Toward an exact

THEORY OF LIGHTBEAMS

Nicholas Wheeler, Reed College Physics Department
March 2002

Introduction. The "plane wave" is an abstraction of such indispensable utility that in many contexts it would seem picayune to remark that such a thing is never encountered in Nature. It is from the theory of plane waves that we inherit—and to which we look to illustrate the meaning of—much of the language (frequency and wavelength, direction of propagation, polarization, intensity) that we use to classify/discuss waves-in-general—solutions-in-general of the diverse linear wave equations encountered in a great variety of subject areas.

But when we walk into the laboratory, with heads full of such "plane wave language and imagery," we encounter waves of *finite* temporal duration, of *finite* spatial extent, waves that transport *finitely* much energy/momentum/angular momentum, waves that are in this or that sense "normalizable." So habituated are we to the practice of construing the real things before us as "assembled populations of spooks" (wavepackets) that we are heedless of odd light which our practice casts upon any claim that our science is rooted in "reductionism."

My intention here is not to argue that one *should* pay heed to that small philosophical conundrum (much less to suggest how the point might be resolved!), but to engage a bit in the very practice that calls the conundrum into being.

Recently I developed (renewed) interest in these questions: *How does one lend substance to the frequently-encountered claim that "angular momentum is transported in the fringes of a lightbeam"? And how is that angular momentum* *related to the polarizational state of the beam? Do Stokes parameters—which serve so elegantly to describe the polarizational states of plane waves—serve as well to describe the polarization of laterally-confined optical beams?* To approach such questions I had to decorate the standard "scalar theory of Gaussian beams" with $\boldsymbol{E}$ and $\boldsymbol{B}$ vectors, since those are basic to any account of the mechanical properties of electromagnetic fields: I had, in short, to develop a theory of "beams as electromagnetic objects," a theory of "lightbeams."[1] This I was able to do in a degree of detail sufficient to my immediate needs,[2] but the theory proceeds from a seemingly innocuous approximation which it is my present intention to try to remove, and in its present (sketchy) state of development fails to address certain formal/physical fine points, which I hope here to do.

In order more clearly to separate points of principle from the clutter of details I will look first to the 2-dimensional theory, then to the complications introduced by a third space-dimension. But a preliminary word about the (degenerate) one-dimensional theory may be in order:

In one dimension (which is to say: in 2-dimensional spacetime) the "beam problem" is, for obvious reasons, trivial, and the plane wave concept empty. Looking for monochromatic solutions

$$\varphi(t,z) = e^{i\omega t}\cdot\phi(z)$$

of the wave equation

$$\square\varphi = 0 \quad\text{with}\quad \square \equiv \big(\tfrac{1}{c}\tfrac{\partial}{\partial t}\big)^2 - \big(\tfrac{\partial}{\partial z}\big)^2$$

we encounter the "one-dimensional Helmholtz equation"

$$\big(\partial_z^2 + k^2\big)\phi = 0 \quad\text{with}\quad k^2 = (\omega/c)^2$$

Immediately $\phi(z) \sim e^{\pm ikz}$, so we don't have much to work with:

$$\varphi(t,z) = \text{linear combination of } e^{i(\omega t - kz)} \text{ and } e^{i(\omega t + kz)}$$

From this material we can construct running waves $\varphi \sim \cos(\omega t - kz)$ and standing waves (in which we have no present interest), but to construct waves of finite duration or—which comes to the same thing—of limited spatial reach we must *abandon the monochromaticity assumption*. Note also that it is, even in such an impoverished context, entirely possible to speak of "vectorial waves" $\varphi \sim \boldsymbol{A}\cos(\omega t - kz)$, the general point being that $\boldsymbol{A}$ need not live in spacetime, but can inhabit a vector space of arbitrary dimension . . . and that it is, in particular, entirely possible to speak of the "polarization of a transverse wave" (as one would find it entirely natural to do in a "theory of guitar strings.")

---

[1] Webster recognizes "lightbulb" and "lighthouse," but insists upon "light beam." The eccentric usage which I allow myself is responsive, I guess, to the same Germanic sensibility that prefers "wavepacket" over "wave packet": the sense that unitary entities are entitled to single-word names.

[2] See Chapter 5, §5 in CLASSICAL ELECTRODYNAMICS (2002).

2-DIMENSIONAL THEORY

1. Essentials of the Gaussian theory of scalar beams. In 3-dimensional spacetime the scalar wave equation reads

$$\Big\{\big(\tfrac{1}{c}\tfrac{\partial}{\partial t}\big)^2 - \big(\tfrac{\partial}{\partial x}\big)^2 - \big(\tfrac{\partial}{\partial z}\big)^2\Big\}\varphi(t,x,z) = 0 \tag{1}$$

The monochromaticity assumption $\varphi(t,x,z) = e^{i\omega t}\cdot\phi(x,z)$ directs our attention to the Helmholtz equation

$$\Big\{\big(\tfrac{\partial}{\partial x}\big)^2 + \big(\tfrac{\partial}{\partial z}\big)^2 + \kappa^2\Big\}\phi(x,z) = 0 \quad \text{with} \quad \kappa^2 = (\omega/c)^2 \tag{2}$$

Exponential solutions are of the form

$$\phi(x,z) \sim e^{-i(px+kz)} \quad \text{with} \quad p^2 + k^2 = \kappa^2 \tag{3}$$

but the "theory of Gaussian beams" proceeds from a quest for solutions of the specialized form

$$\phi(x,z) = e^{-i\kappa z}\cdot\psi(x,z) \tag{4}$$

which requires

$$\Big\{\big(\tfrac{\partial}{\partial x}\big)^2 + \big(\tfrac{\partial}{\partial z}\big)^2\Big\}\psi(x,z) = 2i\kappa\tfrac{\partial}{\partial z}\psi(x,z) \tag{5}$$

The theory derives its distinctive coloration from an assumption that the red term can be abandoned. The surviving equation

$$\Big\{\big(\tfrac{\partial}{\partial x}\big)^2\Big\}\psi(x,z) = 2i\kappa\tfrac{\partial}{\partial z}\psi(x,z) \tag{6}$$

is structurally identical to the 1-dimensional free particle Schrödinger equation, and from its innumerable solutions one selects the "diffusive solution"

$$\psi(x,z) = \frac{1}{\sqrt{1-i(z/Z)}}\exp\Big\{-a\frac{x^2}{1-iz/Z}\Big\} \quad : \quad 2aZ \equiv \kappa \tag{7}$$

that "evolves in time z" from $\psi(x,0) = e^{-ax^2}$. Here a (which has obvious physical dimension) can be assigned any positive real value, and $Z \equiv \kappa/2a$ can be looked upon as a handy abbreviation. Notice now that, as *Mathematica* is quick to confirm,

$$= \int_{-\infty}^{+\infty} \tfrac{1}{\sqrt{4\pi a}}e^{-\frac{1}{4a}p^2}\cdot e^{-i\big\{px-(p^2/4aZ)z\big\}}\,dp \tag{8}$$

The functions $e^{-i\big\{px-(p^2/4aZ)z\big\}}$ are readily seen to be exact solutions of the "Schrödinger equation" (6): exact solutions, that is to say, of the *wrong* *equation*! To summarize: Gaussian beam theory, in its simplest manifestation, purports to extract informative physics from functions of the form

$$\varphi(t,x,z) = \int_{-\infty}^{+\infty} G(p)\cdot e^{i\left\{\omega t - px - kz\right\}}\,dp \quad : \quad k = \kappa - p^2/2\kappa \tag{9}$$
$$G(p) = \tfrac{1}{\sqrt{4\pi a}}e^{-\frac{1}{4a}p^2}$$

—functions which are, in fact, *not solutions of the wave equation* (1). How can such a claim be supported?

We have

$$\begin{aligned} p^2 + k^2 &= p^2 + \left[\kappa - p^2/2\kappa\right]^2 \\ &= \kappa^2 + \left[p^2/2\kappa\right]^2 \end{aligned} \tag{10}$$

whereas

$$= \kappa^2 \quad \text{is stipulated at (3)}$$

Moreover, the latter condition requires that $-\kappa \leqslant p \leqslant +\kappa$ while the integral (9) ranges on $-\infty \leqslant p \leqslant +\infty$. The reason the red term hurts less and less as the parameter $1/4a$ becomes large—the reason we "can have it both ways"—is exposed in the following figure:

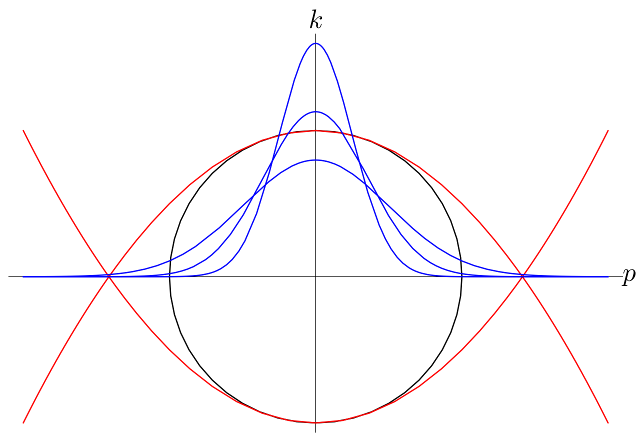

FIGURE 1: *The black circle locates the points that satisfy the exact relation $p^2 + k^2 = \kappa^2$; the red arcs locate the points that satisfy the approximate relation* (10). *The blue Gaussians become narrower as $1/4a$ becomes larger, and—by discriminating against p-values where the red and black curves differ significantly—render the distinction essentially invisible.*

2. Exact theory of 2-dimensional scalar beams. Parameterize the solutions of $p^2 + k^2 = \kappa^2$ by writing

$$\left.\begin{aligned} p &= \kappa\sin\theta \\ k &= \kappa\cos\theta \end{aligned}\right\} \tag{11}$$

and note in passing that

$$k = \kappa\sqrt{1-(p/\kappa)^2} = \kappa - (p^2/2\kappa) + \cdots$$

gives back precisely the approximate relation (10) when the higher-order terms are abandoned (which, of course, is why the black and red curves in Figure 1 conform so neatly when p is small).

The exact solutions of (1) can in this notation be represented

$$\varphi(t,x,z) = \int_{-\pi}^{+\pi} g(\theta)\cdot e^{i\left\{\omega t - \kappa[x\sin\theta + z\cos\theta]\right\}}\,d\theta \tag{12}$$

Though the weight function $g(\theta)$ is arbitrary, we have interest in assigning to it a "Gaussian" design. How is that to be accomplished on the restricted compass of a circle? One (in my view especially "natural") way to accomplish that objective would be to adopt the stereographic procedure described in the following figure:

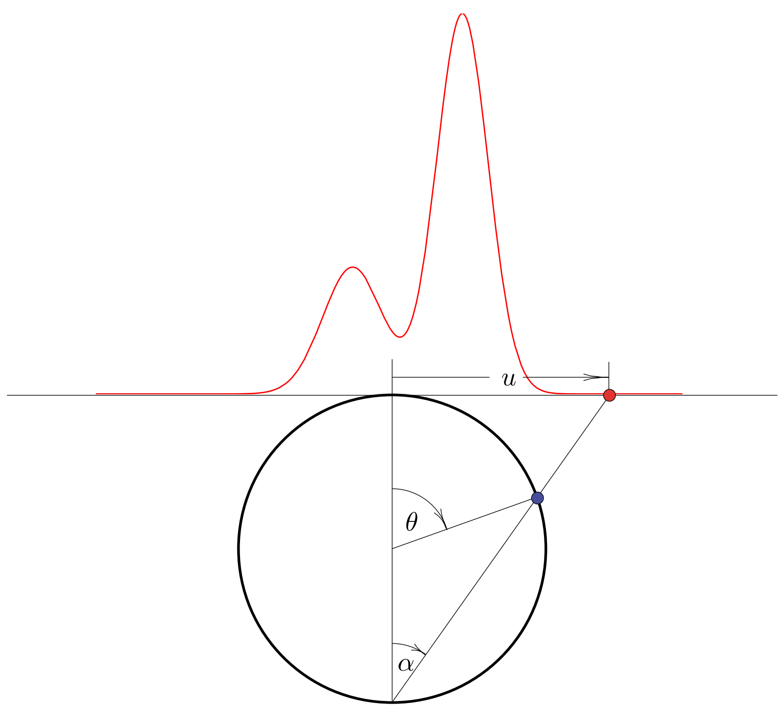

FIGURE 2: *South-polar stereographic projection* • → • *from the line tangent at the North Pole onto a circle of radius* κ. *Distributions* $G(u)$ *written onto the line become distributions* $g(\theta)$ *written onto the circle by the rule* $G(u)\,du = g(\theta)\,d\theta$.

Working from the figure, we have $\tan\alpha = u/2\kappa = \tan\frac{1}{2}\theta$, giving

$$u = 2\kappa\tan\tfrac{1}{2}\theta = \kappa\big\{\theta + \tfrac{1}{12}\theta^3 + \tfrac{1}{120}\theta^5 + \cdots\big\} \tag{13.1}$$

and

$$du = \kappa\sec^2\tfrac{1}{2}\theta\, d\theta = \kappa\big\{1 + \tfrac{1}{4}\theta^2 + \tfrac{1}{24}\theta^4 + \cdots\big\}d\theta \tag{13.2}$$

If, in particular, we take $G(u)$ to be the normalized Gaussian

$$G(u) = \tfrac{1}{\sqrt{4\pi a}}e^{-\frac{1}{4a}u^2} \tag{14.1}$$

then we find

$$\begin{aligned} g(\theta) &= \kappa\tfrac{1}{\sqrt{4\pi a}}\sec^2\tfrac{1}{2}\theta\cdot e^{-\frac{1}{a}\kappa^2\tan^2\frac{1}{2}\theta} \\ &= \kappa\tfrac{1}{\sqrt{4\pi a}}\sec^2\tfrac{1}{2}\theta\cdot e^{-\frac{1}{a}\kappa^2[\sec^2\frac{1}{2}\theta - 1]} \\ &= \kappa\tfrac{1}{\sqrt{4\pi a}}\big\{1 + \tfrac{1}{4}\theta^2 + \tfrac{1}{24}\theta^4 + \cdots\big\}\cdot e^{-\frac{1}{a}\kappa^2[\frac{1}{4}\theta^2 + \frac{1}{24}\theta^4 + \cdots]} \end{aligned} \tag{14.2}$$

We are gratified—but not at all surprised—to be informed by *Mathematica* that

$$\texttt{NIntegrate[}g(\theta)\texttt{], \{}\theta\texttt{,}-\pi\texttt{,}+\pi\texttt{\}] = 1.}$$

when the parameters κ and a are assigned representative numerical values. We note also that as a becomes small we in leading approximation have

$$\begin{aligned} \int_{-\pi}^{+\pi} g(\theta)\,d\theta = \int_{-\pi}^{+\pi}\kappa\tfrac{1}{\sqrt{4\pi a}}e^{-\frac{1}{4a}\kappa^2\theta^2}\,d\theta &= \texttt{Erf}\big[\tfrac{\pi\kappa}{2\sqrt{a}}\big] \\ &\downarrow \\ &= 1 \quad \text{when } \pm\pi \text{ replaced by } \pm\infty \end{aligned}$$

We are brought thus to the construction

$$\varphi(t,x,z) = \int_{-\pi}^{+\pi}\kappa\tfrac{1}{\sqrt{4\pi a}}\sec^2\tfrac{1}{2}\theta\cdot e^{-\frac{1}{a}\kappa^2\tan^2\frac{1}{2}\theta}\cdot e^{i\big\{\omega t - \kappa[x\sin\theta + z\cos\theta]\big\}}\,d\theta \tag{15}$$

of the fundamental object in what might plausibly be called the "exact theory of approximately Gaussian beams," a theory that gives back the more familiar "approximate theory of exactly Gaussian beams" as a becomes small:

$$\begin{aligned} &\downarrow \\ &= \int_{-\pi}^{+\pi}\kappa\tfrac{1}{\sqrt{4\pi a}}\cdot e^{-\frac{1}{4a}\kappa^2\theta^2}\cdot e^{i\big\{\omega t - x\kappa\,\theta - z\kappa[1-\frac{1}{2}\theta^2]\big\}}\,d\theta \end{aligned}$$

The preceding equation is brought into *precisely* agreement with (9) if one adjusts the limits on the integral ($\pi\mapsto\infty$) and also the notation: $\kappa\theta\mapsto p$ (which, according to (13.1), makes good sense when θ is small; *i.e.*, in the vicinity of the North Pole).

The situation with regard to (15) is clarified if, within that exact context, one reverts (by change of variable) to u-notation: drawing upon (13) and the elementary fact[3] that if $t \equiv \tan\frac{1}{2}\theta = u/2\kappa$ then

$$\begin{aligned}\sin\theta &= \frac{2t}{1+t^2} = \frac{4\kappa u}{4\kappa^2+u^2} = \tfrac{1}{\kappa}u\Big\{1-(u/2\kappa)^2+(u/2\kappa)^4-\cdots\Big\}\\ \cos\theta &= \frac{1-t^2}{1+t^2} = \frac{4\kappa^2-u^2}{4\kappa^2+u^2} = 1-2(u/2\kappa)^2+2(u/2\kappa)^4-\cdots\end{aligned}$$

we obtain

$$\varphi(t,x,z) = \int_{-\infty}^{+\infty}\tfrac{1}{\sqrt{4\pi a}}e^{-\frac{1}{4a}u^2}\cdot\exp\Big\{i\Big[\omega t-\kappa\frac{4\kappa u}{4\kappa^2+u^2}x-\kappa\frac{4\kappa^2-u^2}{4\kappa^2+u^2}z\Big]\Big\}\,du \tag{16}$$

> REMARK: The stereographic projection trick led us to this result by a series of natural steps. But we would have been led *directly* to (16) if, in place of (11), we had adopted the following non-obvious parameterization
>
> $$\begin{aligned}p &= \kappa\frac{4\kappa u}{4\kappa^2+u^2} = u+\cdots\\ k &= \kappa\frac{4\kappa^2-u^2}{4\kappa^2+u^2} = \kappa-(u^2/2\kappa)+\cdots\end{aligned}$$
>
> of the solutions of $p^2+k^2=\kappa^2$. The labor invested in development of the polar representation is, however, not labor wasted, for that representation remains indispensable in other connections: see, for instance, Figure 3.

When a is sufficiently small the integral at (16) is readily performed in analytical closed form: *Mathematica* supplies

$$\begin{aligned}\varphi(t,x,z) &= \int_{-\infty}^{+\infty}\tfrac{1}{\sqrt{4\pi a}}e^{-\frac{1}{4a}u^2}\cdot\exp\Big\{i\Big[\omega t-ux-\big[\kappa-(u^2/2\kappa)\big]z\Big]\Big\}\,du\\ &= e^{i(\omega t-\kappa z)}\cdot\frac{1}{\sqrt{1-iz/Z}}\exp\Big\{-a\frac{x^2}{1-iz/Z}\Big\}\end{aligned} \tag{17}$$

with $Z \equiv \kappa/2a$. We have recovered the standard (approximate) Gaussian beam. But exact analytical evaluation of the integral (15/16) appears to present an intractable problem: the best I can presently hope to do is to discuss some of its properties. Look, for example, to its asymptotics: if we write

$$\left.\begin{aligned}x &= r\sin\alpha\\ z &= r\cos\alpha\end{aligned}\right\} \tag{18}$$

---

[3] M. Abramowitz & I. Stegun, *Handbook of Mathematical Functions* (1964), **4.3.23**.

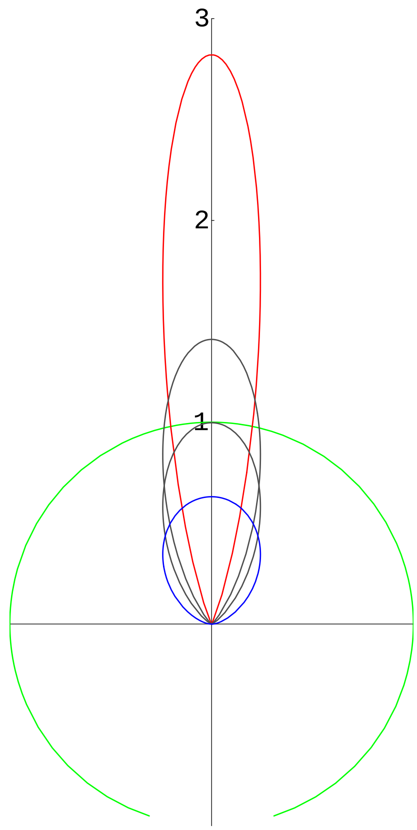

FIGURE 3: *Polar plots, in the cases*

$$\kappa = 1 \quad \textit{and} \quad a = \big\{0.01, 0.04, 0.08, 0.20\big\}$$

of the function

$$g(\theta;\kappa,a) \equiv \kappa\tfrac{1}{\sqrt{4\pi a}}\sec^2\tfrac{1}{2}\theta\cdot e^{-\frac{1}{a}\kappa^2\tan^2\frac{1}{2}\theta}$$

introduced at (14.2); *i.e., of what "Gaussian weighting on the circle" does to unit* $\boldsymbol{k}$*-vectors that point in various directions. Polar plots of distributions are visually deceptive: one has*

$$\int_{-\pi}^{+\pi} g(\theta;\kappa,a)\,d\theta = 1 \quad : \quad \text{all } \kappa, \text{ all } a$$

but the area within the loops is given by

$$\int_{-\pi}^{+\pi} \tfrac{1}{2}[g(\theta;\kappa,a)]^2\,d\theta = \big\{0.99985, 0.50366, 0.35967, 0.23417\big\}$$

then (15) becomes

$$\varphi(t,x,z) = e^{i\omega t}\int_{-\pi}^{+\pi} g(\theta)\, e^{-ir\kappa\cos(\theta-\alpha)}\, d\theta \tag{19.1}$$

which yields in the the limit $r\to\infty$ to the "method of stationary phase:"[4] we expect on that basis to have

$$\sim \sum_{\substack{\text{zeros } \theta_0\\ \text{of } g'(\theta)}} e^{i\omega t} g(\theta_0)\, e^{-ir\kappa\cos(\theta_0-\alpha)}\sqrt{\frac{2\pi}{r\kappa\cos(\theta_0-\alpha)}}\; e^{\pm\frac{1}{4}i\pi}$$

where $\frac{d}{d\theta}[-\kappa\cos(\theta-\alpha)]=0$ supplies $\theta_0=\alpha+n\pi$ (effectively: $\theta_0=\alpha$ else $\theta_0=\alpha\pm\pi$, of which only one falls within the range of the $\int$) and where we are to take the upper or lower sign according as $\frac{d^2}{d\theta^2}[-\kappa\cos(\theta-\alpha)]=\kappa\cos(\theta-\alpha)$ is positive or negative at θ_0. The specific implication of this line of argument is that

$$\begin{aligned}&\sim \sqrt{2\pi/\kappa r}\,g(\alpha)\, e^{i(\omega t-\kappa r+\frac{1}{4}\pi)}\\ &\qquad+\sqrt{2\pi/\kappa r}\,g(\alpha-\pi)\, e^{i(\omega t+\kappa r-\frac{1}{4}\pi)}\end{aligned} \tag{19.2}$$

The first term describes an out-rushing monochromatic cylindrical wave, the amplitude of which falls off as $r^{-\frac{1}{2}}$ (mechanical attributes of the wave, since quadratic in amplitude, therefore fall off as r^{-1}, which is to say: "geometrically"). The second term describes a phase-shifted in-rushing cylindrical wave. The angular attenuation factor $g(\alpha)$ was seen at (14.2) to be symmetric and periodic, so $g(\alpha+\pi)=g(\alpha-\pi)$, and it has the form shown in Figure 3. The implication is that $g(\alpha)$-term predominates in the forward direction (where the $g(\alpha-\pi)$-term can in excellent approximation be ignored), and that the terms exchange roles in the backward direction.

The result achieved at (19.2) conforms beautifully to physical intuition, yet conceals some points worthy of comment:

- Retreat in the direction α to a point far from the origin. There you see what looks locally like a (phase-shifted) version of the plane wave

$$e^{i(\omega t-\boldsymbol{k}\cdot\boldsymbol{x})}\quad\text{with}\quad \boldsymbol{k}=\begin{pmatrix}\kappa\sin\alpha\\ \kappa\cos\alpha\end{pmatrix}$$

  But you have not entirely escaped from the efffects of the other Fourier components of the wave field—those with $\boldsymbol{k}$-vectors *not* parallel to your direction of retreat (each of which, after all, extends infinitely far in all spacetime directions): *they have conspired to produce the angular/radial attenuation factor and phase shift.*
- The argument that led (19.1)→(19.2) exploited special properties of the "circular Gaussian" $g(\theta)$ only in its final steps: we can therefore expect something very like it to pertain to *any* $g(\theta)$; *i.e.*, to *any 2-dimensional superposition of plane waves*.

---

[4] See F. W. J. Olver, *Asymptotics and Special Functions* (1997), page 96.

In the approximation to Gaussian beam theory we at (17) had a result that can be expressed

$$\begin{aligned}\varphi(t,x,z) &= \frac{1}{\sqrt{1+(z/Z)^2}}\exp\Big\{-a\frac{x^2}{1+(z/Z)^2}\Big\}\cdot e^{i\,\wp(t,x,z)}\\ &= [\text{amplitude function } A(x,z)]\cdot[\text{oscillatory factor }\cos\wp(t,x,z)]\end{aligned}$$

The (t-independent!) amplitude function is more susceptible to close study than the oscillatory factor, and is plotted in the following figure:

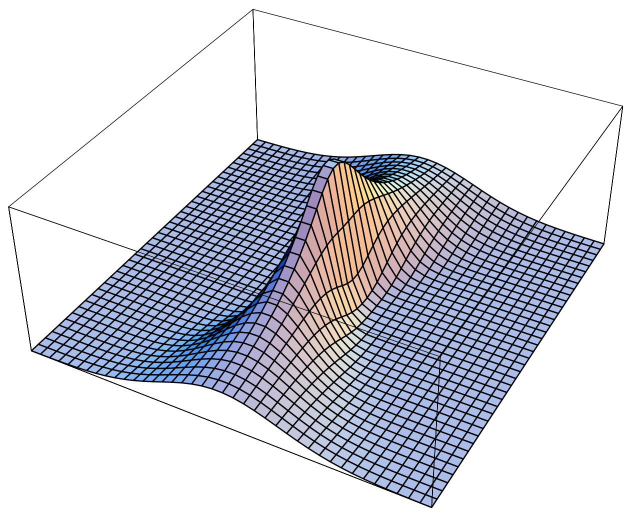

FIGURE 4: *Graph of* $A(x,z;a,Z)$ *in the case* $a=1$ *and* $Z=5$. *For clarity, the transverse* x*-axis has been stretched by a factor of two:* x *ranges on* $[-10,+10\,]$, z *ranges on* $[-20,+20\,]$.

It will, however, remain difficult to discuss the (small) changes brought about in the exact theory until some way is found to perform the integral encountered on the right side of (16).

2. Fourier analysis of the "circular Gaussian distribution function." We described at (14.2), and again in the caption to Figure 3, the result of projecting a Gaussian onto a circle of radius κ. It will facilitate the work at hand if we adjust our notation, writing

$$\begin{aligned}g(\theta;\beta) &\equiv \sqrt{\beta/4\pi}\,\sec^2\tfrac{1}{2}\theta\cdot e^{-\beta\tan^2\frac{1}{2}\theta}\\ &= \sqrt{\beta/4\pi}\big\{1+\tfrac{1}{4}\theta^2+\tfrac{1}{24}\theta^4+\cdots\big\}\cdot e^{-\beta[\frac{1}{4}\theta^2+\frac{1}{24}\theta^4+\cdots]}\end{aligned}$$

where $\beta\equiv\kappa^2/a$ is a dimensionless lumped parameter. The notation emphasizes that where formerly we employed the phrase "as a becomes small" we could just as well have alluded to "the high-frequency limit" $\kappa\uparrow\infty$: that these two notions are equivalent becomes clear when one looks again to Figure 2. The limiting form of the statement

$$\int_{-\pi}^{+\pi} g(\theta;\beta)\,d\theta = \int_{-\pi}^{+\pi}\sqrt{\beta/4\pi}\sec^2\tfrac{1}{2}\theta\cdot e^{-\beta\tan^2\frac{1}{2}\theta}\,d\theta = 1 \quad : \quad \text{all } \beta \tag{20.1}$$

can—for diagramatically evident reasons (the distribution has become localized to the region where the circle and its tangent are nearly coincident)—be cast in this "classically Gaussian" form

$$\int_{-\infty}^{+\infty}\sqrt{\beta/4\pi}\, e^{-\frac{1}{4}\beta\theta^2}\,d\theta = 1 \tag{20.2}$$

which, though valid for all β, can be extracted from (20.1) only as β becomes large.

So much by way of preparation. I present the following figure to emphasize that $g(\theta;\beta)$ is, for all β, an *even* function of θ, and lives on the interval $[-\pi,+\pi\,]$.

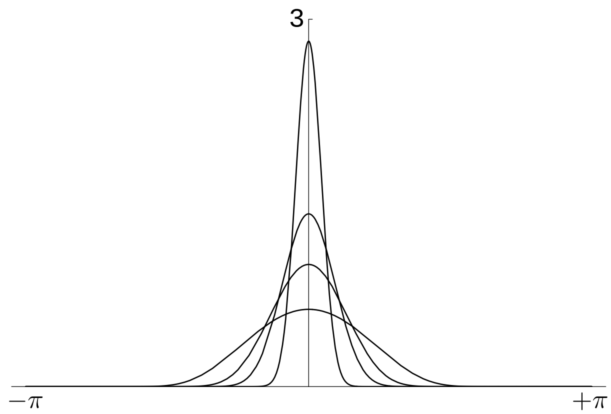

FIGURE 5: *Cartesian plots of precisely the functions $g(\theta;\kappa,a)$—now called $g(\theta;\beta)$ with $\beta\equiv\kappa^2/a$—of which polar plots are presented in Figure 3. In this display the area under each curve is unity. The peaks become narrower/taller in the high-frequency limit $\beta\uparrow\infty$, and it becomes evident that*

$$\lim_{\beta\uparrow\infty} g(\theta;\beta) = \delta(\theta) \tag{21}$$

The functions

$$\begin{aligned} C_0(\theta) &\equiv \tfrac{1}{\sqrt{2\pi}} \\ C_n(\theta) &\equiv \tfrac{1}{\sqrt{\pi}}\cos n\theta \quad : \quad n = 1, 2, 3, \ldots \end{aligned}$$

are well known to be orthonormal

$$\int_{-\pi}^{+\pi} C_m(\theta)\, C_n(\theta)\, d\theta = \delta_{mn}$$

and complete within the space of such functions. We expect, therefore, to be able to write

$$g(\theta;\beta) = \sum_m g_m(\beta)\, C_m(\theta) \tag{22.1}$$

$$g_m(\beta) = \int_{-\pi}^{+\pi} g(\theta;\beta)\, C_m(\theta)\, d\theta \tag{22.2}$$

The evaluation of $g_0(\beta)$ is, by (20.1), immediate

$$g_0(\beta) = \tfrac{1}{\sqrt{2\pi}}$$

but the Fourier coefficients of higher order are intricate: *Mathematica* supplies

$$\begin{aligned}
g_m(\beta) &= \int_{-\pi}^{+\pi} \sqrt{\beta/4\pi}\,\sec^2\tfrac{1}{2}\theta\cdot e^{-\beta\tan^2\frac{1}{2}\theta}\,\tfrac{1}{\sqrt{\pi}}\cos n\theta\, d\theta\\
&\Downarrow\\
g_1(\beta) &= \quad -\tfrac{1}{\sqrt{\pi}} \qquad\qquad\qquad\qquad + 2\cdot\sqrt{\beta}e^{\beta}\big\{1-\mathrm{erf}\big(\sqrt{\beta}\big)\big\} && (23.1)\\
g_2(\beta) &= \quad +\tfrac{1+8\beta}{\sqrt{\pi}} \qquad\qquad\quad - (4+8\beta)\cdot\sqrt{\beta}e^{\beta}\big\{1-\mathrm{erf}\big(\sqrt{\beta}\big)\big\} && (23.2)\\
g_3(\beta) &= -\tfrac{1+24\beta+16\beta^2}{\sqrt{\pi}} + (6+32\beta+16\beta^2)\cdot\sqrt{\beta}e^{\beta}\big\{1-\mathrm{erf}\big(\sqrt{\beta}\big)\big\} && (23.3)\\
&\vdots
\end{aligned}$$

which—ugly though they are—are found at $\beta = 10$ to agree precisely with the results of numerical integration. But if we work in the high-frequency (or Gaussian) approximation—which is to say: from

$$g_m(\beta) \approx \int_{-\infty}^{+\infty} \sqrt{\beta/4\pi}\, e^{-\frac{1}{4}\beta\theta^2}\,\tfrac{1}{\sqrt{\pi}}\cos n\theta\, d\theta$$

—then we obtain results of striking simplicity:

$$g_1(\beta) \approx \tfrac{1}{\sqrt{\pi}} e^{-1/\beta} \tag{24.1}$$

$$g_2(\beta) \approx \tfrac{1}{\sqrt{\pi}} e^{-4/\beta} \tag{24.2}$$

$$g_3(\beta) \approx \tfrac{1}{\sqrt{\pi}} e^{-9/\beta} \tag{24.3}$$

$$\vdots$$

I interrupt this discussion to indicate how the expressions on the right side of (23) come to share the asymptotic properties of their counterparts in (24). We are informed at **7.1.23** in Abramowitz & Stegun that

$$\sqrt{\beta}e^{\beta}\big\{1-\mathrm{erf}\big(\sqrt{\beta}\big)\big\} = \tfrac{1}{\sqrt{\pi}}\Big\{1 - \tfrac{1}{2\beta} + \tfrac{3}{(2\beta)^2} - \tfrac{15}{(2\beta)^3} + \tfrac{35}{(2\beta)^4} - \cdots\Big\}$$

Returning with that information to (23) and simplifying, we obtain

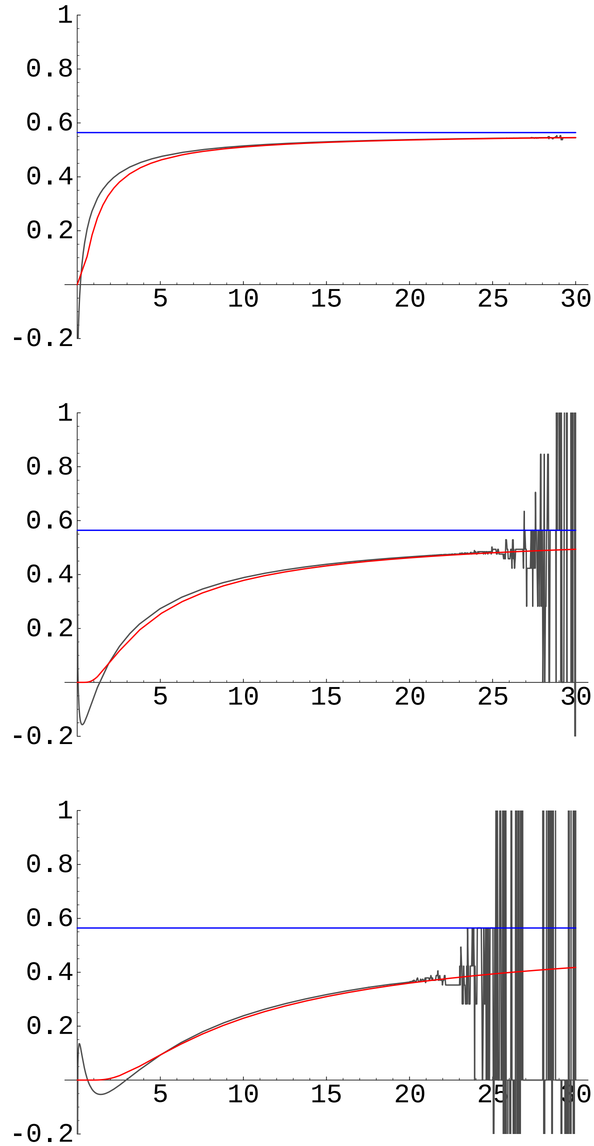

FIGURE 6: *Comparison (reading from top ↓ bottom) of the functions* $g_1(\beta)$, $g_2(\beta)$ *and* $g_3(\beta)$ *that appear on the right sides of* (23) *with (in red) their counterparts in* (24)*. The blue line marks the asymptote at* $1/\sqrt{\pi}$*. It is evident that Mathematica encountered a numerical instability problem when compiling the data: its source is discussed in connection with the next figure.*

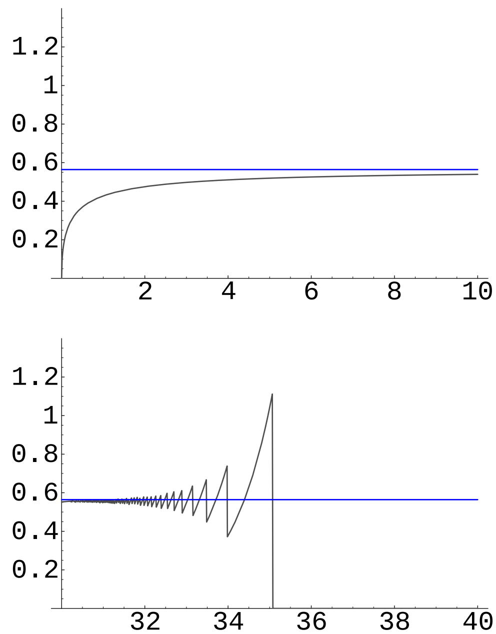

FIGURE 7: *Graphs of the function*

$$\sqrt{\beta}e^{\beta}\big\{1-\text{erf}\big(\sqrt{\beta}\,\big)\big\}\equiv\sqrt{\beta}e^{\beta}\text{erfc}\big(\sqrt{\beta}\,\big)$$

which occurs as a factor in (23). *The blue line marks the asymptote at* $1/\sqrt{\pi}$*, my conjecture being that*

$$\lim_{\beta\uparrow\infty}\sqrt{\beta}e^{\beta}\big\{1-\text{erf}\big(\sqrt{\beta}\,\big)\big\}=\tfrac{1}{\sqrt{\pi}}$$

[*This is, in fact, a classic result: see Abramowitz & Stegun* **7.1.23**.] *Since* $\sqrt{\beta}e^{\beta}\uparrow\infty$ *while* $\big\{1-\text{erf}\big(\sqrt{\beta}\,\big)\big\}\downarrow 0$, *it might seem remarkable that their product approaches a finite limit, but this is in fact a common occurrence: consider the example* $\frac{1}{x}\sin x$. *On evidence of the lower figure,* Mathematica *appears to find its work to be computationally delicate for* β *larger than about 30, and I take this fact to lie at the heart of the instability evident in Figure 6.*

$$\begin{aligned}
g_1(\beta)&=\tfrac{1}{\sqrt{\pi}}\big\{1-\tfrac{1}{\beta}+\tfrac{3}{2\beta^2}-\tfrac{15}{4\beta^3}+\cdots\big\}\\
g_2(\beta)&=\tfrac{1}{\sqrt{\pi}}\big\{1-\tfrac{4}{\beta}+\tfrac{12}{\beta^2}-\tfrac{45}{\beta^3}+\cdots\big\}\\
g_3(\beta)&=\tfrac{1}{\sqrt{\pi}}\big\{1-\tfrac{9}{\beta}+\tfrac{99}{2\beta^2}-\tfrac{795}{4\beta^3}+\cdots\big\}\\
&\;\;\vdots
\end{aligned}$$

which establishes the point at issue.

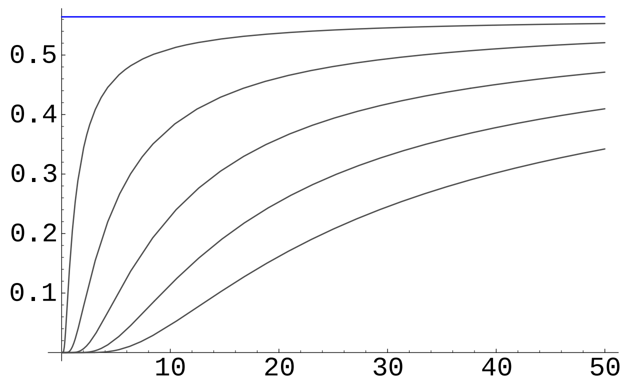

FIGURE 8: *Superimposed plots of the functions $g_m(\beta)$ defined at (24): $m = 1, 2, 3, 4, 5$, and of their shared asymptote at $1/\sqrt{\pi}$. It is seen that the Fourier coefficients $g_m(\beta)$ switch on sequentially as β ascends. Specifically: $g_m(\beta)$ is activated as β approaches/passes the value m^2.*

Working most conveniently from (24) we find that the Fourier coefficients "switch on sequentially" as β increases (see the figure), and that in the limit $\beta \to \infty$ (22.1) becomes

$$\lim_{\beta\uparrow\infty} g(\theta;\beta) = \delta(\theta - 0) = \tfrac{1}{\sqrt{2\pi}}C_0(\theta) + \sum_{m=1}^{\infty} \tfrac{1}{\sqrt{\pi}}C_m(\theta) = \sum_{m=0}^{\infty} C_m(\theta)C_m(0)$$

which can be read as an allusion to the *completeness* of the functions $\{C_m(\theta)\}$.[5]

Whether it is exactly or only approximately that the g_m are known, and whether they refer to the "circular Gaussian" or to some other (symmetric) distribution function, the equation

$$\varphi(t,x,z) = \sum_m g_m \int_{-\pi}^{+\pi} C_m(\theta)e^{i[\,\omega t - x\kappa\sin\theta - z\kappa\cos\theta]}\,d\theta$$

describes an invariably *exact* solution of the wave equation. Our job is to perform the integrations. To that end we adopt the polar notation introduced at (18), writing

$$\varphi_m = \begin{cases} \frac{1}{\sqrt{2\pi}}\displaystyle\int_{-\pi}^{+\pi}\cos[\omega t - \kappa r\cos(\theta-\alpha)]\,d\theta & : \ m = 0 \\ \frac{1}{\sqrt{\pi}}\displaystyle\int_{-\pi}^{+\pi}\cos[m\theta]\cos[\omega t - \kappa r\cos(\theta-\alpha)]\,d\theta & : \ m = 1, 2, 3, \ldots \end{cases}$$

[5] The argument suggests a technique for establishing completeness in much more general situations: expand a representation of the δ-function, then pass to the limit.

Mathematica supplies

$$\varphi_0 = \sqrt{2\pi}\cos\omega t\cdot J_0(\kappa r) \tag{25.0}$$

while in higher order we write

$$\begin{aligned}
\int_{-\pi}^{+\pi}\cos[m\theta]\cos[\omega t-\kappa r\cos(\theta-\alpha)]\,d\theta &= \int_{-\pi-\alpha}^{+\pi-\alpha}\cos[m(\vartheta+\alpha)]\cos[\omega t-\kappa r\cos\vartheta]\,d\theta\\
&= \int_{-\pi}^{+\pi}\cos[m(\vartheta+\alpha)]\cos[\omega t-\kappa r\cos\vartheta]\,d\theta \quad \text{because integrand periodic}\\
&= \int_{-\pi}^{+\pi}\big\{\cos[m\alpha]\cos[m\vartheta]-\sin[m\alpha]\sin[m\vartheta]\big\}\cos[\omega t-\kappa r\cos\vartheta]\,d\theta\\
&= \cos[m\alpha]\cdot\int_{-\pi}^{+\pi}\big\{\cos[m\vartheta]\cos[\omega t-\kappa r\cos\vartheta]\,d\theta \quad \text{because } \sin[m\vartheta] \text{ is odd}
\end{aligned}$$

and obtain

$$\begin{aligned}
\varphi_1 &= +2\sqrt{\pi}\,\sin\omega t\cos 1\alpha\cdot J_1(\kappa r) && (25.1)\\
\varphi_2 &= -2\sqrt{\pi}\,\cos\omega t\cos 2\alpha\cdot J_2(\kappa r) && (25.2)\\
\varphi_3 &= -2\sqrt{\pi}\,\sin\omega t\cos 3\alpha\cdot J_3(\kappa r) && (25.3)\\
\varphi_4 &= +2\sqrt{\pi}\,\cos\omega t\cos 4\alpha\cdot J_4(\kappa r) && (25.4)\\
\varphi_5 &= +2\sqrt{\pi}\,\sin\omega t\cos 5\alpha\cdot J_5(\kappa r) && (25.5)\\
&\ \vdots
\end{aligned}$$

When r is large we have (see Abramowitz & Stegun, **9.2.1**)

$$\begin{aligned}
\varphi_0 &\sim \quad \sqrt{2\pi}\cos\omega t\cdot\sqrt{2/\pi\kappa r}\cos(\kappa r-\tfrac{1}{4}\pi) \quad : \quad \text{omnidirectional}\\
\varphi_1 &\sim +2\sqrt{\pi}\,\sin\omega t\cos 1\alpha\cdot\sqrt{2/\pi\kappa r}\cos(\kappa r-\tfrac{3}{4}\pi)\\
\varphi_2 &\sim -2\sqrt{\pi}\,\cos\omega t\cos 2\alpha\cdot\sqrt{2/\pi\kappa r}\cos(\kappa r-\tfrac{5}{4}\pi)\\
\varphi_3 &\sim -2\sqrt{\pi}\,\sin\omega t\cos 3\alpha\cdot\sqrt{2/\pi\kappa r}\cos(\kappa r-\tfrac{7}{4}\pi)\\
\varphi_4 &\sim +2\sqrt{\pi}\,\cos\omega t\cos 4\alpha\cdot\sqrt{2/\pi\kappa r}\cos(\kappa r-\tfrac{9}{4}\pi)\\
\varphi_5 &\sim +2\sqrt{\pi}\,\sin\omega t\cos 5\alpha\cdot\sqrt{2/\pi\kappa r}\cos(\kappa r-\tfrac{11}{4}\pi)\\
&\ \vdots
\end{aligned}$$

That the functions (25) are, in fact, exact solutions of the wave equation follows quickly from the observation that in polar coordinates the wave operator

$$\Big\{\big(\tfrac{1}{c}\tfrac{\partial}{\partial t}\big)^2-\big(\tfrac{\partial}{\partial x}\big)^2-\big(\tfrac{\partial}{\partial z}\big)^2\Big\} \quad \text{becomes} \quad \Big\{\big(\tfrac{1}{c}\tfrac{\partial}{\partial t}\big)^2-\big(\tfrac{\partial}{\partial r}\big)^2-\tfrac{1}{r}\tfrac{\partial}{\partial r}-\tfrac{1}{r^2}\big(\tfrac{\partial}{\partial \alpha}\big)^2\Big\}$$

which when applied to φ_m gives[6] $\Box\varphi = [\kappa^2-(\omega/c)^2]\varphi = 0$. The functions (25) are precisely the functions that would have resulted had we undertaken to solve $\Box\varphi = 0$ by separation of variables in polar coordinates, and retained only the solutions that conform to the side condition $\varphi(t,r,\alpha) = \varphi(t,r,-\alpha)$.

---

[6] Use $\Big\{\big(\tfrac{\partial}{\partial r}\big)^2+\tfrac{1}{r}\tfrac{\partial}{\partial r}\Big\}J_m(r) = \big(\tfrac{m^2}{r^2}-1\big)J_m(r)$ or *Mathematica*'s `FullSimplify` command.

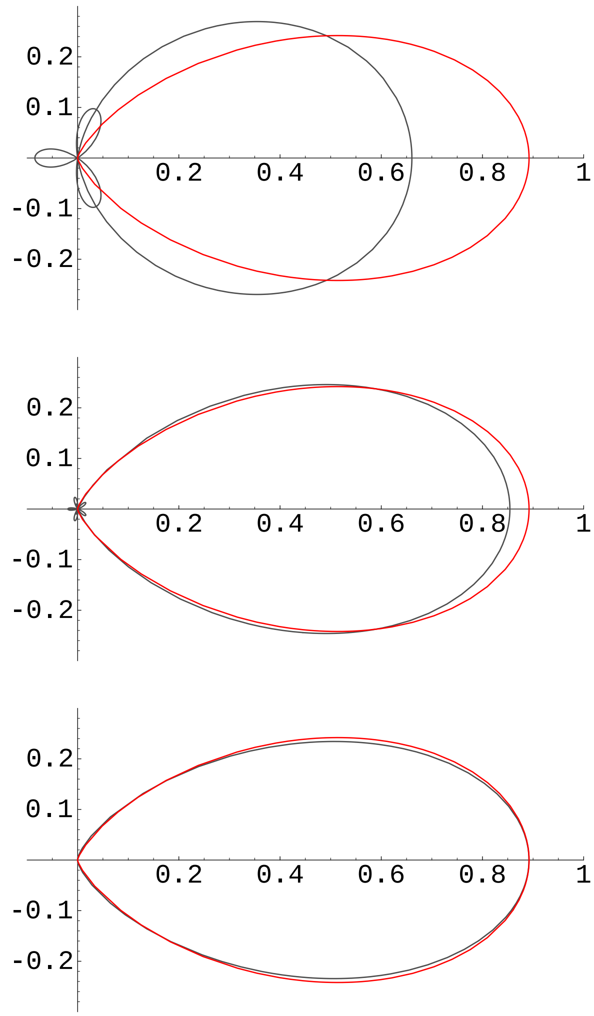

FIGURE 9: *In each polar plot the red loop derives from*

$$g(\theta;\beta) \equiv \sqrt{\beta/4\pi}\,\sec^2\tfrac{1}{2}\theta\cdot e^{-\beta\tan^2\frac{1}{2}\theta}$$

(see again page 10) with $\beta = 10$. *The black loops derive from*

$$g(\theta;\beta,n) \equiv \tfrac{1}{2\pi} + \tfrac{1}{\pi}\sum_{m=1}^{n} e^{-m^2/\beta}\cos m\theta$$

with (reading from top to bottom) $n = 2, 4, 8$. *The surviving error can be attributed to the fact that we took the Fourier coefficients to be given by* (24) *rather than by the exact equations* (23).

The preceding figure demonstrates the effectiveness of our Fourier analysis of the "circular Gaussian distribution function" $g(\theta;\beta)$. But the objects of primary interest are "beams;" we have only incidental interest in "Gaussian beams" . . . so it is gratifying to discover that we are in position now to construct beams in great variety by assigning values to the coefficients $g_m(\beta)$ that differ how ever we please from those described at (23/24). In each case we are able to describe exactly (if as a series of Bessel functions) the resulting field.

But within the universe of beams the Gaussian beam—of which I a moment ago spoke dismissively—occupies a special place. That is because it supports an infinite population of sibling beams, "Gaussian beams of higher modal order" . . . which collectively provide a natural basis in "beam space." I turn now to a description of how this comes about.

3. Higher modes of a Gaussian beam. At (7) we looked to a particular solution of

$$\Big\{\big(\tfrac{\partial}{\partial x}\big)^2\Big\}\psi(x,z) = 2i\kappa\tfrac{\partial}{\partial z}\psi(x,z) \tag{6}$$

—namely, the "diffusive solution"

$$\psi(x,z) = \frac{1}{\sqrt{1-i(z/Z)}}\exp\Big\{-a\frac{x^2}{1-iz/Z}\Big\} \quad : \quad 2aZ\equiv\kappa \tag{7}$$

that "evolves in time z" from $\psi(x,0) = e^{-ax^2}$. Concerning that evolution process:

We know from the quantum mechanics of a free particle that the solution of

$$\big(\tfrac{\partial}{\partial x}\big)^2\psi(x,t) = -\tfrac{2m}{\hbar}i\tfrac{\partial}{\partial t}\psi(x,t)$$

that evolves dynamically from $\psi(x,0)$ can be described

$$\psi(x,t) = \int_{-\infty}^{+\infty} K(x,t;y,0)\,\psi(y,0)\,dy$$

where the "free particle propagator" is given by

$$K(x,t;y,0) = \sqrt{m/2\pi i\hbar t}\,\exp\Big\{\frac{i}{\hbar}\frac{m}{2}\frac{(x-y)^2}{t}\Big\}$$

In the present context we expect therefore to have

$$\psi(x,z) = \int_{-\infty}^{+\infty} \mathcal{K}(x,z;y,0)\,\psi(y,0)\,dy \tag{26}$$

with (replace $m/\hbar \longmapsto -\kappa = -2aZ$)

$$\mathcal{K}(x,z;y,0) = \sqrt{iaZ/\pi z}\,\exp\Big\{-iaZ\,\frac{(x-y)^2}{z}\Big\}$$

and by computation find that indeed

$$\int_{-\infty}^{+\infty} \mathcal{K}(x,z;y,0)\,e^{-ay^2}\,dy = \text{expression on the right side of (7)}$$

The process just illustrated will acquire special importance in a moment.

It is evident that if $\psi(x,z)$ is a solution of (6) then so also is

$$\psi_m(x,z) \equiv a^{-\frac{m}{2}}\big(-\tfrac{\partial}{\partial x}\big)^m \psi(x,z)$$

where the factor $a^{-\frac{m}{2}}$ has been introduced so as to make ψ and ψ_m have the same physical dimension. But by Rodrigues' construction of the Hermite polynomials (known to *Mathematica* as `HermiteH[m,ξ]`) we have

$$\big(-\tfrac{d}{d\xi}\big)^m e^{-\xi^2} = H_m(\xi)e^{-\xi^2} \quad \text{with} \quad \begin{aligned} H_0(\xi) &= 1\\ H_1(\xi) &= 2\xi\\ H_2(\xi) &= 4\xi^2 - 2\\ H_3(\xi) &= 8\xi^3 - 12\xi\\ H_4(\xi) &= 16\xi^4 - 48\xi^2 + 12\\ &\vdots \end{aligned}$$

It now follows that

$$\psi_m(x,0) = H_m(\sqrt{a}x)e^{-ax^2} \tag{27.0}$$

and that

$$\begin{aligned} \psi_m(x,z) &= a^{-\frac{m}{2}}\big(-\tfrac{\partial}{\partial x}\big)^m \frac{1}{\sqrt{1-i(z/Z)}}\exp\Big\{-a\frac{x^2}{1-iz/Z}\Big\}\\ &= \bigg[\frac{1}{\sqrt{1-i(z/Z)}}\bigg]^{1+m} H_m\bigg[\frac{\sqrt{a}\,x}{\sqrt{1-i(z/Z)}}\bigg]\exp\Big\{-a\frac{x^2}{1-iz/Z}\Big\} \end{aligned} \tag{27.1}$$

Mathematica confirms (in low-order cases) that $\psi_m(x,z)$ is in fact a solution of the beam equation (6):

$$\Big\{\big(\tfrac{\partial}{\partial x}\big)^2 - 4iaZ\tfrac{\partial}{\partial z}\Big\}\psi_m(x,z) = 0$$

The functions (27.1) give back (27.0) at $z=0$, and (much less obviously, but which *Mathematica* is quick again to confirm) can be recovered *from* the latter means of the propagation formula (26).

We are informed, however, that when work-a-day laser physicists refer to the higher modes of a Gaussian beam they have in mind a related but *distinct* population of "Gaussian siblings," namely the functions[7]

---

[7] See page 176 in O. Svelto's classic text *Principles of Lasers* (3rd edition 1989) and §3.3 in H. Kogelnik & T. Li, "Laser beams and resonators," Appl. Opt. **5**, 1550 (1966). I have pulled back their 3-dimensional result to two space dimensions.

$$\Psi_m(x,z) = \left[\frac{1}{1+(z/Z)^2}\right]^{\frac{1}{4}} H_m\left[\frac{\sqrt{2a}\,x}{\sqrt{1+(z/Z)^2}}\right] e^{i[\frac{1}{2}+m]\arctan(z/Z)} \cdot \exp\left\{-a\frac{x^2}{1-iz/Z}\right\} \tag{28.1}$$

That these functions do in fact satisfy the beam equation is readily verified (in any specific low order) by explicit calculation. At the beam waist $z = 0$ we have

$$\downarrow$$
$$= H_m(\sqrt{2a}\,x)e^{-ax^2} \tag{28.0}$$

from which we do in fact recover (28.1) if we use (26) to propagate up and down the z-axis. Comparison of (28.0) with (27.0) shows how very closely related the ψ_m and Ψ_m families are . . . and yet there is a world of difference:

- The red factor in (28.1) is manifestly real: the functions Ψ_m are readily brought to polar form, with explicit phase factors, while the ψ_m are profoundly complex.
- Write

$$e^{i[\frac{1}{2}+m]\arctan(z/Z)}\exp\Big\{-a\frac{x^2}{1-iz/Z}\Big\}$$
$$= \exp\Big\{-a\frac{x^2}{1+(z/Z)^2}\Big\}\cdot\exp\Big\{i\Big[\big(\tfrac{1}{2}+m\big)\varPhi - a\frac{x^2}{1+(z/Z)^2}(z/Z)\Big]\Big\}$$

with $\varPhi(z) \equiv \arctan(z/Z)$. Construct (see again (4))

$$\phi_m(x,z) = e^{-i\kappa z}\cdot\Psi_m(x,z)$$

and observe that, while the equi-phase contours

$$\kappa z - \big(\tfrac{1}{2}+m\big)\varPhi + a\frac{x^2}{1+(z/Z)^2}(z/Z) = \text{constant} \quad : \quad \kappa = 2aZ$$

of the functions ϕ_m are m-dependent, their *curvature as they cross the z-axis* is, by an argument which I will omit,[8] m-independent—the same in all modes. The ϕ_m derived from (27.1) are, in consequence of their profound complexity, readily seen to possess no such property. The point acquires technological importance from the necessarily fixed shape of the mirrors at the ends of a laser cavity.

Svelto presents (28.1) with an "it can be shown" and an allusion to the obscure depths of the theory of resonators. Kogelnik & Li omit the details of their argument "because of space limitations." In §4 I sketch an analytical line of argument from which the functions $\Psi(x,z)$ emerge with a kind of inevitability as "natural constructs."

[8] See CLASSICAL ELECTRODYNAMICS (2002) pages 325–326 for the details.

The “higher modes of a Gaussian beam” is evidently a phrase to which one can attach a *variety* of specific meanings. If the Ψ_m modes are to be assigned pride of place, it must be on *physical* grounds: namely, that those are—for reasons traceable (as suggested above) to the detailed physics of beam production—the modes encountered in the laboratory.

The alternative modal definitions share (among others) this important property: starting from either, one can use the orthogonality of the Hermite polynomials

$$\int_{-\infty}^{+\infty} H_m(\xi)H_n(\xi)e^{-\xi^2}\,d\xi = \sqrt{\pi}m!2^m\delta_{mn}$$

to construct modal superpositions that possess any prescribed structure at the beam waist. What we have now in hand are, in effect, two alternative “bases in beam space”—two to be joined soon by infinitely many others.

4. Development of modal properties by the generating function method. We have several times drawn general inferences from evidence supplied by *Mathematica* in a few low-order cases. It is obvious, however, that such a procedure, for all its heuristic value, can never provide formal proof of infinite sets of propositions. I describe here how, in the circumstances at hand, the elegant “generating function method” can be used to bridge the gap. By way of orientation . . .

It is well known (and known, in particular, to *Mathematica*) that

$$\sum_{m=0}^{\infty} \tfrac{1}{m!}H_m(\xi)u^n = e^{-u^2+2\xi u}$$

Arguing as follows

$$\begin{aligned}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\tfrac{1}{m!}\tfrac{1}{n!}\Big\{\int_{-\infty}^{+\infty}&H_m(\xi)H_n(\xi)e^{-\xi^2}\,d\xi\Big\}u^nv^n\\ &= \int_{-\infty}^{+\infty} e^{-u^2+2\xi u}e^{-v^2+2\xi v}e^{-\xi^2}\,d\xi\\ &= \sqrt{\pi}e^{2uv}\\ &= \sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\tfrac{1}{m!}\tfrac{1}{n!}\Big\{\sqrt{\pi}m!2^m\delta_{mn}\Big\}u^nv^n\end{aligned}$$

we establish—in a single blow—the infinite set of orthogonality statements presented at the top of the page.

Taking inspiration now from (27.0), we construct the generating function

$$\begin{aligned}g(x,0;u) \equiv \sum_{m=0}^{\infty}\tfrac{1}{m!}\psi_m(x,0)u^m &= \sum_{m=0}^{\infty}\tfrac{1}{m!}H_m(\sqrt{a}x)e^{-ax^2}u^m\\ &= e^{-u^2+2\sqrt{a}\,xu-ax^2}\end{aligned}$$

and use (26) to propagate that function up and down the z-axis:

$$
\begin{aligned}
g(x,z;u) &= \int_{-\infty}^{+\infty} \mathcal{K}(x,z;y,0)\, g(y,0;u)\, dy \\
&\qquad\qquad \mathcal{K}(x,z;y,0) = \sqrt{iaZ/\pi z}\,\exp\Big\{-iaZ\,\frac{(x-y)^2}{z}\Big\} \\
&= \frac{1}{\sqrt{1-iz/Z}}\exp\Big\{-\frac{(u-\sqrt{a}\,x)^2}{1-iz/Z}\Big\} \\
&= \frac{1}{\sqrt{1-iz/Z}}\,e^{-\mathsf{u}^2+2\sqrt{a}\,\mathsf{u}\,\mathsf{x}}\exp\Big\{-a\frac{x^2}{1-iz/Z}\Big\} \\
&= \frac{1}{\sqrt{1-iz/Z}}\Big\{\sum_{m=0}^{\infty}\tfrac{1}{m!}H_m(\sqrt{a}\,\mathsf{x})\,\mathsf{u}^m\Big\}\exp\Big\{-a\frac{x^2}{1-iz/Z}\Big\}
\end{aligned}
$$

Here $\mathsf{u} \equiv u/\sqrt{1-iz/Z}$ and $\mathsf{x} \equiv x/\sqrt{1-iz/Z}$, so we have

$$
= \sum_{m=0}^{\infty}\tfrac{1}{m!}\bigg[\frac{1}{\sqrt{1-iz/Z}}\bigg]^{1+m} H_m\bigg[\frac{\sqrt{a}\,x}{\sqrt{1-iz/Z}}\bigg]\exp\Big\{-a\frac{x^2}{1-iz/Z}\Big\}u^m
$$

which is to say: we have recovered precisely the beam modes (27.1). That each is in fact a solution of the beam equation follows simply and elegantly from the observation that

$$
\Big\{\big(\tfrac{\partial}{\partial x}\big)^2 - 4iaZ\tfrac{\partial}{\partial z}\Big\}g(x,z;u) = 0 \quad : \quad \text{all } u
$$

Which brings me finally to the main point of this discussion. Taking our inspiration now from (28.0) we construct

$$
\begin{aligned}
G(x,0;u;b) &\equiv \sum_{m=0}^{\infty}\tfrac{1}{m!}H_m(b\sqrt{a}x)e^{-ax^2}u^m \\
&= e^{-u^2+2b\sqrt{a}\,xu-ax^2}
\end{aligned}
$$

which gives back $g(x,0;u)$ at $b=1$ and

$$
\begin{aligned}
&\downarrow \\
&= \sum_{m=0}^{\infty}\tfrac{1}{m!}\Psi_m(x,0)\,u^m \quad \text{at} \quad b=\sqrt{2}
\end{aligned}
$$

Drawing again upon (26) to propagate up and down the z-axis, we find

$$
G(x,z;u;b) = \frac{1}{\sqrt{1-iz/Z}}\exp\Big\{-\frac{iZ-(b^2-1)z}{iZ+z}u^2 + \frac{2b\sqrt{a}ux}{1-iz/Z} - \frac{ax^2}{1-iz/Z}\Big\}
$$

which is found to satisfy

$$\Big\{\big(\tfrac{\partial}{\partial x}\big)^2 - 4iaZ\tfrac{\partial}{\partial z}\Big\}G(x,z;u;b) = 0 \quad : \quad \text{all } u\text{, all } b$$

and from which it follows in particular that

$$\begin{aligned} G(x,z;u;1) &= g(x,z;u) \quad : \quad \text{described above}\\ G(x,z;u;\sqrt{2}) &= \frac{1}{\sqrt{1-iz/Z}}\exp\Big\{-\frac{1+iz/Z}{1-iz/Z}u^2+\frac{2\sqrt{2a}\,ux}{1-iz/Z}-\frac{ax^2}{1-iz/Z}\Big\}\end{aligned}$$

But if $1+iz/Z=\sqrt{1+(z/Z)^2}e^{i\Phi}$ then $1-iz/Z=\sqrt{1+(z/Z)^2}e^{-i\Phi}$ and we have

$$\begin{aligned} G(x,z;u;\sqrt{2}) &= \bigg[\frac{1}{\sqrt{1+(z/Z)^2}}\bigg]^{\frac{1}{2}}e^{\frac{1}{2}i\Phi}\\ &\qquad\cdot\exp\Big\{-e^{2i\Phi}u^2+2\,e^{i\Phi}u\frac{\sqrt{2a}\,x}{\sqrt{1+(z/Z)^2}}\Big\}\exp\Big\{-\frac{ax^2}{1-iz/Z}\Big\}\\ &= \sum_{m=0}^{\infty}\tfrac{1}{m!}\bigg\{\bigg[\frac{1}{1+(z/Z)^2}\bigg]^{\frac{1}{4}}H_m\bigg[\frac{\sqrt{2a}\,x}{\sqrt{1+(z/Z)^2}}\bigg]\\ &\qquad\qquad\cdot\exp\Big\{-\frac{ax^2}{1-iz/Z}+i\big(\tfrac{1}{2}+m\big)\Phi\Big\}\bigg\}u^m\\ &= \sum_{m=0}^{\infty}\tfrac{1}{m!}\Psi_m(x,z)\,u^m\end{aligned}$$

To summarize: the generating function $G(x,z;u;b)$ at the bottom of the preceding page presents a b-parameterized *infinitude* of alternative beam-modal definitions. But special simplifications attach to the expression

$$\frac{iZ-(b^2-1)z}{iZ+z}$$

in the cases $b=0$ (trivial), $b=1$ and $b=\sqrt{2}$. Exploitation of those simplifications has been shown to lead to the ψ_m-modes (27.1) and Ψ_m-modes (28.1), respectively—constructions which in this sense occupy preferred places within the range of possibilities.

Since the present essay is addressed primarily to the construction of an *exact* theory of (electromagnetic) beams, I should perhaps emphasize that the discussion this and the preceding section has been concerned with an aspect of the standard *approximate theory of (scalar) beams*. It has been developed in response to some perceptive critical remarks by Morgan Mitchell, who brought to my attention a defect in an earlier version of this material.

3-DIMENSIONAL THEORY

5. Spherical shadow of a bivariate Gaussian. Figure 2 alludes to a technique for erecting distributions on the circle (of radius κ) by stereographic projection from distributions on the line. Here I describe how, by essentially the same technique, distributions on the plane can be stereographically reinterpreted as distributions on the sphere. Let $\{u, \phi\}$ refer to a polar coordinatization of the plane, and let

$$G(u,\phi)\,u\,dud\phi = \begin{cases}\text{weight assigned to the differential}\\ \text{neighborhood } u\,dud\phi \text{ of the planar point}\\ \text{with radial/angular address } \{u,\phi\}\end{cases}$$

We will have interest mainly in rotationally-invariant (or axially-symmetric, or ϕ-independent) distributions $G(u)$: for those we have

$$G(u)\,2\pi u\,du = \begin{cases}\text{weight assigned to the differential ring of radius } u\\ \text{and width } du\text{, concentric about the origin}\end{cases}$$

And of those we will restrict our attention mainly (compare (14.1)) to the Gaussian case

$$G(u) = \tfrac{1}{4\pi a}e^{-\frac{1}{4a}u^2}$$

. . . in which connection we observe that

$$\int_0^\infty \tfrac{1}{4\pi a}e^{-\frac{1}{4a}u^2}\cdot 2\pi u\,du = 1$$

We place the origin of the $\{u, \phi\}$-system at the North Pole of the sphere, on which we inscribe spherical coordinates:

$$\begin{aligned}\theta &\equiv \text{co-latitude (see again Figure 2)}\\ \phi &\equiv \text{longitude}\end{aligned}$$

The $du \times ud\phi$ neighborhood of the point $\{u, \phi\} = \{2\kappa \tan \frac{1}{2}\theta, \phi\}$ on the plane projects to the $\kappa d\theta \times \kappa \sin\theta\, d\phi$ neighborhood of the point $\{\theta, \phi\}$ on the sphere. To obtain the projected distribution $g(\theta, \phi)$ we write

$$\begin{aligned}g(\theta,\phi)\kappa^2 \sin\theta\, d\theta d\phi &= G(u,\phi)\,u\,dud\phi\\ &= G(2\kappa\tan\tfrac{1}{2}\theta,\phi)\,2\kappa\tan\tfrac{1}{2}\theta\cdot\kappa\sec^2\tfrac{1}{2}\theta\,d\theta d\phi\end{aligned}$$

giving

$$\begin{aligned}g(\theta,\phi) &= \frac{2\kappa\tan\frac{1}{2}\theta\cdot\kappa\sec^2\frac{1}{2}\theta}{\kappa^2\sin\theta}\cdot G(2\kappa\tan\tfrac{1}{2}\theta,\phi)\\ &= \sec^4\tfrac{1}{2}\theta\cdot G(2\kappa\tan\tfrac{1}{2}\theta,\phi)\end{aligned}$$

In the presence of rotational invariance nothing happens except that the ϕ's on left and right become moot, and for Gaussian distributions centered at the pole we have (compare (14.2) and the bottom of page 10)

$$\begin{aligned} g(\theta,\phi) &= \Big[\kappa \tfrac{1}{\sqrt{4\pi a}} \sec^2 \tfrac{1}{2}\theta\Big]^2 e^{-\frac{1}{a}\kappa^2 \tan^2 \frac{1}{2}\theta} \quad : \quad \phi\text{-independent} \\ &= \Big[\sqrt{\beta/4\pi}\, \sec^2 \tfrac{1}{2}\theta\Big]^2 e^{-\beta \tan^2 \frac{1}{2}\theta} \\ &\qquad \beta \equiv \kappa^2/a \end{aligned} \tag{29}$$

To test this result we look to

$$\int_0^\pi \left\{\int_0^{2\pi} g(\theta,\phi) \sin\theta \, d\phi\right\} d\theta = \int_0^\pi (\beta/4\pi) \sec^4 \tfrac{1}{2}\theta \cdot e^{-\beta \tan^2 \frac{1}{2}\theta}\, 2\pi \sin\theta \, d\theta$$

and are gratified to be informed by *Mathematica* that

NIntegrate[Evaluate[etc, $\{\theta, 0, \pi\}$]] = 1.

where representative values are assigned to β. Note the adjusted lower limit on the θ-integral: formerly negative values of θ are now subsumed in the ϕ-sweep *around* the sphere.

When presented with a function of the generic form $g(\theta,\phi)$ it is natural to contemplate writing

$$g(\theta,\phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{m=+\ell} g_\ell^m Y_\ell^m(\theta,\phi) \tag{30.1}$$

and using the orthonormality of the spherical harmonics to obtain

$$g_\ell^m = \int_0^\pi \int_0^{2\pi} g(\theta,\phi) \big[Y_\ell^m(\theta,\phi)\big]^* \sin\theta \, d\phi d\theta \tag{30.2}$$

How does the normalization condition

$$\int_0^\pi \int_0^{2\pi} g(\theta,\phi)\, \sin\theta \, d\phi d\theta = 1$$

fit within such a scheme? We have[9]

$$\int_0^\pi \int_0^{2\pi} Y_\ell^m(\theta,\phi)\, \sin\theta \, d\phi d\theta = \begin{cases} \sqrt{4\pi} & \text{if} \quad m = \ell = 0 \\ 0 & \text{otherwise} \end{cases}$$

so normalization forces $g_0^0 = \frac{1}{\sqrt{4\pi}}$ but places no restriction on the other coefficients g_ℓ^m, essentially because $Y_\ell^m(\theta,\phi) \perp Y_0^0(\theta,\phi) \equiv \frac{1}{\sqrt{4\pi}}$.

[9] *Mathematica*—which responds with

$$\sqrt{\tfrac{2\ell+1}{4\pi} \tfrac{(\ell-m)!}{(\ell+m)!}}\, P_\ell^m(\cos\theta) e^{im\phi}$$

to the command SphericalHarmonicY[ℓ, m, θ, ϕ]—can be used to construct experimental evidence for all such claims.

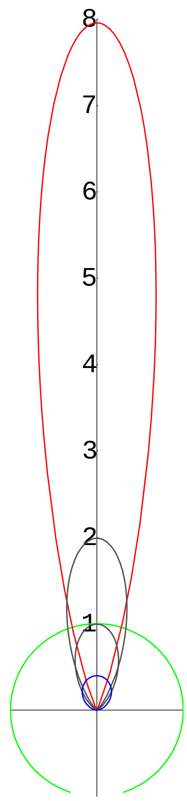

FIGURE 10: *Stepping up from 2 to 3 dimensions is shown on the preceding page to entail an adjustment of the form*

$$\sqrt{\beta/4\pi}\,\sec^2\tfrac{1}{2}\theta \quad\longrightarrow\quad \Big[\sqrt{\beta/4\pi}\,\sec^2\tfrac{1}{2}\theta\Big]^2$$

the effect of which—as illustrated in this cross-section of the spherical plot (compare Figure 3: the same parameter values have been used here as there)—is to sharpen the distribution.

If $g(\theta,\phi)$ is in fact ϕ-independent (*i.e.*, if the distribution is axially symmetric) than only the "zonal harmonics"[10]

$$Y_\ell^0(\theta) = \sqrt{\tfrac{2\ell+1}{4\pi}}P_\ell(\cos\theta)$$

[10] See Philip Morse & Herman Feshbach, *Methods of Theoretical Physics* (1953), page 1264 for discussion of this and some related terminology.

contribute to the sum (30.1): we have

$$g(\theta) = \tfrac{1}{4\pi} + \sum_{\ell=1}^{\infty} g_\ell Y_\ell^0(\theta) \tag{31.1}$$

$$g_\ell = \int_0^\pi g(\theta)\, Y_\ell^0(\theta)\, 2\pi \sin\theta\, d\theta \tag{31.2}$$

If, in particular, $g(\theta)$ has the β-dependent Gaussian design (29) then

$$g_\ell(\beta) = \int_0^\pi \Big[\sqrt{\beta/4\pi}\, \sec^2 \tfrac{1}{2}\theta\Big]^2 e^{-\beta \tan^2 \frac{1}{2}\theta} \cdot \sqrt{\tfrac{2\ell+1}{4\pi}} P_\ell(\cos\theta) \cdot 2\pi \sin\theta\, d\theta$$

Alternatively: if $\cos\theta \equiv w$ then $\sin\theta\, d\theta = -dw$ and[11]

$$\sec^2 \tfrac{1}{2}\theta = \frac{2}{1+w} \quad \text{and} \quad \tan^2 \tfrac{1}{2}\theta = \frac{1-w}{1+w}$$

so we can write

$$g_\ell(\beta) = \int_{-1}^{+1} \Big[\sqrt{\beta/4\pi}\, \tfrac{2}{1+w}\Big]^2 e^{-\beta \frac{1-w}{1+w}} \cdot \sqrt{\tfrac{2\ell+1}{4\pi}} P_\ell(w) \cdot 2\pi\, dw$$

The integral is a little delicate because, though the integrand is well-behaved on the interval, it has an essential singularity at $w = 1$. A second change of variable ameliorates the problem: write $\frac{1-w}{1+w} \equiv u$. Then $w = \frac{1-u}{1+u}$, $dw = -\frac{2}{(1+u)^2} du$, $\frac{2}{1+w} = 1 + u$ and

$$\begin{aligned} &= \int_0^\infty \Big[\sqrt{\beta/4\pi}\,(1+u)\Big]^2 e^{-\beta u} \cdot \sqrt{\tfrac{2\ell+1}{4\pi}} P_\ell(\tfrac{1-u}{1+u}) \cdot 2\pi\, \tfrac{2}{(1+u)^2} du \\ &= \tfrac{1}{\sqrt{4\pi}} \int_0^\infty \beta\sqrt{2\ell+1}\, e^{-\beta u} P_\ell(\tfrac{1-u}{1+u})\, du \end{aligned} \tag{32}$$

The former weak pathology at $w = -1$ has now been removed to $u = \infty$ and rendered quite benign: *Mathematica* now does the integration uncomplainingly, but produces longer and longer strings of incomplete gamma functions and Meijer G-functions.[12,13] Specifically

---

[11] See Abramowitz & Stegun, **4.3.21** & **4.3.22**. These "half-angle formulas" are in obvious conformity with $\sec^2 z - \tan^2 z = 1$.

[12] Than which there are few things more horrible. See Gradshteyn & Ryzhik, *Table of Integrals, Series, and Products* (1965), §**9.3**.

[13] A final change of variable recommends itself: write $u = v^2$. Then $du = 2v\, dv$ and we have

$$g_\ell = \tfrac{1}{2}\sqrt{\beta(2\ell+1)} \int_{-\infty}^{+\infty} \Big[\sqrt{\beta/\pi}\, e^{-\beta v^2}\Big] P_\ell(\tfrac{1-v^2}{1+v^2})\, |v|\, dv$$

The expression in square brackets $\to \delta(v)$ as $\beta \uparrow \infty$, but because of the dangling $\sqrt{\beta}$ leads us not very usefully to an expression of the indefinite form $\infty \cdot 0$.

$$
\begin{aligned}
g_0(\beta) &= \tfrac{1}{\sqrt{4\pi}}\Big\{1\Big\} \quad : \quad \text{all } \beta\\
g_1(\beta) &= \tfrac{1}{\sqrt{4\pi}}\Big\{\sqrt{3}\beta\big[-\beta^{-1}+2\,e^{\beta}\text{Gamma}[0,\beta]\big]\Big\}\\
g_2(\beta) &= \tfrac{1}{\sqrt{4\pi}}\Big\{\sqrt{5}\beta\big[+\beta^{-1}+6-(6+\tfrac{9}{2}\beta)\,e^{\beta}\text{Gamma}[0,\beta]\big]\\
&\qquad\qquad\qquad -\tfrac{3}{2}e^{\beta}\texttt{MeijerG[}\{\{\},\{2\}\},\{\{1,1\},\{\}\},\beta\texttt{]}\Big\}\\
g_3(\beta) &= \tfrac{1}{\sqrt{4\pi}}\Big\{\sqrt{7}\beta\big[-\beta^{-1}-20+(\text{quadratic in }\beta)\,e^{\beta}\text{Gamma}[0,\beta]\\
&\qquad\qquad\qquad +e^{\beta}(\text{linear combination of 3 }\texttt{MeijerG}\text{ functions})\big]\Big\}
\end{aligned}
$$

Mathematica-assisted numerical/graphical analysis of the complicated expressions on the right suggests (if somewhat ambiguously: see below) that

$$\lim_{\beta\uparrow\infty} g_\ell(\beta) = \sqrt{\tfrac{2\ell+1}{4\pi}} \tag{33}$$

Such a result would be in gratifyingly precise conformity with the following point of general principle: Suppose functions $\varphi_n(x)$ are orthonormal and complete on the interval $a \leqslant x \leqslant b$:

$$f(x) = \sum_n \Big\{\int_a^b f(x)\varphi_n^*(x)\,dx\Big\}\varphi_n(x) \quad : \quad \text{all nice functions } f(x)$$

Then

$$\delta(x-x_0) = \sum_n \varphi_n^*(x_0)\,\varphi_n(x) \quad : \quad x\,\&\,x_0 \in [a,b]$$

For spherical harmonics we expect therefore to have this *representation of the "spherical δ-function"*:

$$\delta(\theta-\theta_0)\delta(\phi-\phi_0) = \sum_\ell \sum_{m=-\ell}^{m=+\ell}\big[Y_\ell^m(\theta_0,\phi_0)\big]^* Y_\ell^m(\theta,\phi)$$

Within the space of nice *axially-symmetric* spherical functions $f(\theta)$ the "zonal harmonics" are by themselves complete, and we expect to have

$$\delta(\theta-\theta_0) = \sum_\ell \big[Y_\ell^0(\theta_0)\big]^* Y_\ell^0(\theta)$$

which at the North Pole becomes

$$
\begin{aligned}
\delta(\theta) &= \sum_\ell \big[Y_\ell^0(0)\big]^* Y_\ell^0(\theta)\\
&= \sum_\ell \sqrt{\tfrac{2\ell+1}{4\pi}}\,Y_\ell^0(\theta)
\end{aligned}
\tag{34}
$$

But if we return with (33) to (32)—which in the Gaussian case reads

$$g(\theta,\beta) = \sum_{\ell=0}^{\infty} g_\ell(\beta) Y_\ell^0(\theta) \tag{35}$$

—and if we take into account the notion that

$$\text{spherical Gaussian } g(\theta,\beta) \xrightarrow[\beta\uparrow\infty]{\hspace{3cm}} \text{spherical } \delta(\theta)$$

then we recover precisely (34). Which inspires increased confidence in the accuracy of (33).

But to deposit the distribution $\delta(\theta)$ on the κ-sphere in $\boldsymbol{k}$-space is to identify a plane wave (one that runs parallel to the 3-axis). If our goal is a theory of beams then we have interest in structured "pencils" of $\boldsymbol{k}$-vectors ... must deposit something like a narrow Gaussian distribution (or "fat" δ-function) on the κ-sphere. We are forced, therefore, to look more closely to the structure of the function $g(\theta,\beta)$ with β large but finite. This I for the moment interpret as an obligation to look more closely to the functions coefficients $g_\ell(\beta)$. To reduce the clutter we write

$$g_\ell(\beta) = \sqrt{\tfrac{2\ell+1}{4\pi}}\, G_\ell(\beta)$$

and look to the functions

$$\begin{aligned} G_\ell(\beta) &\equiv \int_0^\infty \beta e^{-\beta u} P_\ell\big(\tfrac{1-u}{1+u}\big)\, du \\ &= \int_0^\infty e^{-t} P_\ell\big(\tfrac{\beta-t}{\beta+t}\big)\, dt \end{aligned} \tag{36}$$

Immediately

$$G_0(\beta) = 1 \quad : \quad \text{all } \beta$$

but already at $\ell = 1$ the situation becomes typically non-trivial: *Mathematica* supplies (as reported already on page 28)

$$G_1(\beta) = -1 + 2\beta e^\beta \Gamma(0,\beta)$$

When we ask *Mathematica* to report the value of $G_1(0)$ she complains

"Indeterminate expression $0\,\infty$ encountered"

though it is an obvious implication of (36) that

$$G_\ell(0) = \int_0^\infty e^{-t} P_\ell(-1)\, dt = (-)^\ell \int_0^\infty e^{-t}\, dt = (-)^\ell$$

Looking to the asymptotic properties of $G_\ell(\beta)$, in which we have greater interest: when we use the command `Series[`$G_1(\beta)$`,{`β`,Infinity,3}]` to ask for the asymptotic behavior as $\beta \uparrow \infty$ we get the response

```
"Essential singularity encountered..."
```

On the other hand, Spanier & Oldham report[14] the "important asymptotic expansion"

$$z^{1-a}e^z\Gamma(a,z) \approx 1 - \frac{1-a}{z} + \frac{(1-a)(2-a)}{z^2} - \frac{(1-a)(2-a)(3-a)}{z^3} + \cdots$$

to which they attach no qualifications, and from which it would follow that

$$G_1(\beta) \approx 1 - 2\beta^{-1} + 2!2\beta^{-2} - 3!2\beta^{-3} + \cdots$$

This result can be reproduced/generalized by the following elementary procedure: noting that the e^{-t} factor discriminates against large values of t, we introduce into (36) the expansion

$$P_1\big(\tfrac{\beta-t}{\beta+t}\big) = 1 - \tfrac{2}{\beta}t + \tfrac{2}{\beta^2}t^2 - \tfrac{2}{\beta^3}t^3 + \cdots$$

and by term-by-term integration obtain

$$G_1(\beta) \approx 1 - \tfrac{2}{\beta} + \tfrac{4}{\beta^2} - \tfrac{12}{\beta^3} + \cdots$$

Similarly

$$\begin{aligned}
P_2\big(\tfrac{\beta-t}{\beta+t}\big) &= 1 - \tfrac{6}{\beta}t + \tfrac{12}{\beta^2}t^2 - \tfrac{18}{\beta^3}t^3 + \cdots\\
P_3\big(\tfrac{\beta-t}{\beta+t}\big) &= 1 - \tfrac{12}{\beta}t + \tfrac{42}{\beta^2}t^2 - \tfrac{92}{\beta^3}t^3 + \cdots\\
P_4\big(\tfrac{\beta-t}{\beta+t}\big) &= 1 - \tfrac{20}{\beta}t + \tfrac{110}{\beta^2}t^2 - \tfrac{340}{\beta^3}t^3 + \cdots\\
&\vdots
\end{aligned}$$

which give

$$\begin{aligned}
G_2(\beta) &\approx 1 - \tfrac{6}{\beta} + \tfrac{24}{\beta^2} - \tfrac{108}{\beta^3} + \cdots\\
G_3(\beta) &\approx 1 - \tfrac{12}{\beta} + \tfrac{84}{\beta^2} - \tfrac{552}{\beta^3} + \cdots\\
G_4(\beta) &\approx 1 - \tfrac{20}{\beta} + \tfrac{220}{\beta^2} - \tfrac{2040}{\beta^3} + \cdots\\
&\vdots\\
G_\ell(\beta) &\approx 1 - \tfrac{\ell(\ell+1)}{\beta} + \cdots
\end{aligned} \tag{37}$$

these remove the ambiguity from (33)

It is interesting that *Mathematica* describes $G_\ell(\beta)$ in closed form, but in terms of such complicated combinations of such fancy functions that it gets confused when trying to evaluate $G_\ell(\beta)$ when β is larger than some relatively small ℓ-dependent critical value (see Figure 11). It does, however, seem to encounter no difficulty when asked to plot $G_\ell(\beta)$ with $0 \leqslant \beta \leqslant 1$.

[14] *An Atlas of Functions* (1987), 45:6:6 page 440.

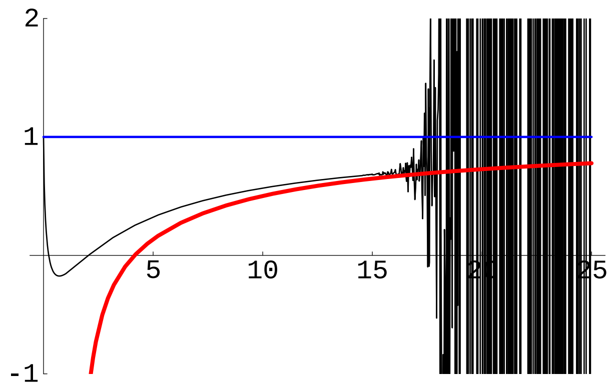

FIGURE 11: *Record of Mathematica's failed attempt to plot its exact analytic description of $G_2(\beta)$, which marks the lowest-order occurance of Meijer G-functions. Shown in red is the a graph of the asymptotic approximant (37), while the unit asymptote is shown in blue. Note that—rather surprisingly—*Mathematica *experiences no difficulty when β is small.*

Suppose we were, on the basis of our imperfect (leading-order asymptotic) knowledge of $G_\ell(\beta)$, to construct (see again (31.1))

$$g(\theta,\beta) \equiv \sum_{\ell=1}^{\infty} g_\ell(\beta) Y_\ell^0(\theta) \tag{38.1}$$

$$g_\ell(\beta) = \sqrt{\tfrac{2\ell+1}{4\pi}}\Big\{1 - \tfrac{\ell(\ell+1)}{\beta}\Big\} \tag{38.2}$$

Then

$$\int_0^\pi g(\theta,\beta)\, 2\pi \sin\theta \, d\theta = g_0(\beta)\sqrt{4\pi} = 1 \quad : \quad \text{all } \beta$$

$$\lim_{\beta\uparrow\infty} g(\theta,\beta) = \sum_{\ell=1}^{\infty} \sqrt{\tfrac{2\ell+1}{4\pi}}\, Y_\ell^0(\theta) = \sum_{\ell=1}^{\infty} \big[Y_\ell^0(0)\big]^* Y_\ell^0(\theta) = \delta(\theta)$$

show that in (38) we have a β-parameterized class of normalized functions that yield the δ-function in the limit $\beta \uparrow \infty$. But what the preceding equations really indicate is that <u>we have labored long and hard to achieve a disappointing result</u>. We have learned that (38) can be written

$$g(\theta,\beta) = \delta(\theta) + \beta^{-1} \cdot (\text{function that averages to zero})$$

The expression the right is not everywhere non-negative (therefore of no interest as a probability distribution) and describes not a "fat" δ-function but a trivially "decorated" δ-function: it presents an explicit δ-function, so it can have nothing useful to contribute to the representation theory of δ-functions, and so far as

I can see it has nothing useful to contribute either to the physical theory of optical beams. The introduction of spherical harmonics has in this instance led us astray. It is as though we had proceeded from the identity

$$\sqrt{\beta/\pi}\, e^{-\beta x^2} = \tfrac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-k^2/4\beta}\cos xk\,dk$$

(note the normalized Gaussian on the left) to write

$$\begin{aligned}
&= \tfrac{1}{2\pi}\int_{-\infty}^{+\infty}\big\{1 - \tfrac{1}{2!}(k^2/4\beta)^2 + \tfrac{1}{4!}(k^2/4\beta)^4 - \cdots\big\}\cos xk\,dk\\
&= \lim_{K\uparrow\infty}\int_{-K}^{+K}\tfrac{1}{2\pi}\cos xk\,dk - \lim_{K\uparrow\infty}\int_{-K}^{+K}\tfrac{1}{64\pi\beta^2}\,k^4\cos xk\,dk + \cdots\\
&= \lim_{k\uparrow\infty}\frac{\sin kx}{\pi x} - \beta^{-2}\bigg\{\lim_{K\uparrow\infty}\int_{-K}^{+K}\tfrac{1}{64\pi}\,k^4\cos xk\,dk + \cdots\bigg\}\\
&= \delta(x) - \beta^{-2}\cdot(\text{complicated “decorations”})
\end{aligned}$$

It is interesting that we have managed to convert one representation of $\delta(x)$ into another

$$\lim_{\beta\uparrow\infty}\sqrt{\beta/\pi}\, e^{-\beta x^2} = \lim_{k\uparrow\infty}\frac{\sin kx}{\pi x} = \delta(x)$$

but in most contexts any argument that would describe representations of delta functions in terms of naked delta functions is an argument that marches in the wrong direction.

The short of it: spherical harmonics may well have something to contribute to the theory of lightbeams, but their introduction at (30) led us astray. Better to work with the Gaussian *left* side of that equation.

6. ?.