

From Ray Geometrical to Wave Diffraction Imaging

Virendra N. Mahajan

The University of Arizona, USA

It is quite common to start the study of geometrical optics with Fermat's principle and derive its laws of rectilinear propagation, refraction, and reflection from it in 2D. Then, using the paraxial or the small-angle approximation of the rays, launch into the imaging equations for refraction and reflection. Sometimes, instead of deriving the equations for reflection independently, they are obtained from those of refraction by giving a value of -1 for the refractive index and replacing the angle of refraction with a minus value as the angle of reflection. As soon as these equations are established, the fact that they have been obtained in the small-angle approximation is forgotten. Has anyone who has obtained the image of an object by a graphical construction wondered if the rays used in the construction actually make small angles? In fact, the small-angle approximation is never quantified, except that it is imposed by replacing the sines and tangents of the angles, however large, by the angles (in radians). Even the curved refracting and/or reflecting surfaces are replaced by their paraxial counterparts in the name of *tangent planes*. The image of an object obtained in this manner is called a Gaussian image. Unfortunately, when done in this manner, Gauss, who introduced the *paraxial* or *Gaussian* approximation, does not get credit for how he came up with the idea of his approximation.

Rays do not travel only in a plane. Hence, it is important to consider propagation of skew rays and derive the laws of geometrical optics in 3D.^{1,2} When this is done, we realize that, to trace a ray exactly from one point to another on an imaging surface, the transverse coordinates of the surface point depend on the direction cosines of the ray and the distance between the two. However, the distance itself depends on the coordinates of the incident point. Hence, the two equations are coupled and must be solved simultaneously. The ray is refracted or reflected according to the law of refraction or reflection. In an imaging system

consisting of multiple refracting and/or reflecting surfaces, where a ray originates at a point in the object plane, coupled equations are solved when the ray leaves one surface and meets another, finally reaching the image plane. The refracted and reflected rays lie in the plane of incidence, i.e., the plane containing the incident ray and the normal to the refracting or reflecting surface at the point of incidence. In the small-angle approximation, projections of a *skew ray* in two orthogonal planes propagate independently of each other. A ray in the tangential plane, for example, remains in that plane after refraction or reflection. Because of the rotational symmetry of an imaging system, the consequence is that we need to trace rays only in one of these planes. It is a common practice to trace rays in the tangential plane of an imaging system. Because the sine of an angle is replaced by the angle itself, which is a *first-order approximation*, ray tracing in this approximation is called *first-order optics*, and the process of determining the image in this manner, regardless of the magnitude of the angles and sizes, is called *Gaussian optics*.

Gaussian imaging is used to determine the image location and its size in terms of the object location and its size. It depends only on the vertex radius of curvature of a surface. Thus, the Gaussian image formed by a conic surface of some vertex radius of curvature is the same as that formed by a spherical surface of the same radius of curvature. How does the difference in the two surfaces manifest itself? While the two Gaussian images are the same, their actual images and qualities are not. In Gaussian optics, the image is aberration free; in reality, that is generally not the case. The rays from a point object incident on an optical system do not all pass through its Gaussian image point. Their distribution is called a *spot diagram*, and their separations in the image plane from the Gaussian image point are called *transverse ray aberrations*. The spherical wavefront incident from the point object is not refracted by it as a spherical wavefront. The deviations of the wavefront along the exact rays from a spherical surface with its center of

curvature at the Gaussian image point and passing through the center of the exit pupil of the imaging system are called *wave aberrations*. Such deviations are different for spherical and paraboloidal refracting surfaces. A lens designer designs an optical imaging system with multiple surfaces to control the aberrations to required tolerances over a certain field of view of an object.

For a rotationally symmetric imaging system, the aberrations at a point (r, θ) on its pupil for the imaging of a point object at a height h from its optical axis depend on the integral powers of three *rotational invariants*: h^2 , r^2 , and $hr\cos\theta$.¹ The order of such aberrations is even. This is how the fourth-order, or primary, or *Seidel* wave aberrations r^4 , $hr^3\cos\theta$, $h^2r^2\cos^2\theta$, h^2r^2 , and $h^3r\cos\theta$ come about. These are spherical aberration, coma, astigmatism, field curvature, and distortion, respectively. The higher-order aberrations can be similarly written. Since the ray aberrations represent the gradients of the wave aberrations, their order is one less than that of a corresponding wave aberration. Thus, Seidel ray aberrations, for example, are of the third order. If the image is observed in a plane that is displaced along the optical axis from the Gaussian image plane, a defocus wave aberration varying as r^2 is introduced. Moreover, if the aberration is considered with respect to a point in the Gaussian image plane other than the Gaussian image point, a wavefront tilt aberration varying as $r\cos\theta$ is introduced. These are referred to as classical aberrations. A wave aberration of a certain order can be combined with those of lower orders to minimize its variance across the pupil. These aberrations are called *balanced aberrations* and are represented by *Zernike polynomials*. These polynomials are orthogonal to each other across a circular pupil. A wave aberration can also be balanced so that the variance of the corresponding ray aberrations, or the spot size, is minimized. However, such wave aberrations are not orthogonal to each other, but their gradients are.

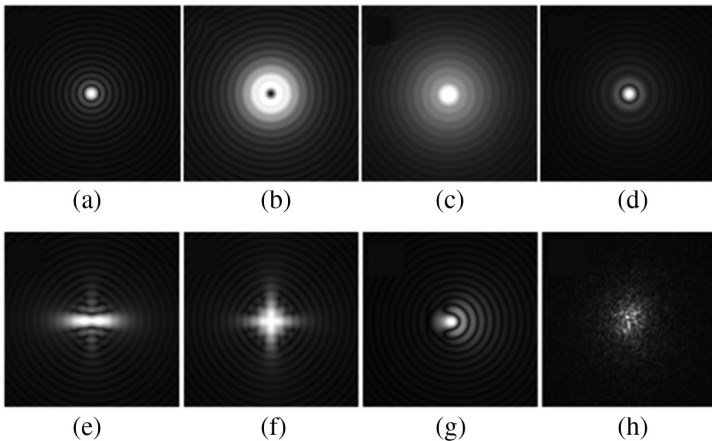
In principle, a system can be designed to yield the image of a point object to be small to some prescribed tolerance.

Even if the rays transmitted by the system intersect at the Gaussian image point, the observed image, instead of being a point, is a light distribution, called the *Airy pattern*.^{1,3} Owing to the circular symmetry of the pupil, the pattern consists of a bright circular spot, called the *Airy disc*, surrounded by alternating dark and bright circular diffraction rings of decreasing brightness. The Airy disc contains 83.8% of the total amount of light. Its radius is $1.22\lambda F$, where λ is the wavelength of light, and F is the focal ratio (distance between the pupil and image planes divided by diameter of the pupil) of the image-forming light cone. The Airy pattern also represents the squared modulus of the Fourier transform of the uniform distribution of light across the circular pupil because of the diffraction of light. The brightest point lies at the Gaussian image point. This point is where the center of the pattern lies and is equidistant from the points on the spherical wavefront; it is also where Huygens' secondary wavelets interfere constructively. When aberrations are present in a system, the points on the wavefront exiting from its pupil are not equidistant from the center of the pattern, and secondary wavelets interfere partially destructively, thus resulting in a reduction of the brightness at this point. When the defocus wave aberration, varying as r^2 , is an integral number of waves, secondary wavelets interfere destructively at the center, and the irradiance reduces to zero.

For small aberrations, the relative brightness at the center, called the *Strehl ratio*, is given approximately by $\exp(-\sigma^2_\phi)$, where σ^2_ϕ is the variance of the phase aberration.^{1,3} The fabrication tolerance for a single mirror for a Strehl ratio of 0.8, for example, is approximately $\lambda/30$ (where we have doubled the surface error to obtain the wavefront error because of reflection of light). The irradiance distribution of the image of a point object is called the *point-spread function*, and its Fourier transform is called the *optical transfer function (OTF)*.³ The OTF is a complex function, and the integral of its real part yields the Strehl ratio. Its modulus is called the *modulation transfer function (MTF)*.

In the presence of aberrations, whereas the spot diagram grows linearly in size, the distribution of light in the observed image changes without such increase, as shown in the figure.³ Although the spot diagram does not represent the true distribution of light in the image, lens designers use it as a convenient tool in the early stages of a design. As the spot diagram becomes small in the range of the Airy disc, they resort to diffraction calculations of the image to ascertain the image quality. Two commonly used image quality criteria are the fraction of light on a pixel in the image plane and the MTF of the system.

While ray geometrical optics determines the location and size of the image of an object, its quality is determined by wave diffraction optics, and the aberrations provide a bridge between the two. It is suggested that, in a course on geometrical optics, the paraxial approximation be derived from 3D ray tracing and the Airy pattern be discussed at least at an elementary level, including how it is impacted by aberrations.



Aberrated PSF: (a) Airy pattern, (b) defocus, (c) spherical, (d) balanced spherical, (e) astigmatism, (f) balanced astigmatism, (g) coma, and (h) random aberration.

References

1. V. N. Mahajan, *Fundamentals of Geometrical Optics*, SPIE Press (2014) [doi: 10.1117/3.1002529].
2. M. V. Klein and T. E. Furtak, *Optics*, John Wiley & Sons (1988).
3. V. N. Mahajan, *Optical Imaging and Aberrations, Part II: Wave Diffraction Optics*, SPIE Press (2001); 2nd ed. (2011) [doi: 10.1117/3.898443].