Lecture Notes 7 Introduction to Fourier Optics

Benjamin Schreyer

University of Maryland, College Park, Maryland

June 23rd 2024

1. LIMITATIONS OF THE SINGLE FOURIER TRANSFORM PROPAGATION

The single Fourier Transform propagation allows the general computation of many linear optical phenomena, especially far field effects. It will be seen that some phenomena do not mesh so well with a constantly expanding simulation window. For example an optical cavity which can consist of a pair of lensing mirrors (pictured in figure 1) can "trap" light of a laser, and keep the width of the beam small even after many reflections off the cavity mirrors, and thus a large propagation distance. Trying to model an approximately constant size laser beam, with an ever larger simulation grid is not easy, because the effective sample density will degrade with each calculation.

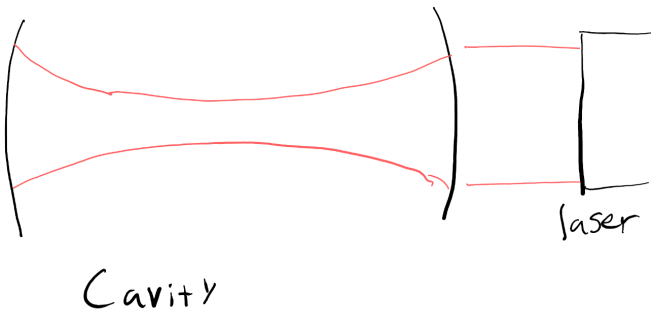

Figure 1. The rough width profile of a laser source shining into a cavity. The laser source can enter the cavity because the mirror is only partially reflective, half silvered.

Imagine doing propagation calculations for such a cavity with Single Fourier Transform Fresnel propagation. One cannot just propagate the beam the total path distance, as the propagation is periodically interrupted when the slightly curved mirrors are computed as a phasing of the amplitude. The field size would become larger and larger, but the focusing cavity mirror walls would keep the light pattern roughly the same size. The detail would become progressively worse, until sampling became insufficient, and the simulation failed. It turns out many optical phenomena, such as the laser cavity, require many linear propagation steps. If each propagation increases the field size this can be a big computational cost.

This is where the transfer function approach to linear optics comes in. It requires instead of a single Fourier transform and scaling, a Fourier transform, a multiplication by the transfer function, and then an inverse Fourier transform. This is double the computation, but as will be seen will maintain the grid size after propagation, a significant advantage in the aforementioned low diffraction scenarios.

Correspondence E-mail: bschrey1@terpmail.umd.edu

2. THE GREEN'S FUNCTION, AGAIN

One should recall the Green's function as the spherical wave. It was discussed in the first lecture, that each point of a directional wavefront emits a spherical wave because of symmetry. Here the light frequency $f = \frac{c}{\lambda}$ is ignored, since the radiation field is always assumed to be in a steady state.

$$S(r) = \frac{1}{r} e^{2\pi i \frac{r}{\lambda}} \tag{1}$$

How does one obtain the field at a further plane, given the field at the plane $A(x, y)$, and the result of a single point of radiation in the point $A(x, y)$? Since the wave equation is linear, the answer is simply to add up all points of the amplitude $A(x, y)$ as if they were point sources, then evaluate this sum of (x', y', z') where z' is constant because the goal is to propagate between planes.

$$A'(x', y') = \psi(x', y', z') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(x, y) S(\sqrt{(x' - x)^2 + (y' - y)^2 + z'^2}) dx dy \tag{2}$$

This may look scary, what is expressed here is a sum over each point of the initial condition wavefront, assuming its resulting field is a spherical wave (This was derived in the first lecture). Now, it is useful to express $S(r)$ as $S(x, y, z)$ to reveal a convolutional structure. ψ the solution for all z' is dropped, because its argument z' will be a constant, the propagation distance L for all use cases discussed.

$$A'(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(x, y) S(x - x', y - y', z') dx dy \tag{3}$$

This shows that the resulting field at constant $z' = L$ is a 2d convolution between the impulse function S and the amplitude at $z = 0$ $A(x, y)$. Since A, S are solutions to the wave equation, the variables x, y separable, and the multiple dimensions of convolution is not a concern. What is useful about this expression is that convolution may be expressed as multiplication in the Fourier domain. To propagate between planes, one could directly do this integral, for each sampled point of $A'(x', y')$ However consider this would be extremely slow, for each point N^2 samples contribute to the double integral (assume a square sampling grid), and this sum of N^2 must be taken N^2 times assuming the number of samples is the same in the sampled result plane $A'(x', y')$. The total number of additions is N^4, which will quickly become slow even on today's computers. It is hoped that the Fourier domain expression will be more computationally efficient.

The two dimensional convolution of S and A is now converted into the Fourier domain.

$$FT_{x',y'}\{A'(x', y')\} \propto FT_{x,y}\{A(x, y)\} \cdot FT_{x,y}\{S(x, y, L)\} \tag{4}$$

To recover $A'(x', y')$, invert the Fourier transform.

$$A'(x', y') \propto FT^{-1}_{x',y'}\{FT_{x,y}\{A(x, y)\} \cdot FT_{x,y}\{S(x, y, L)\}\} \tag{5}$$

Now heuristically, consider making this algorithm numerical. There are three Fourier transform equivalent (in terms of complexity) computations, and a multiplication point wise. The pointwise multiplication is N^2 multiplications, and the 2d Fourier transforms are $N^2 logN$. This looks much faster!

First, the term $FT_{x,y}\{S(x, y, L)\}$ must be specifically considered. Currently no approximation is being made on the spherical wave S in terms of coordinates x, y, z or x, y, L. It is not strictly necessary to make such an approximation, but to make error considerations easier, the Fresnel approximation will again be made for $S(x, y, L)$. The first step is to note the circular symmetry of the half spherical wave S, so that $S(x, y, L)$ can be written $S(r, L)$. This is explicitly the equation first given (6).

Before the fresnel approximation, it is assumed the factor $\frac{1}{r}$ varies negligibly with x, y compared to the effect of phase variation due to x, y on the diffraction calculation, so it is repalced with $\frac{1}{L}$

$$S(r) \approxeq \frac{1}{L} e^{2\pi i \frac{r}{\lambda}} \tag{6}$$

Then to execute the Fresnel approximation $r = \sqrt{(x-x')^2 + (y-y')^2 + L^2}$ must be treated as in lecture 4. It is useful to take the variable $R_{circ}^2 = (x-x')^2 + (y-y')^2$ which is allowable form the circular symmetry of S in the x, y plane and makes the manipulations of approximation in the variable R_{circ}. The result follows of the same approximation (Fresnel) made in 1d in lecture 4.

$$r \approxeq L + \frac{1}{2L}(R_{circ}^2) = L + \frac{1}{2L}[(x-x')^2 + (y-y')^2] \tag{7}$$

With this approximation the propagation, known as the transfer function method may be simplified.

$$A'(x', y') \propto FT_{x',y'}^{-1}\{FT_{x,y}\{A(x,y)\} \cdot FT_{x,y}\{\frac{1}{L} e^{\frac{2\pi i}{\lambda} L + \frac{1}{2L}[x^2+y^2]}\}\} \tag{8}$$

This does not look simpler, but that is only because previously alot of the expression was wrapped up in S. Note the Fourier transform that is not on $A(x, y)$ does not depend on any initial condition, so it could be calculated analytically once, or numerically, which will have different implications on propagation error.

Since L is fixed, it is sometimes appropriate to drop the term L, since global phase may not be a concern in simple propagation, where the beam does not reflect back onto itself.

3. NUMERICAL FORMULATION

Before diving into code, a proof is provided for the fact that a discrete circular convolution may be computed by three Fourier transforms and a multiplication. For large grids, doing three Fourier transforms is much faster than directly computing a convolution. This is also true for the case of continuous Fourier transform, for which the proof is very similar. The proof for 2d grids, or functions of two continuous variables is not discussed, but quickly follows the 1d case, since Fourier transforms on different variables may be separated by multiplication.

This straightforward proof is provided on wikipedia, and explained here.

First write down the discrete Fourier transform of a circular convolution explicitly.

$$DFT\{A \circledast B\}_k = \sum_{n=0}^{N-1} (\sum_{m=0}^{N-1} A_m B_{n-m \ (mod \ N)}) e^{-\frac{i2\pi kn}{N}} \tag{9}$$

By manipulating the exponential, and factoring out from sums:

$$DFT\{A \circledast B\}_k = \sum_{m=0}^{N-1} A_m e^{-\frac{i2\pi km}{N}} (\sum_{n=0}^{N-1} B_{n-m \ (mod \ N)} e^{-\frac{i2\pi k(n-m)}{N}}) \tag{10}$$

Because of the modulo on the indexing of B, the term in parenthesis is a discrete Fourier transform of B. The other term is likewise a discrete Fourier transform of A.

$$DFT\{A \circledast B\}_k = DFT\{A\}_k \cdot DFT\{B\}_k \tag{11}$$

Finally, the circular convolution of A and B can be found by an inverse discrete Fourier transform.

$$(A \circledast B)_i = DFT^{-1}\{DFT\{A\} \cdot DFT\{B\}\}_i \tag{12}$$

The next step is to numerically formulate the transformation from A to A' under free space propagation. The analytic Fourier transform integrals are not easy for arbitrary cases, but with fast enough computers approximations of them are not unreasonable to calculate.

```
#This could be optimized by passing in a transfer function so it
#doesnt need to be computed every time
def propagate_fresnel_transfer_2d(A, coordinates, photon_lambda, L):
#d is the delta of the grid, which is extracted from the provided coordinates
#Generate the 2d frequency space coordates that form the transfer function
#The frequency range goes from 0 (constant), to 1/dx
#np.abs(coordinates[0][0][0] - coordinates[1][1][0]) is dx
  fs_2d=generate_2d_coordinates(len(coordinates[0,:,0]),(1/np.abs(coordinates[0][0][0]
    - coordinates[1][1][0])))


#Analytically calculated fresnel transfer function, by Fourier transform of the Green'
    s
#funciton

  H_fresnel = np.fft.fftshift(np.exp(1.0j * 2 * np.pi / photon_lambda * L) * np.exp
    (-1.0j * np.pi*photon_lambda * L * np.square(distances_2d(fs_2d))))

  Ap = np.fft.ifftshift(np.fft.ifft2(np.fft.fft2(np.fft.fftshift(A)) * H_fresnel))
  Ap /= np.sqrt(np.sum(np.abs(np.square(Ap))))
  return Ap
```

Now it is appropriate to do a test calculation with the propagation. Code is already written to produce Gaussian laser beams and apply lenses to them to make them focus. While this is the chosen example, obviously the power of this numerical computation is that we may choose to deviate from ideal Gaussian beams or lensing, and still get a valid result, which is much more difficult with analytic computation.

The first order of business is to generate initial conditions for a Gaussian laser beam that has passed through a thin lens. A diagram of such a system is provided, propagation code is needed only after the lens, since the lens is assumed to be a thin lens.

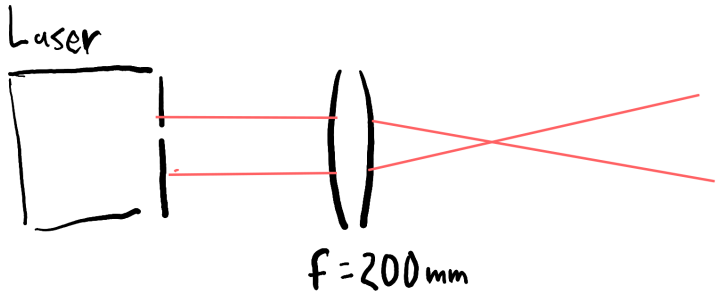

Figure 2. A laser beam enters a thin lens.

```
#Do an example calculation with then new transfer function
#a 500nm Gaussian laser beam going through a 200mm lens
coordinates_example = generate_2d_coordinates(256, 8 * millimeter)
amplitude_example = gaussian_2d(coordinates_example, 0.3 * millimeter) * lens_2d(
    coordinates_example, 200 * millimeter,500 * nanometer)
```

Let us plot the initial conditions.

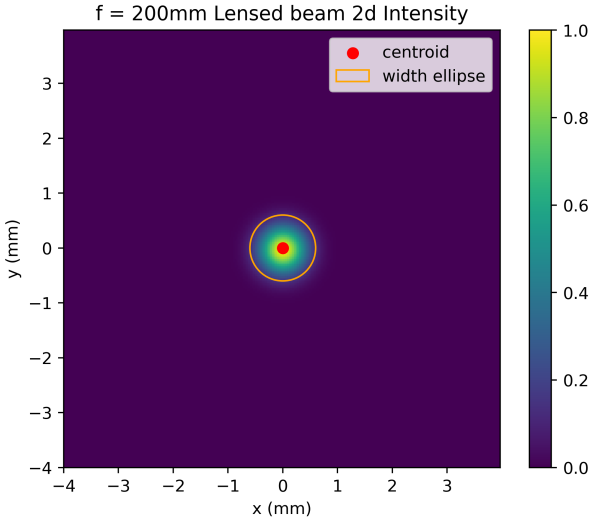

Figure 3. The initial intensity of the Gaussian beam.

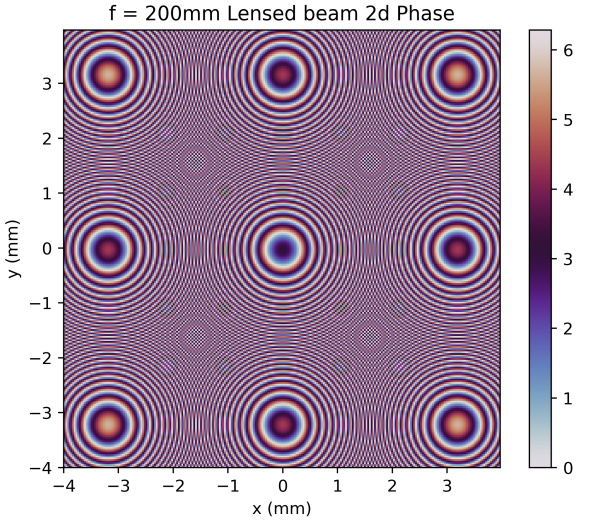

Figure 4. The initial lensed phasing of the Gaussian beam.

Now apply the propagation $100mm$, since the focal length of the lens is $200mm$, it is expected that the beam should be roughly half its size. No new coordinates are created, since the transfer function approach maintains the grid size and sample count.

```
#Calculate the new field at 100mm, this function doesnt need new coordinates, since
    the
#simulation window stays the same size
amplitude_prime_example = propagate_fresnel_transfer_2d(amplitude_example,
    coordinates_example, 500 * nanometer, 100 * millimeter)
```

The resulting amplitude is plotted, and matches the expectation for lensing of the ideal Gaussian beam.

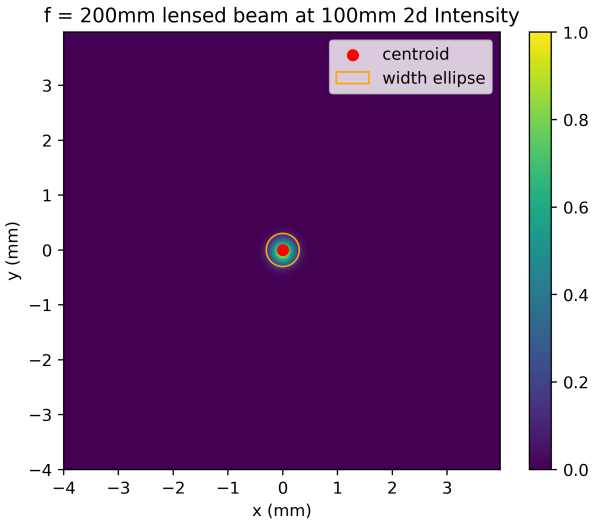

Figure 5. Propagated intensity of the Gaussian beam.

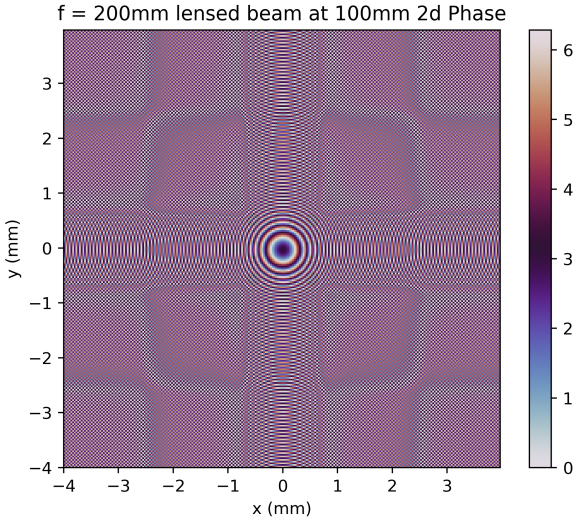

Figure 6. Propagated phasing of the Gaussian beam.

There are clearly artifacts in the phase, since a square pattern is appearing, despite nothing regarding the initial conditions or ideal propagation has the properties of a square. This results from the square simulation

window, and is an artifact. Since the intensity is low in the regions where this error occurs, there is no effect on the accuracy for our purposes. Such square artifacts will also be visible in the intensity for further propagations, and one solution is to increase window size and sampling.

4. PARABOLIC PROPAGATION

Now with this new algorithm in hand, an important property of beam propagation will be shown (not proven) via the results of simulations. When a beam propagates through a linear homogeneous medium, width of the beam will follow a parabolic profile through space along the propagation axis. In this case the width of the beam is rigorously qualified, as previously described as follows:

$$W_{x/y}^2 \propto < (x/y - \mu_{x/y})^2 >_A \tag{13}$$

Where A is the field amplitude along a plane perpendicular to he beam propagation.

The width will vary over the z coordinate according to the following model, even under lensing.

$$W(z) = \sqrt{W_0^2 + M^4[\frac{\lambda}{\pi W_0}(z - z_0)]^2} \tag{14}$$

Explaining the analytic calculations that yield this nice result for the beam width is above my pay-grade. Instead simulations will be undertaken, and compared to this model. $W(z)$ may be the width taken with respect to any transverse axis. The parameter z_0 is the focal point, or where the beam is smallest. The minimum width, or spot size is W_0, and $M^2 = \sqrt{M^4}$ is the beam quality, which in rough terms determines the ratio of the size of the input beam to the spot size. The lower the parameter M^2, the better the beam may be approximated by a Gaussian beam in its propagation. An intuitive part of this model is that for large $|z|$, the width diverges linearly, which corresponds to the inverse square law scaling expected. This linear divergence should not be suprising given the properties of the Fraunhofer far field single Fourier transform propagation which should match the transfer function method, and exhibits such scaling.

With two initial conditions the above model will be shown graphically to hold. The figures 2, 3 give the first initial condition that will be compared to the model (14). In fact since these initial conditions are ideal (Gaussian in intensity, and phase of a thin lens), the parameter M^2 will be 1 within numerical error. The second set of initial conditions will be modified, to be a non-ideal laser beam, and follow in plots. To see each propagated amplitude plotted, running the code will plot all 20 planes for each of the two initial conditions. Importantly, the approach to such many plane propagations should not be to stack propagations on top of each other, this would cause error to build up unnecessarily. Instead do single propagations to each desired distance from the initial condition. A general rule of thumb is to minimize the number of recursive propagations that are applied to reach the result, only being forced to make such recursive propagations when the path is not free, such as a lens, or other optical component.

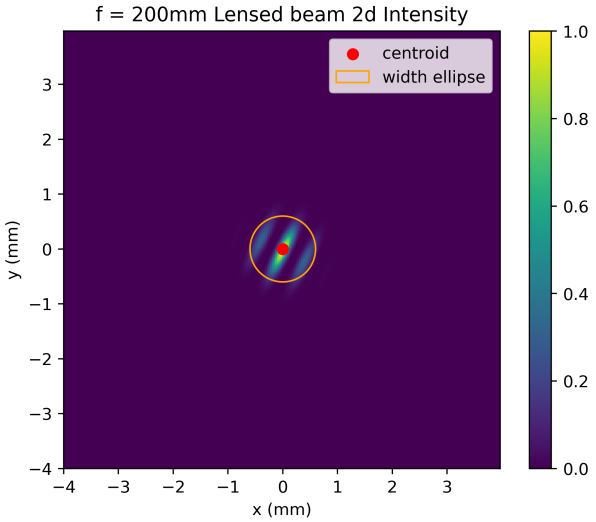

Figure 7. The initial intensity of the non-ideal beam.

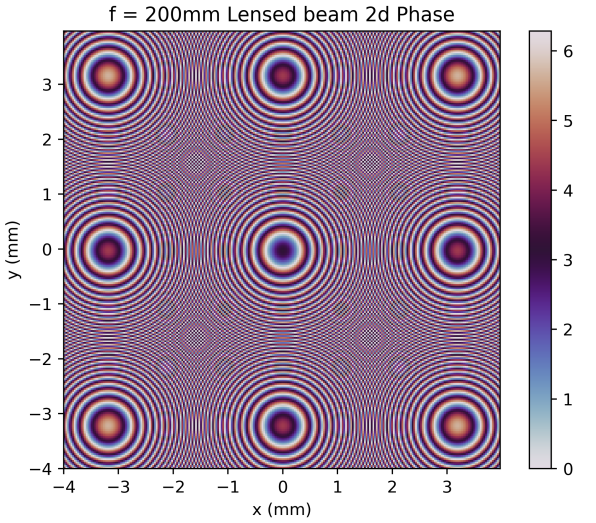

Figure 8. The initial lensed phasing of the non-ideal beam.

The phasing is left alone, and what makes this specific beam non-ideal is its intensity profile, modulated with a cosine squared. The beam could equally well be made non-ideal by giving it phasing, so long as this phasing is not lensing.

The ideal, and non-ideal beam are propagated over a range of z coordinates, reaching over $200mm$ (the focal length of the lens), to reveal the parabolic shape of the beam pattern.

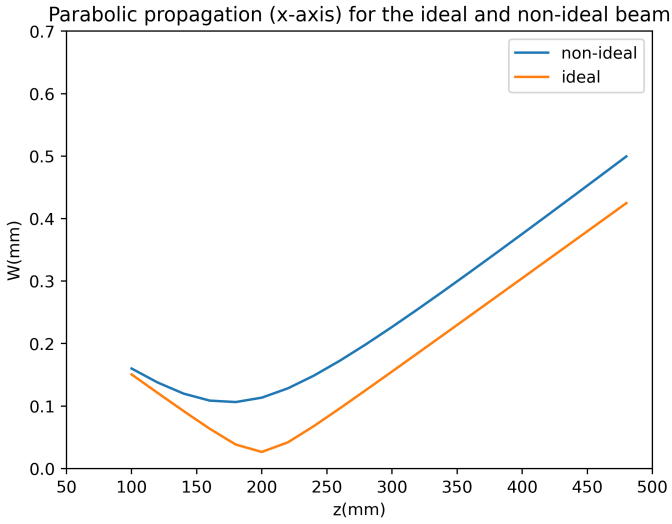

Figure 9. Widths comparison for the x axis. This exhibits greater deviation in the width profile from ideal, than the y width, since the modulation of the intensity was mostly in the x axis.

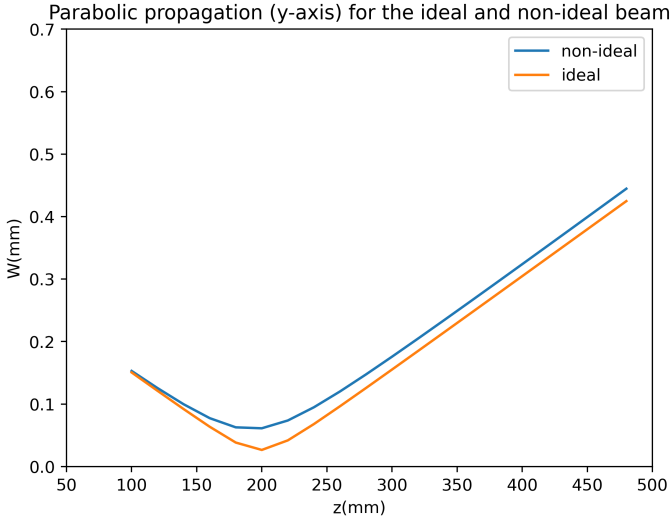

Figure 10. Widths comparison for the y axis.

This demonstrated parabolic width profile has immediate practical application. In the lab, so long as one can collimate their laser beam, even if it is not perfectly ideal, its propagation and lensing characteristic will match an ideal laser beam, but will not focus to a point so small as the ideal beam. It will be a homework assignment to confirm the computations actually fit (by curve fitting) to the parabolic profile.

5. ERROR CONSIDERATIONS

5.1 Error general to transfer function approach

The transfer function approach is much faster than direct integration, and often provides more convenient coordinates than a single Fourier transform propagation. The speed up over direct integration requires circular

convolution, which suffers from artifacts whenever significant beam intensity occurs near the edge of the grid. Because the convolved Green's function is wide itself, the intensity need not reach the edge of the grid, just close enough for the circular convolution of the Green's function and EM field amplitude to have significant periodic overlap.

To illustrate such an error a reasonably nice beam will be reflected off of a tilted mirror, so that it propagates towards the edge of the simulation grid. The beam will now be focusing and additionally traveling towards the edge of the grid. In a physical case that such a propagation exactly occurs, consider a laser beam passing through a thin slanted glass plate.

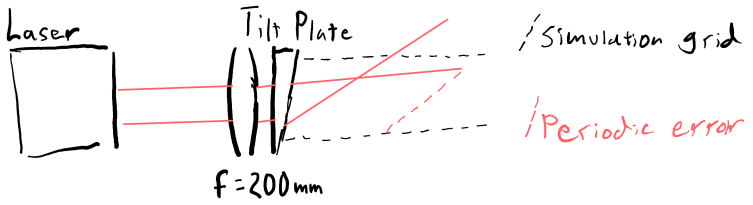

Figure 11. A laser beam enters a thin lens, then a thin tilt plate.

The initial intensity is unchanged, and looks like figure 6. The phase is changed, since now the beam not only passes a lens, but also a tilted mirror.

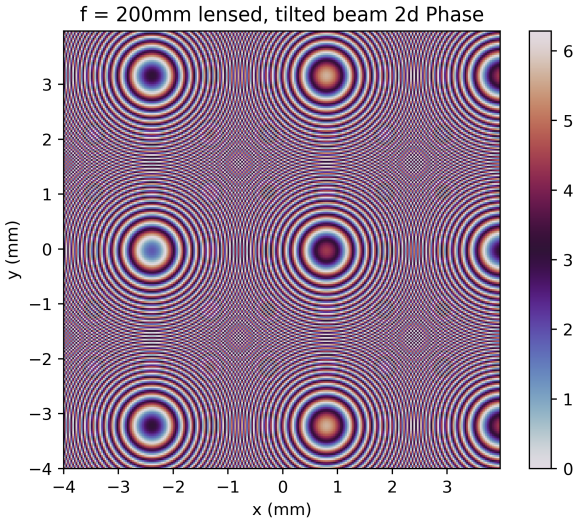

Figure 12. The initial condition phase for the tilted, lensed, beam

One should compare the titled and untilted phasing patterns for the lensed beam. It is not what may be immediately expected, given a tilted phasing pattern alone looks like a stripe pattern.

The simulation initially is correct, the beam focuses, while also being deflected by the tilted mirror.

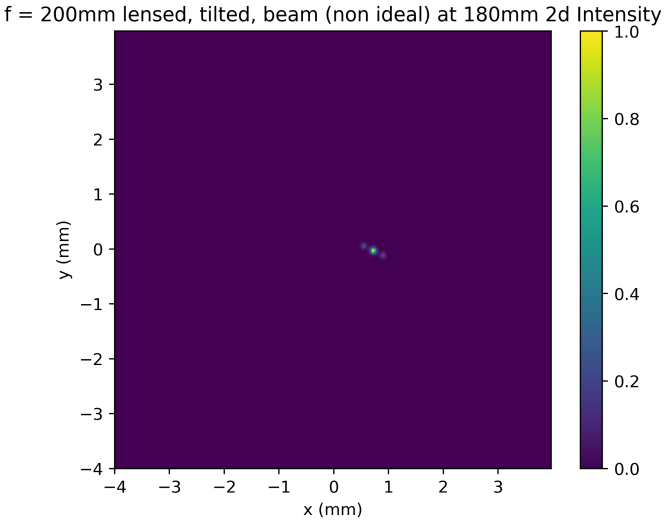

Figure 13. The focal point with tilt, in the same location in z, but now displaced in the transverse axes due to tilt.

Further propagation than the focus is also not an issue.

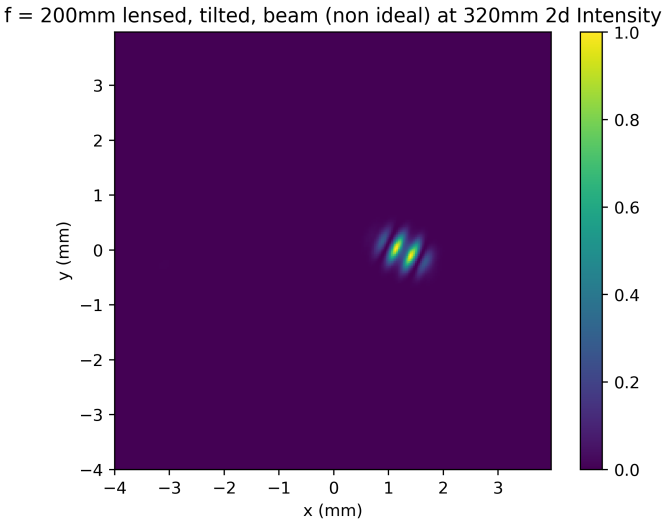

Figure 14. Expanding intensity pattern of the beam after the focus.

The issue occurs when the beam travels further and reaches close to the boundary. The error first manifests as high frequency, asymmetric, features that are not expected in the propagation.

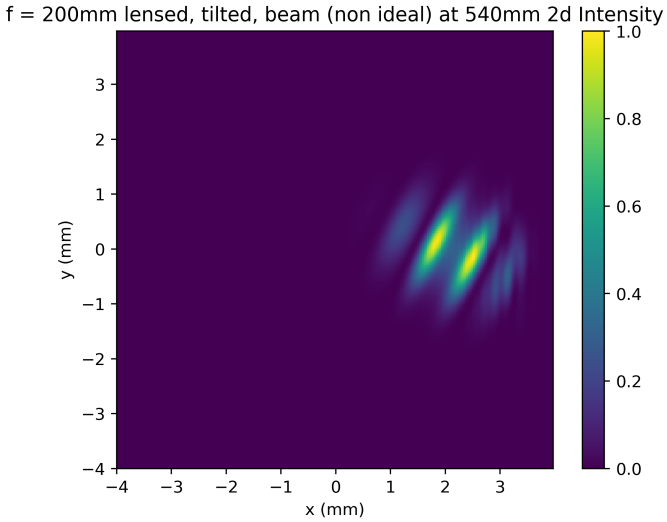

Figure 15. First sign of error in the propagation calculation.

Eventually the boundary effects are seen in their periodic form, where despite the beam traveling away from the left side of the image, in error, intensity appears there in calculation.

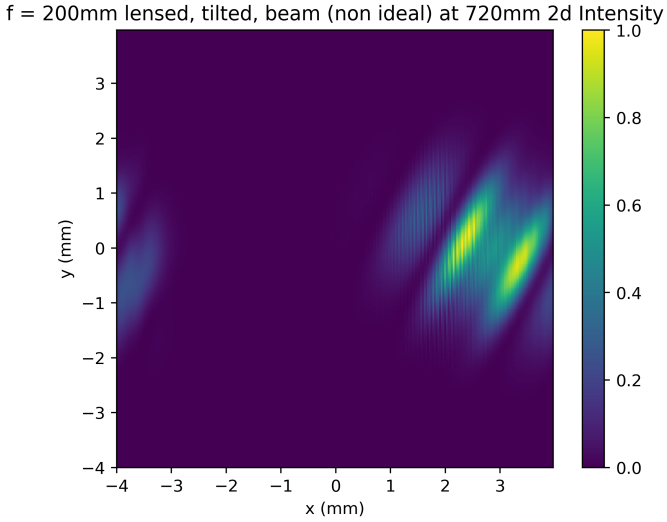

Figure 16. Serious error due to periodic continuation.

Propagating even further the error is more pronounced, with half of the intensity rolling over to the left region of the grid where it should never reach. The simplest approach to avoid such errors is a big grid in dimension. When this is unfeasible, absorbant regions may be placed near the boundary to damp the intensity that would have rolled over.

Another significant general error is unsurprisingly sampling. The finest features of the simulation must be captured by the grid, by achieving at least Nyquist sampling. Sampling concerns are not only of the amplitude, but also the transfer function, since the transfer function, or greens function (depending on implementation) is

also sampled by the simulation grid. A very short focal length lens could for example be aliased and result in a failed simulation. This example will be shown below, this time with an ideal beam. The same ideal beam and lens physical parameters will be taken as was first shown, this time, with less grid sample points.

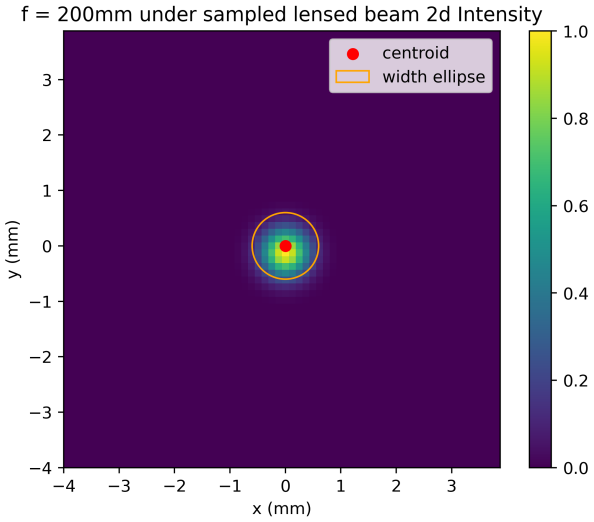

Figure 17. The ideal Gaussian intensity, this time with lower sampling.

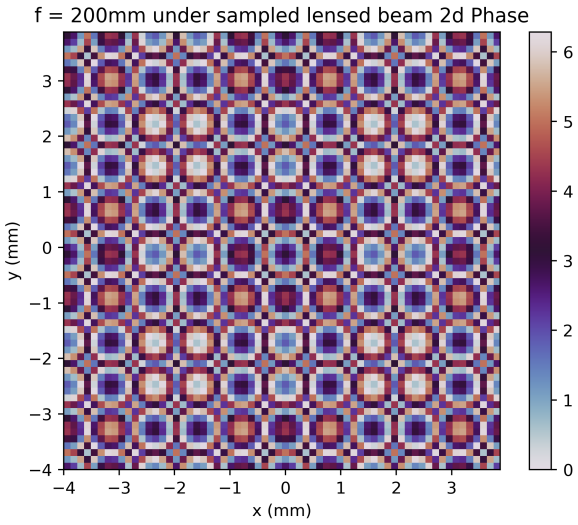

Figure 18. Under sampled lens phasing, the pattern is not captured by so few grid points.

The under sampling is only clear in the phase, which should resemble figure 3. Under sampling causes the failure to capture the lensing phase pattern. Next is to observe how this affects the propagation. The first propagation done above is repeated, so that it may be compared to the sufficiently sampled result.

The incorrectly sampled lensing pattern causes failure of simulation, as can be seen by the fact that the Gaussian beam loses its shape (Gausian intensity beams remain Gaussian shaped under lensing), and the width

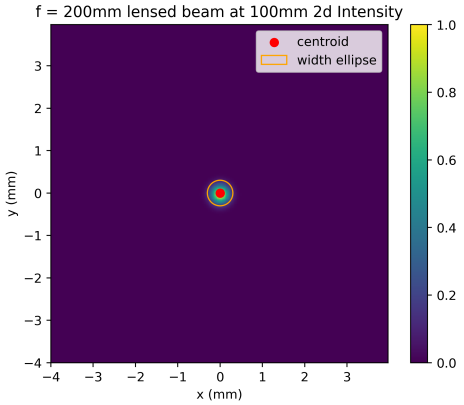

Figure 19. Sufficiently sampled propagation of the lensed ideal beam

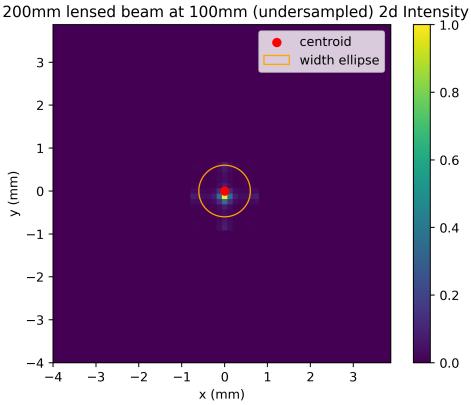

Figure 20. Under sampled propagation of the lensed ideal beam

Figure 21. Contrast between two propagation results, one under sampled. Lensing does not have the intended effect if the sampling fails to resolve the lensing phase.

does not decrease significantly. Note that the lensing pattern sampled sufficiently, is still not sampled perfectly. This can be seen by the repeating circular forms. If perfectly sampled, the pattern should just be concentric annuli. This aliasing present even in the more sampled phase is acceptable however, since negligible intensity falls upon regions where this aliasing occurs.

5.2 Error specific to the analytically Fourier transformed Fresnel transfer function

The Fresnel approximation must hold to make any calculation. Outside of the Fresnel approximation the Rayleigh-Sommerfeld transfer function can be used, which applies to shorter propagation distances.

The choice of whether to numerically Fourier transform the Green's function, or apply an analytical calculation to reach the transfer function also matters. Voelz describes the differences in the errors resulting from the methods well in his book *Computational Fourier Optics*, chapter five. The usual choice is to use the analytically calculated transfer function.

6. CITATIONS/RESOURCES

All the code to generate plots is made available with this PDF.