \frac{1}{R}\right)x_1. \tag{7.51}$$

Once again, the maximum spatial frequency occurs at $x_1 = \pm D_1/2$ because this factor is multiplied by the source-plane pupil function. Applying the Nyquist sampling gives

$$\frac{1}{\lambda}\left|\frac{1-\delta_2/\delta_1}{\Delta z} + \frac{1}{R}\right|\frac{D_1}{2} \leq \frac{1}{2\delta_1}. \tag{7.52}$$

After some algebra, we obtain

$$\left(1+\frac{\Delta z}{R}\right)\delta_1 - \frac{\lambda\Delta z}{D_1} \leq \delta_2 \leq \left(1+\frac{\Delta z}{R}\right)\delta_1 + \frac{\lambda\Delta z}{D_1}. \tag{7.53}$$

The phase of the second quadratic phase factor (the amplitude transfer function) is

$$\phi = \pi^2\frac{2\Delta z}{mk}|\mathbf{f}_1|^2 \tag{7.54}$$

$$= \pi^2\frac{2\delta_1\Delta z}{\delta_2 k}|\mathbf{f}_1|^2. \tag{7.55}$$

The local spatial frequency f'_{lx} (prime notation to avoid confusion with the variable in the quadratic phase factor) is

$$f'_{lx} = \frac{1}{2\pi}\frac{\partial}{\partial f_{1x}}\phi \tag{7.56}$$

$$= \frac{\delta_1\lambda\Delta z}{\delta_2}f_{1x}. \tag{7.57}$$

This is a maximum at the edge of the spatial-frequency grid where $f_{1x} = \pm 1/(2\delta_1)$. Applying Nyquist sampling criterion gives

$$\frac{\lambda\Delta z}{2\delta_2} \leq \frac{N\delta_1}{2} \tag{7.58}$$

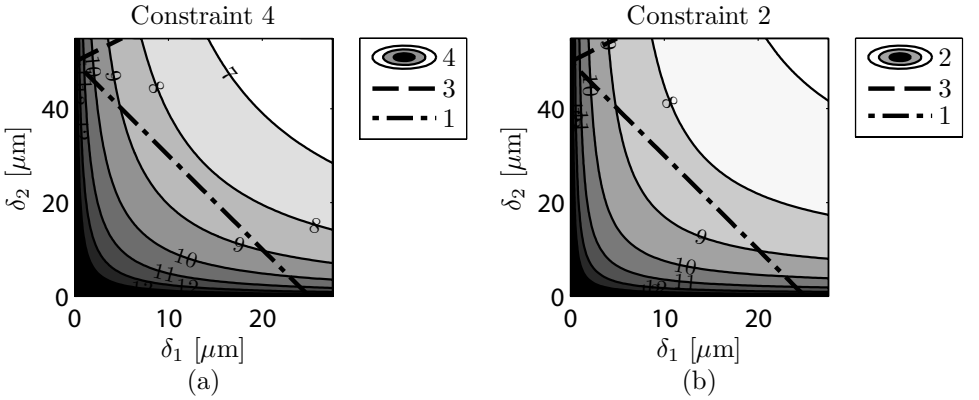

Figure 7.5 Sampling constraints for the angular-spectrum propagation method: (a) constraints 4, 3, and 1; (b) constraints 2, 3, and 1.

$$N \geq \frac{\lambda \Delta z}{\delta_1 \delta_2}. \tag{7.59}$$

Because there are four inequalities, the procedure here is more complicated than for Fresnel-integral propagation. Again, the simplest way to illustrate this procedure is by example. Let us restate the sampling constraints grouped together:

1. $\delta_2 \leq -\frac{D_2}{D_1}\delta_1 + \frac{\lambda \Delta z}{D_1}$,
2. $N \geq \frac{D_1}{2\delta_1} + \frac{D_2}{2\delta_2} + \frac{\lambda \Delta z}{2\delta_1 \delta_2}$,
3. $\left(1 + \frac{\Delta z}{R}\right)\delta_1 - \frac{\lambda \Delta z}{D_1} \leq \delta_2 \leq \left(1 + \frac{\Delta z}{R}\right)\delta_1 + \frac{\lambda \Delta z}{D_1}$,
4. $N \geq \frac{\lambda \Delta z}{\delta_1 \delta_2}$.

Consider an example of evaluating Eq. (7.44) for the following parameters: $D_1 = 2$ mm, $D_2 = 4$ mm, $\Delta z = 0.1$ m, and $\lambda = 1\,\mu$m. Solving four inequalities simultaneously is challenging. The simplest approach is to graphically display the bounds for these inequalities in the (δ_1, δ_2) domain. These are shown in Fig. 7.5. Plot (a) shows a contour plot of the lower bound on $\log_2 N$ from constraint 4 (solid black lines). Also on the plot are the upper bounds on δ_2 given by constraints 1 (dash-dot line) and 3 (dashed line barely visible in the upper-left corner). Constraint 1 is clearly more restrictive than constraint 3 where δ_2 is concerned. When choosing values for δ_1 and δ_2, this limits us to the lower-left corner of the plot below the dotted line. The required number of grid points in this region of the contour plot is at least $2^{8.5}$. However, we realistically must pick an integer power of two to take advantage of the FFT algorithm, so it looks like we must choose $N = 2^9 = 512$ grid points. Somewhat arbitrarily choosing $\delta_1 = 9.48\,\mu$m and $\delta_2 = 28.12\,\mu$m, the minimum required number of grid points is $2^{8.55}$. Consequently, we must choose $N = 2^9 = 512$ grid points unless constraint 2 is more restrictive. Plot (b) indicates

Listing 7.2 Example of evaluating the Fresnel diffraction integral in MATLAB using the angular-spectrum method.

```
% example_square_prop_ang_spec.m

D1 = 2e-3;    % diameter of the source aperture [m]
D2 = 4e-3;    % diameter of the observation aperture [m]
wvl = 1e-6;  % optical wavelength [m]
k = 2*pi / wvl;
Dz = 0.1;      % propagation distance [m]
delta1 = 9.4848e-6;
delta2 = 28.1212e-6;
Nmin = D1/(2*delta1) + D2/(2*delta2) ...
     + (wvl*Dz)/(2*delta1*delta2);
% bump N up to the next power of 2 for efficient FFT
N = 2^ceil(log2(Nmin));

[x1 y1] = meshgrid((-N/2 : N/2-1) * delta1);
ap =rect(x1/D1) .* rect(y1/D1);
[x2 y2 Uout] = ang_spec_prop(ap, wvl, delta1, delta2, Dz);

% analytic result for y2=0 slice
Uout_an ...
     = fresnel_prop_square_ap(x2(N/2+1,:), 0, D1, wvl, Dz);
```

that the required number of grid points according to constraint 2 is only $2^{8.51}$. As a result, picking $N = 512$ is sufficient, given that $\delta_1 = 9.48\,\mu$m and $\delta_2 = 28.12\,\mu$m.

Listing 7.2 gives the MATLAB code for the simulation in this example. The code numerically evaluates the angular-spectrum method [Eq. (7.44)] to simulate propagation from a square aperture. The simulation uses the parameters from this discussion of sampling. Given all of this consideration to sampling, one expects that the amplitude and phase of the simulated result should match the analytic results closely. These results are shown in Fig. 7.6 with a $y_2 = 0$ slice of the irradiance shown in plot (a) and a $y_2 = 0$ slice of the wrapped phase shown in plot (b). Indeed, the simulation result does match the analytic result closely.

7.3.3 General guidelines

We can now formulate this problem more generally. First, it can be shown that constraint 4 is more restrictive than the combination of constraints 1 and 2. Therefore, only Fig. 7.5(a) needs to be analyzed, and plot (b) may be ignored. Further, constraints 2 and 3 are simple linear inequalities. Constraint 1 has a slope of $-D_2/D_1$ and a δ_2-intercept of $\lambda\Delta z/D_1$, as shown in Fig. 7.7. Constraint 3 is more interesting, however. The upper bound has a slope of $1 + \Delta z/R$ and a δ_2-

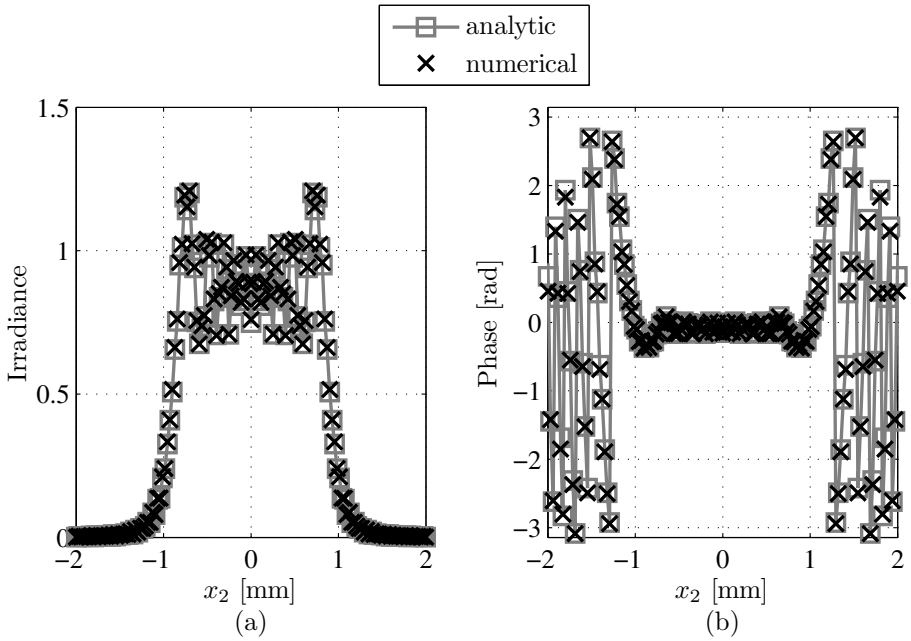

Figure 7.6 Fresnel diffraction from a square aperture, angular-spectrum simulation and analytic: (a) observation-plane irradiance and (b) observation-plane phase.

intercept of $\lambda\Delta z/D_1$. Comparing Fig. 7.7 (a) and (b) with plot (c) shows that if $-D_2/D_1 < 1 + \Delta z/R$, the upper bound on constraint 3 is not a consideration because it has the same δ_2-intercept and a greater slope than constraint 2. The lower bound of constraint 3 has a slope of $1 + \Delta z/R$ and a δ_2-intercept of $-\lambda\Delta z/D_1$. The δ_2-intercept is unphysical, so we disregard it and instead focus on the δ_1 intercept, which is $\lambda\Delta z/\left[D_1\left(1 + \Delta z/R\right)\right]$. Therefore, comparing plots (a) and (b) reveals that when $1 + \Delta z/R < D_2/D_1$, the lower bound of constraint 3 is not a factor.

To summarize the above discussion of constraint 3, when

$$\left|1 + \frac{\Delta z}{R}\right| < \frac{D_2}{D_1}, \tag{7.60}$$

constraint 3 is not a factor. Interestingly, the physical interpretation is that the geometric beam is contained within a region of diameter D_2. This includes diverging source fields and converging source fields that are focused in front of and behind the observation plane.

This analysis of sampling constraints should serve as a guideline for wave-optics simulations, but not as unbreakable rules. The most important lesson from this chapter is that quadratic phase factors, which are ubiquitous in Fourier optics, pose great challenges to numerical evaluation, so simulations must be approached carefully and validated fully. When attempting to simulate a Fourier-optics propagation problem that does not have a known analytic solution, one must consider

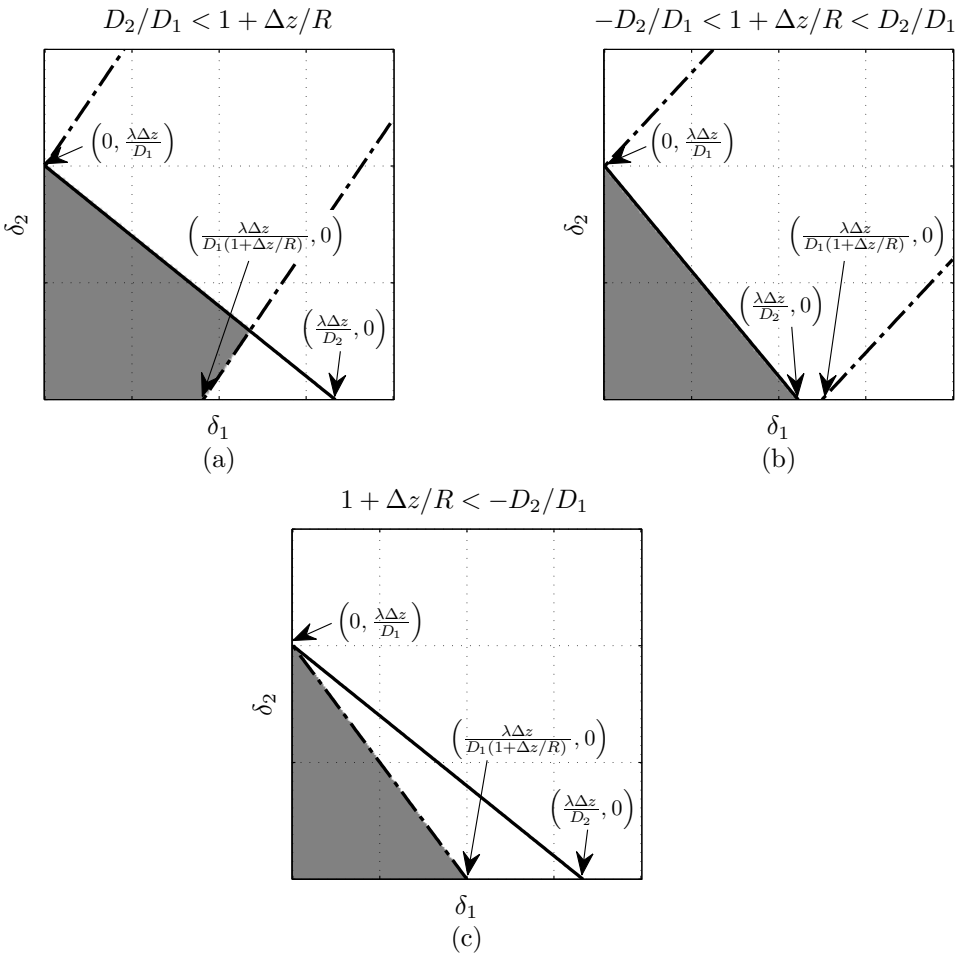

Figure 7.7 General sampling constraints for angular-spectrum propagation.

sampling first as a general guideline for choosing the propagation grids. Then, the accuracy of the simulation setup must be validated through the simulation of a similar problem with a known solution. That is why this book makes such heavy use of the square-aperture propagation problem.

7.4 Problems

1. Consider the signal

$$g(x) = \exp\left(i\pi a^2 x^2\right) \tag{7.61}$$

with $a = 4$ sampled on a grid with $N = 128$ points and $L = 4$ m total grid size. Without performing any FTs, analytically show that the sampled signal has aliasing.

2. Show the sampling diagram for a point source with wavelength of a 1 μm

propagating a distance 100 km to a 2-m-diameter aperture.

3. Show the sampling diagram for a source with a wavelength of 0.5 μm and a diameter of 1 mm propagating a distance 2.0 m to a 2-m-diameter aperture.

4. Modify Listings 7.2 and B.5 to use a converging/diverging source of the form

$$U(x_1, y_1) = \text{rect}\left(\frac{x_1}{D_1}\right) \text{rect}\left(\frac{y_1}{D_1}\right) e^{i\frac{k}{2R}\left(x_1^2 + y_1^2\right)}. \tag{7.62}$$

   (a) Rework the analytic solution for Fresnel diffraction by a square aperture given in Eq. (1.60) to include the diverging/converging wavefront in Eq. (7.62). Just a little algebraic manipulation obtains an analytic result similar to Eq. (1.60), but slightly more general to account for the diverging/converging source. See Ref. 5 for details on the derivation of Eq. (1.60).

   (b) Let $D_1 = 2$ mm, $D_2 = 4$ mm, $\Delta z = 0.1$ m, $\lambda = 1\,\mu$m, and $R = -0.2$ m (just like in the example, but with a converging source). In preparation for carrying out an angular-spectrum simulation, generate plots similar to Fig. 7.5 to show your careful method of picking values for δ_1, δ_2, and N.

   (c) Carry out the simulation, and produce plots of the $y_2 = 0$ slice of the amplitude and phase. Evaluate the analytic result you obtained in part (a) for the given parameters, and include the analytic result on those same plots.

5. Show diagrammatically that Eq. (7.60) means that the geometric beam is contained with a region of diameter D_2. Show the ray diagrams for diverging source fields and converging source fields that are focused in front of and behind the observation plane.

6. Show algebraically that constraint 4 is more restrictive than the combination of constraints 1 and 2.