DIFFRACTION AND FOURIER OPTICS: THE MATHEMATICS REQUIRED

Anna Consortini
Department of Physics - University of Florence -Via S. Marta 3
50139 Florence, Italy

Abstract

Mathematical needs for Diffraction and Fourier Optics are examined and some suggestions and examples given. The purpose is to suggest the mathematics needed by optics students for both understanding and working with Diffraction and Fourier Optics.

1 -INTRODUCTION

The general question of how much mathematics optics students should know was examined in a paper which I presented at the previous Conference of this series in St Petersburg[1]. In the present paper specific needs for Diffraction and Fourier Optics are examined and some suggestions and examples given. The purpose is to suggest the mathematics needed by optics students for both understanding and working with Diffraction and Fourier Optics.
As already stated in the previous paper[1], by "optics students" I mean University undergraduates and graduates in Physics or Engineering. Some concepts could also be applied to students of other faculties such as Medicine, Biology and so on, and to high school students.
From a mathematical point of view, the study of diffraction means solving Maxwell's equations by taking into account boundary conditions. However this difficult task can be accomplished only in a small number of particular cases, and in most cases one can only obtain approximate solutions more or less consistent with the requirements not only of mathematics, but also of physics. Diffraction can be studied at several different levels and approximations, starting from a very elementary one, based on the simplest form of the Huygens-Fresnel principle, up to the top level where a rigorous theory can be developed in terms of electromagnetic field, as for example Sommerfeld's theory of diffraction from the half plane or Mie theory of diffraction by a sphere. Without entering into the problems of vectors, here use will be made of the so called "optics approximation", which allows one to neglect the vectorial nature of the field and works very well in optics. The mathematics required is different for different levels and, in general, the comprehension of the subject is strictly connected to, and often limited by, mathematics. As an example, we can mention the diffraction of plane waves by openings in screens. While diffraction by a slit can be easy to handle mathematically, even with little knowledge of mathematics, diffraction by a circular aperture requires the more sophisticated mathematics of Bessel Functions. In this paper we will consider the mathematics required for a medium-high level with attention to lower levels and also some information about higher levels. A number of mathematical approximations used in optics will be pointed out, and this will be done, when possible, at the lower level. The starting point is the mathematical analysis of the wave equation and a clarification of the term that is neglected when the geometrical optics approximation is introduced. Fourier Optics will be introduced by means of diffraction in terms of plane wave expansion of the field on the aperture. that also constitutes the basis for the optical elaboration of images, filtering, and holography.
The mathematics considered here is not intended to be exhaustive, but just that required to accomplish the different tasks described.

2 - BASIC REQUIREMENTS AND APPROXIMATIONS

The level considered here implies that the students already know the basics of electromagnetism and Maxwell's equations. Here we assume that medium-high level students have already studied differential and integral calculus, and vector calculus, necessary in the study of the fundamental laws of electromagnetism and in the development of Maxwell's equations. In particular they are

0-8194-1884-6/95/$6.00

assumed to know partial derivatives and a number of theorems of the integral calculus of vectors and of properties which allow one to pass from the integral form to the differential forms of Maxwell's equations in free space or homogeneous media, and to obtain the wave equation for vectors. Students should also know how to describe waves in general, at least by means of trigonometric expressions, and should know some basic concepts such as those of amplitude, phase, wavelength, velocity and so on. They also need to be aware of some different forms of wave equations in different reference systems and their solutions, for future handling of waves of different types.

The optics approximation is in general applicable to diffraction in optics and noticeably simplifies the problem, because it consists of treating the phenomena without vectors, by considering a Cartesian component v(P,t) of the field as representative of the entire field. The student will easily understand that energy is proportional to the square of the component. There are problems that require use of the entire field and therefore cannot be treated by the optics approximation; however they are not of interest here.

Use of the complex form to describe waves is necessary. Although some authors do not use it, but simply the trigonometric form, in my opinion the didactics of this point is greatly simplified by the complex procedure and avoids burdening students with a number of cumbersome trigonometric formulas. The complex form allows one to introduce the time dependence of a monochromatic wave simply by means of a factor $\exp(-i\omega t)$ that constitutes a first step towards wave trains, polychromatic signals and impulses. As already noted in Ref 1, concerning the sign of imaginary unit "i" in the exponent, the minus sign is the standard recent use in optics, although some authors use the plus sign. It could also be useful to note that "-j" can be found in texts dealing with electromagnetic waves.

In addition, the complex form permits an easy separation between space and time variables

2.1 $$v(P,t) = u(P)\ e^{-i\omega t}$$

leaving a complete generality for the complex amplitude u(P). In turn u(P) can be easily specialized to represent the different wave forms; plane, spherical, cylindrical waves are described by their corresponding geometrical surfaces having constant phases. Evanescent waves, or leaky or dissociated waves, can be handled too. Gaussian beams, that derive their name from amplitude, and more precisely the fundamental modes emitted by lasers, can be easily described in terms of Gaussian distribution, a generally well known function even at a low level. When laser beams are studied at a higher level, they require some more mathematics. By the way, we mention here that investigation of laser cavities is a difficult diffraction problem, that has received an approximate solution in finite terms only in some particular cases.

It is appropriate to start the study of diffraction from the wave equation in scalar form by examining the term that is neglected when the ray equation of geometrical optics is derived. This can be a useful example of use of mathematics that gives rise to a physical insight. From this mathematics an introduction can be given to diffraction theory, that greatly helps in explaining to the student when and why diffraction takes place. The scalar form of the wave equation in a homogeneous medium is:

2.2 $$\nabla^2 u(P) + n^2 k_0^2 u(P) = 0$$

where, as usual, ∇^2 denotes the Laplacian operator, u(P) the complex amplitude, n the refractive index, and k_o the wavenumber in empty space, $k_o = 2\pi / \lambda_o$ It is useful to recall that operator ∇ is defined as:

2.3 $$\nabla = \underline{i}\frac{\partial}{\partial x} + \underline{j}\frac{\partial}{\partial y} + \underline{k}\frac{\partial}{\partial z}$$

that application of ∇ to a scalar function gives the gradient of the function, $\nabla f = \text{grad } f$, and that $\nabla \cdot \overline{v} = \text{div } \overline{v}$. Let us write the complex amplitude in terms of amplitude A(P) and phase $\varphi(P)$:

$$u(P) = A(P)\ e^{i\varphi(P)} \tag{2.4}$$

Both A(P) and φ(P) are real quantities. A double application of operator ∇ (and correct use of the rules of derivation of products of functions and of function of function) and subsequent introduction into Eq.2.2, gives rise to two separate equations for the real part and the imaginary coefficient:

$$A\nabla^2\varphi + 2\ (\text{grad}\ A)\cdot \text{grad}\ \varphi = 0 \tag{2.5}$$

and

$$\nabla^2 A + A\left[n^2 k_0^2 - |\text{grad}\ \varphi|^2\right] = 0 \tag{2.6}$$

where A and φ stand for A(P) and φ(P), respectively.

Eq.2.6 is of outmost interest to us. In the derivation of the eiconal equation, that constitutes the mathematical basis for geometrical optics, the first term, $\nabla^2 A$, is required to be small with respect to the second term and therefore negligible. This implies that $\nabla^2 A / An^2k_o^2 << 1$. If we consider a distance equal to $1/k_o$, this inequality tells us that, for geometrical optics to hold, amplitude variations (second differences relative to amplitude) taking place in the space of a wavelength must be negligible in order to neglect diffraction. The key point here is that diffraction takes place where term $\nabla^2 A$ is not negligible, that is when

$$\nabla^2 A / An^2 k_0^2 \geq 1 \tag{2.7}$$

In this case, amplitude variations taking place in the space of a wavelength are not small with respect to the amplitude itself. In other words diffraction takes place where there is an abrupt amplitude variation. This fact can be used as a "definition" of diffraction and can be a useful point to start diffraction even at low levels. This can also help to give an intuitive description of diffraction by considering typically what happens at the borders of any aperture on a screen. Please note, "at the borders". The fact that many diffraction phenomena can be better seen with small apertures is certainly useful for students to make diffraction experiments, but does not justify the error (also found in some textbooks) that diffraction takes place when light crosses an aperture smaller than or equal to the wavelength. The border causes diffraction.

Up to now we have intentionally mentioned diffraction from apertures on screens, but an analogous problem is that of surface obstacles. There is complementarity between these two aspects clarified by Babinet's principle. However diffraction by volume obstacles, or scattering, needs much more attention and suitable mathematics. Here we will not enter into the details of scattering, although it is an important diffraction problem both for its classical theories (such as Rayleigh theory, Mie theory) and for recently discovered phenomena such as enhanced backscattering through random media. This also requires the basics of coherence theory.

Here we mostly refer to coherent light, because diffraction and Fourier Optics are typical of it. However it is worthwhile mentioning that partially coherent light also undergoes diffraction and that some concepts of the mathematics of coherence are also appropriate here.

3 - DIFFRACTION FORMULAS

Let us now consider the diffraction of an aperture on an opaque plane screen illuminated by a wave. For simplicity, and without loss of generality, we assume that the wave is plane and impinges orthogonally on the screen surface, Σ, Fig 3.1. The field diffracted at a point P beyond the screen, obtained by three different approximate theories, is given here below:

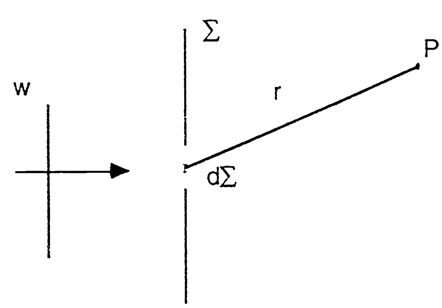

Fig.3.1

3.1 $$u(P) = \frac{-i}{\lambda} A \int_{\Sigma} \frac{e^{ikr}}{r} \, d\Sigma$$ HUYGENS-FRESNEL

3.2 $$u(P) = \frac{-i}{\lambda} \frac{A}{2} \int_{\Sigma} \frac{e^{ikr}}{r} [1 + \cos(\underline{n}, \underline{r})] \, d\Sigma$$ HELMHOLTZ-KIRCHHOFF

3.3 $$u(P) = \frac{-i}{\lambda} A \int_{\Sigma} \frac{e^{ikr}}{r} \cos(\underline{n}, \underline{r}) \, d\Sigma$$ RAYLEIGH-LUNEBERG

Although these expressions differ only in the angular factor, they were obtained from different theoretical formulations: 1) The Huygens-Fresnel formula represents the direct expression of the Huygens-Fresnel Principle, with the phase factor introduced by Fresnel to fit the experimental data. It is historically the first and simplest way to understand diffraction and is considered here due to its simplicity and utility forproducing good results, even for low level students. 2) The Helmholtz-Kirchhoff formula is the result of their, now classical, theoretical formulation and is widely used for many applications. 3) The Rayleigh-Luneberg expression has been reached by several authors, including Sommerfeld. It can also be derived by using the inverse interference principle (Toraldo) or the plane wave expansion (Duffieux). The plane wave expansion is a particular case of the inverse interference principle and will be described here. It is of great interest both because it gives a physical basis to Fourier Optics, (and holography), and also because it requires little mathematics. Here below the mathematics necessary in the three cases and the approximations involved are examined.

3.1-Huygens-Fresnel principle

Formula 3.1 can be immediately understood by students, provided that they know how to write a spherical wave, and realize that the integrand, including the multiplicative constant outside the integral, is the Huygens' secondary wave (wavelet) that originates from the screen aperture, and that the integral represents the interference of the different wavelets coming from the different elements on the surface, as postulated by Fresnel.
As a typical and simple example of how useful the Huygens-Fresnel principle can be for low level students, one can consider the diffraction pattern of a slit. In this case the wavelets are cylindrical waves originating at the slit, with phase factor $-\pi/4$, and the problem becomes a bidimensional one. Let z denote the propagation direction and x the direction orthogonal to the slit. The value u(P) of the field at point $P(x_1,d)$ becomes:

3.4 $$u(P) = \int_{-a/2}^{a/2} \frac{B}{\sqrt{k\rho}} \, e^{ik\rho} \, dx$$

where $a/2$ is the slit half width, B includes the constant part, distance

3.5 $$\rho = \left[(x - x_1)^2 + d^2\right]^{1/2}$$

and the integral must be performed over the slit aperture.
Use of series development of ρ in the exponent allows one to introduce both Fresnel and Fraunhofer approximations of the near and far field, respectively. Although in my opinion these parts should be developed at a later stage, when diffraction has received a more appropriate formulation, they can be usefully understood even at this level and are therefore mentioned here. They also allow us to make some useful considerations on the approximations and different roles played by amplitude and phase in diffraction. If point P is at a distance, d, large with

respect to both x and x_1 one can stop the series for $\sqrt{\rho}$ at the first term, while one has to retain two terms of the series for ρ in the exponent. The difference between the ways amplitude and phase are treated in the series expansion can be suitable for pointing out that, while discontinuities in amplitude determine the phenomenon of diffraction, the main role in deriving the diffracted field from the elementary waves is played by phase, as expected from the Huygens-Fresnel principle.
Stopping the series development of the phase at the second order term constitutes the so called Fresnel approximation:

3.6 $$k\rho = kd\left[1+\frac{(x-x_1)^2}{d^2}\right]^{1/2} \approx kd+\frac{kx^2}{2d}-\frac{kxx_1}{d}+\frac{kx_1^2}{2d}$$

The region where it is valid is called the Fresnel region. The second and third terms are the phase of the integrand in Eq.3.4. Substitution in Eq. 3.4 gives rise to a well known integral:

3.7 $$u(P) \approx \frac{B\ e^{\frac{ikx_1^2}{2d}}}{\sqrt{\lambda d}}\ e^{ikd}\int_{-a/2}^{a/2} e^{-ik\frac{xx_1}{d}}\ e^{ik\frac{x^2}{2d}}\ dx$$

from which it clearly appears that the interference effect involves primarily the phase. Students will learn that, for $x_1=0$, (point P on the axis), the resulting simple integral, can be expressed in terms of its real and imaginary parts, Fresnel's Integrals:

3.8 $$C(s) = \int_0^s \cos\frac{\pi t^2}{2}\ dt \qquad \text{and} \qquad S(s) = \int_0^s \sin\frac{\pi t^2}{2}\ dt$$

Plot of S(s) versus C(s) produces the famous Cornu spiral, a "geometry" exercise for students.

Let us go back to the general case $x_1 \neq 0$. Although Fresnel diffraction is described to students, much more attention is generally paid to Fraunhofer diffraction. Application of Fraunhofer diffraction to the evaluation of the far field of a slit, is a useful exercise for pointing out the mathematical approximations involved and the corresponding physical meaning. Fraunhofer approximation is obtained by a subsequent step in the series development of Eq 3.6, by neglecting the term in x square with respect to the linear term in x. This is possible when the distance, d, is very large with respect to both x and x_1, so that the square terms practically vanish at the borders of the aperture. The corresponding field is called far field. In this case one requires that:

3.9 $$e^{ik\frac{x^2}{2d}} \cong 1 \qquad \text{for} \quad -\frac{a}{2} \le x \le \frac{a}{2}$$

and this requires condition $ka^2/(8d) << 1$ or what amounts to the same

3.10 $$\frac{a^2}{\lambda d} << 1 \qquad \text{Fraunhofer condition}$$

Eq. 3.10 is the very well known Fraunhofer condition. In the region where it is valid, Fraunhofer region, Eq.3.7 becomes

3.11 $$u(P) \approx K\int_{-a/2}^{a/2} e^{-ik\frac{xx_1}{d}}\ dx$$

where complex factor K includes all terms multiplying the integral. Evaluation of this integral is straightforward at any level. Besides, from the mathematical point of view, students can be informed that this integral has the form of a Fourier transform of a constant. Evaluation of the integral gives rise to the Sinc function

3.12 $$u(P) = aK\ \sin\left(\pi\frac{a\,x_1}{\lambda d}\right)\Big/\left(\pi\frac{a\,x_1}{\lambda d}\right) = aK\ \mathrm{Sinc}\left(\pi\frac{a\,x_1}{\lambda d}\right)$$

a well known expression, that can be commented on both in terms of diffraction pattern at infinity ($x_1/d = \vartheta$, the angle under which point P is seen from the slit) and of Fourier transform. And this can be a way of introducing Fourier optics at low level. The proportionality of energy with u(P)u*(P) allows one to analyze in detail the diffraction pattern of the slit without additional mathematics, just looking at the zeros of the Sinc function. These, apart from origin, coincide with the zeros of the sine function. Derivation of the value at the origin is a useful exercise for students for remembering some limit theory. The (angular) position of the first zero

$$\vartheta = \frac{\lambda}{a/2} \tag{3.13}$$

can be utilized here for an introduction to the resolving power of systems.
From the Fraunhofer condition one can derive how large distance d should be for P to be in the far field, in optics. By using λ=0.63 μm (HeNe laser), students will discover, with surprise, that a few meters can be enough, if the aperture is small e.g. 1 mm, and that, due to the square dependence on the aperture, d>> 7 m is needed when a = 2mm.
Coming back to the approximations it is appropriate to note that the Fraunhofer condition, Eq.3.10, is opposite to the condition commonly required for geometrical optics to hold, and that the Fresnel approximation describes an intermediate region between the geometrical optics and the Fraunhofer region.

3.2 -Helmholtz-Kirchhoff theory

Derivation of Eq.3.2 requires a double step. The first one is the exact elegant treatment developed by Helmholtz and generalized by Kirchhoff for deriving the field at a point inside a volume, V, when it is known on a surrounding closed surface, S. The second is the application of the result to the diffraction by an aperture in a screen, a problem which is solved with the help of a number of physical and mathematical approximations, not always consistent with the problem.
The starting point of the theory is the use of Green's formula and Green's method for deriving the field at a point, P, in the volume. It is not the purpose of the present paper to describe in detail the procedure, that can be found in any good book of optics, but we want to underline here some of the main mathematical points. If students do not know Green's formula, it is necessary to introduce it to them. Generally they also do not know the method of Green's function for solving boundary condition problems and it seems advisable to introduce it to them at this stage. by noting that this method is necessary for applications involving sources. It is to be noted that it is not necessary for students to know all the mathematics that lies behind Green's formula and method, but they need a good working knowledge of them; that is they must know not only the 'recipe' of each necessary theorem, but also the conditions for its applicability and validity.
For instance in the case of Green's theorem it is enough for them to know the following:

Let V be a volume, surrounded by a closed surface S. Let u_1 and u_2 be two scalar functions regular in V and on S. (A regular function is continuous and derivable). Green's theorem states that the following formula holds:

$$\int_V (u_1\nabla^2 u_2 - u_2\nabla^2 u_1)\ dV = \int_S \left(u_1\frac{\partial u_2}{\partial n} - u_2\frac{\partial u_1}{\partial n}\right)dS \tag{3.14}$$

where $\partial/\partial n$ denotes the partial derivative in the normal outward direction at a point on S.
In practice the theorem allows one to transform a volume integral into a surface integral.
In the application of the Green method it is also mathematically important to note that a sphere, centered around point P, where the field is to be evaluated, is excluded from the considered volume, Fig 3.2-1.

This choice as well as the choice of the suitable Green function (a spherical wave centered at P) are the basic "tricks" for the formal solution of the problem. It is also interesting to see how the contribution from the spherical surfaces is evaluated and then the field u(P) is derived by making the radius of the sphere tend to zero. At the end of this first part one obtains the exact formula expressing the field at point P in terms of the field and its normal derivative on the surrounding surface:

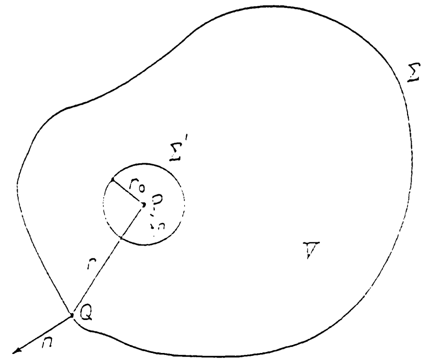

Fig.3.2-1

3.15 $$u(P) = \frac{1}{4\pi}\int_{\Sigma} \frac{e^{ikr}}{r}\left\{\frac{\partial u}{\partial n} - \left(ik - \frac{1}{r}\right) u \cos(\underline{n},\underline{r})\right\} d\Sigma$$

where u =u(Q) at the points of surface Σ and cos($\underline{n}$,$\underline{r}$) is the cosine of the angle between the outward normal $\underline{n}$ and $\underline{r}$ the radius vector from P to Q.

Let us now examine the mathematics for the subsequent second step, for deriving the expression of the field diffracted by an aperture,on a plane screen, which requires a number of assumptions. A suitable surface is chosen, (Fig.3.2-2) constituted by a plane part of the screen and a half-sphere, centered at the aperture and of large radius, so that point P is included in the volume between the two surfaces.

The radius is then made to tend to infinity, and the contribution to the integral from the half-sphere is assumed to be negligible. As shown by Sommerfeld this requires

3.16 $$\lim_{r\to\infty} r\left\{\left(\frac{\partial u}{\partial n} - iku\right)\right\} \to 0$$

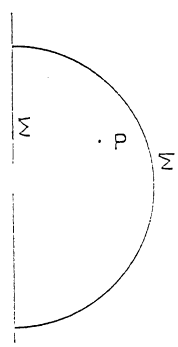

Fig.3.2-2

This expression (Sommerfeld's condition) in practice requires that the field at infinity behaves as a spherical wave, and it is physically reasonable.

On the screen aperture both the field and its normal derivative are assumed to be the same as in the absence of the screen. On the remaining screen part they are assumed to vanish. This is an arbitrary choice because the two quantities are not independent from one another and, in addition, cannot be assigned a priori on the screen surface. As a consequence, the final result, Eq.3.2, is not self consistent, as it immediately appears if point P is made to tend to the screen.

However, in spite of the non physical hypotheses and approximations and inconsistency, Eq.3.2 "works" very well far from the screen, where it is generally used. In addition the distances involved are generally much larger that the aperture dimensions, as for instance in the Fresnel or Fraunhofer approximations, and the angular dependence is generally neglected.

More mathematics is necessary if one goes further to evaluate the radiation pattern of an aperture. A typical example is diffraction from a slit, already examined. More interesting, however, is the diffraction by a circular aperture, illuminated by a plane wave which constitutes a basic step for the study of optical systems. A series development of the distance in the phase term, analogous to that already presented for the slit, allows one to introduce the Fresnel and Fraunhofer approximations in three dimensions. For medium level students these concepts should be introduced here. In the case of Fresnel approximation and observation point Q on the

axis, the integral resulting from Eq.3.2 can be easily evaluated, by means of simple mathematics, giving rise to the well known expression for the Fresnel zones,
More advanced mathematics is required for the Fraunhofer diffraction of the circular aperture, that gives rise to the $J_1(x)/x$ pattern, involving the Bessel function of the first order. If students are not already acquainted with Bessel functions, here they need to learn them and their properties, including their asymptotic behaviour, also because of the importance of the above function in the concept of the resolving power of optical systems. For an insight of the result, and also as a useful exercise, students could analyse a number of properties of the pattern, such as the positions of maxima and zeros, the energy (Airy pattern) flux through the central disc and subsequent rings, and finally the angular width of the central disc, basics for the resolving power.
Let us also note that knowledge of the Bessel function of zero order can also allow one to introduce the so called "non-diffracting beams"
At this point the student should have clear that diffraction takes place at the border of any aperture and could also learn some basis of the ray theory of diffraction, developed by Keller et al. The required mathematics, in the scalar case, is not different from the mathematics required up to now.

4- DIFFRACTION AS DECOMPOSITION IN WAVES (FOURIER TRANSFORM) BY THE INVERSE INTERFERENCE PRINCIPLE

The inverse interference principle allows one to present diffraction in a physical way that easily visualizes the transform properties of diffraction, and introduces students to Fourier optics. It is an alternative way of presenting diffraction with respect to the Helmholtz-Kirchhoff theory, and is of interest here, because it perrmits introduction to Fourier Optics by only requiring the mathematics of the Fourier Transform. It is an approximate theory, but it does not present some of the inconsistencies of the Helmholtz-Kirchhoff formula: it only requires knowledge of the field (and not also of its normal derivative) over the aperture and the result is consistent with this assumption, both on the aperture and on the screen.

The inverse interference principle applies to diffraction from any kind of screen or obstacle, both partially reflecting or partially transparent. Although coherent diffraction by obstacles is of outmost interest for holography, it is not considered here. As this principle is not so widely known we will describe it here. We enunciate it for a partiallly transparent screen of any shape and then we will consider an aperture on a plane screen.

With reference to Fig.4-1, let us consider a screen S illuminated from the left by a field W that produces amplitude $A_\Sigma(P)$ and phase $\varphi_\Sigma(P)$ distributions at points P over the output (right side) surface Σ.
The principle states that if a system of waves outgoing from Σ is found, whose interference over Σ produces the field

4.1 $\quad v_\Sigma(P) = A_\Sigma(P)\, e^{i\varphi_\Sigma(P)}$

these waves are the true diffracted waves.

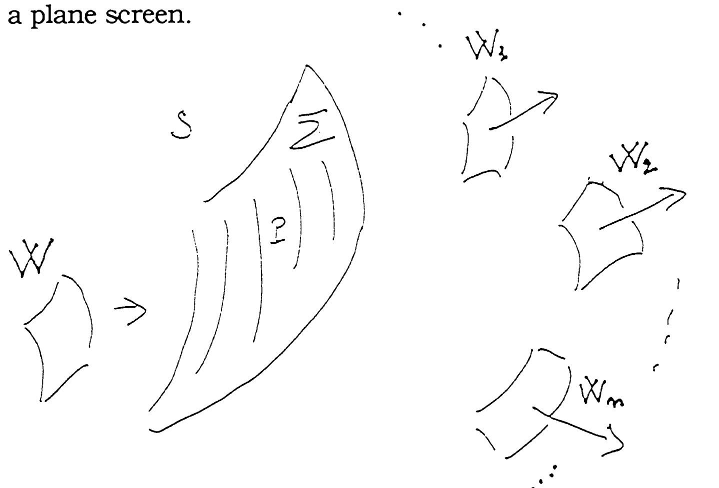

Fig.4.1

The uniqueness of the solution of the boundary problem guarantee that the result is the true one, if the Sommerfeld condition is satisfied. Here the fact that the waves are outgoing from the

surface makes it automatically satisfied. Students should already know this, or they can learn it at this stage.
Up to here there are no approximations, but they are now needed to describe the effect of the screen. Hypotheses about $v_\Sigma(P)$ are now necessary. In general, the effect of a partially transparent screen can be felt on the amplitude, $A_\Sigma(P)$, or on the phase, $\varphi_\Sigma(P)$, or on both. All cases are of interest, for example the phase screen permits investigation of the phase gratings and can be a subsequent useful exercise for students, but now we restrict ourselves to the case of an aperture on a plane screen. Assume that the field on the surface Σ is zero everywhere apart from the aperture, where it is the same as in the absence of the screen. As can be seen in the examples below, knowledge of the field over the aperture is sufficient to solve the problem, because it allows one to find the required wave system. As noted by Toraldo (see:La diffrazione della luce) the choice of the wave system is a problem of eigenvalues and eigenfunctions. In general the solution implies two subsequent steps. The first step is the choice of the values of some parameters, say δ_m, eigenvalues of the problem, on which the orientation and the shape of the waves depend. The second step is the evaluation of the complex amplitudes of the corresponding waves, eigenfunctions. Depending on the aperture one can have a discrete system of eighenvalues and eigenfunctions, or a continuous system or both. As a useful example the student can apply the principle to a periodic aperture and find the spectra diffracted by a grating. This requires use of the Fourier series development, a subject generally well known to students.
Let us consider the non periodic aperture. on the plane screen. The inverse interference rule consists of developing the field diffracted by the aperture in a continuous ensemble (complete including evanescent ones) of plane waves outgoing from the aperture in different directions and requiring that they reproduce the field over the aperture.
As an example let us consider in detail the bidimensional problem (an infinite slit). By using the same geometry as before, let z be the propagation direction, y the axis parallel to the slit and a(x) the complex amplitude on the slit. Let us now consider a plane wave of unit amplitude,

4..2 $$e^{ik(\alpha x+\beta y+\gamma z)}$$

propagating in a direction specified by the direction cosines

4.3 $$\alpha=\cos\theta, \qquad \beta=\cos\gamma, \qquad \gamma=\cos\varphi \qquad\qquad (\alpha^2+\beta^2+\gamma^2=1)$$

and construct a continuous system of plane waves with no dependence on y, because in our case $\beta=0$. Let us consider an elementary angle $d\varphi$. Let $Ad\varphi$ be the amplitude of all waves having propagation direction between φ and $\varphi+d\varphi$. One has

4.4 $$\alpha = \sin\varphi, \quad d\alpha = \cos\varphi \; d\varphi; \quad d\varphi=\frac{d\alpha}{\sqrt{1-\alpha^2}}$$

The field at a point x, z is obtained by integrating all waves

4.5 $$v(x,z)=\int \frac{A\, e^{ik(\alpha x+\gamma z)}}{\sqrt{1-\alpha^2}}\, d\alpha$$

where for a moment the integration limits are not defined. On the aperture the field must concide with a(x):

4.6 $$v(x,0) = a(x)$$

Let us write a(x) by its Fourier transform $\overline{A}(f)$

4.7 $$\overline{A}(f)= \int_{-\infty}^{\infty} a(s)\, e^{-2\pi i f s} ds$$

one has

4.8 $$a(x)= \int_{-\infty}^{\infty} e^{2\pi i f x}\overline{A}(f)df= \int_{-\infty}^{\infty} e^{2\pi i f x}\left(\int_{\infty}^{\infty} a(s)\, e^{-2\pi i f s} ds\right)df$$

By putting

4;9 $f=\alpha/\lambda$

one has

4.10 $$a(x)=\frac{1}{\lambda}\int_{-\infty}^{\infty} e^{2\pi i\frac{\alpha}{\lambda}x}\left(\int_{-\infty}^{\infty} a(s)\, e^{-\frac{2\pi i}{\lambda}\alpha s}ds\right)d\alpha$$

On the aperture the integral of plane waves, Eq.4.5, must give

4.11 $$v(x,0)=\int_{-\infty}^{\infty}\frac{A}{\sqrt{1-\alpha^2}}e^{ik\alpha x}d\alpha$$

and must therefore have infinite limits. This resut shows that, in addition to real waves, there are also evanescent waves (those for which α>1). Comparison of Eq.4.10 and 4.11 gives the complex amplitude, A=A(α):

4.12 $$A(\alpha)=\frac{\sqrt{1-\alpha^2}}{\lambda}\int_{-\infty}^{\infty}a(x)\, e^{-ik\alpha x}dx$$

In terms of φ $A(\alpha)\rightarrow A(\varphi)$

4.13 $$A(\varphi)=\frac{\cos\varphi}{\lambda}\int_{-\infty}^{\infty}a(x)\, e^{-ik\sin\varphi x}dx$$

Let us denote by

4.14 $$A(f)=\frac{\lambda}{\sqrt{1-\alpha^2}}A(\alpha)=\int_{-\infty}^{\infty}a(x)\, e^{-i2\pi fx}dx$$

It appears that A(f) is the Fourier transform of a(x). By recalling that $f=\frac{\alpha}{\lambda}=\frac{\sin\varphi}{\lambda}$ immediately one sees that the amplitude, A(f), of the wave diffracted in the direction φ is the component at frequency $f = \sin\varphi/\lambda$ of the Fourier transform of a(x). For the evanescent waves it is better to consider quantity α, and the result is the same. In other words the system of diffracted waves is the Fourier Transform of the field a(x) on the screen. Diffraction is a Fourier Transform. A(f) is the complex amplitude of the wave diffracted in the direction φ, that is the field at a point at infinity in that direction. If the aperture is limited, it is an easy matter now to derive the far field of the slit, the Sinc function. This can also be a way for the student to see that a limited field has an infinitely extended spectrum and viceversa and to be introduced to the properties of the transform.

If a lens is placed behind the aperture, the lens transfers the field of the point at infinite in direction φ to a point y=F tanφ in the focal plane. This could lead one to think that the lens makes the Fourier transform of the field. This however is not true, because a lens, even of large extension, cannot collect all waves, in particular evanescent waves. In the angular frequency space, (that of the Transforms) the lens carries out a filtering operation because it collects only the frequencies corresponding to a maximum angle. In the image space this corresponds to a convolution. It immediately appears here that students should know not only the Fourier transform but also a number of properties and theorems, including the convolution theorem and Parceval theorem.

Up to here we have considered the bidimensional case, however the tridimensional one is a straightforward generalization. giving rise to a bidimensional Fourier transform, where the two angular frequencies of the transform are now given by:

4.15 $$f_x=\frac{\alpha}{\lambda} \quad \text{and} \quad f_y=\frac{\beta}{\lambda}$$

Derivation of the field diffracted at infinity as well as of Eq. 3.3 is strigthforward.

We also note that this method allows one to easily explain the holographic procedure of wavefront reconstruction. The diffraction from apertures illuminated by partially coherent light can also be easily derived.

5-CONCLUDING REMARKS

Some additional comments on the diffraction for students who will go on to research. It is important that they know that diffraction results are only approximate. They should also know that the different approximate formulas, that have been almost always found to work very well, can sometimes fail when they are used at the limits of their validity. This happened to me in the investigation of open cavities having a low Fresnel number, as described in Ref 1.
As is clear from the name, Fourier optics requires an adequate knowledge of Fourier Transforms, a subject that can be easy handled in the case of Engineering and Physics students and is less easy for students of other faculties (e.g. medicine, biology and so on). By adequate knowledge we mean that students need not only to know the definitions, but to become familiar with a number of properties and theorems (such as sampling theorem, convolution etc..) that are the basis for many practical applications. The ideal preparation on this subject should include the theory of complex variable functions and the theory of distributions (or generalized functions) that allow the student to handle any type of mathematical description of optical signals without trouble about or doubts on the validity of the results. It is hard to exaggerate the importance for students to have a clear mathematical understanding of them in order to be able to handle optics in the laboratory. A number of typical Fourier Transforms that are commonly used can be derived and understood by students. In addition, they are required to know a number of related theorems and to be able to see their practical application. Convolution, and the relation between the spectra in a convolution are basic for understanding the image formation process, the limits in the resolving power of optical systems, the holography and the methods of optical information processing, such as filtering or image enhancement.

LIST OF USEFUL BOOKS

M.Born and E. Wolf, "Principles of Optics", Pergamon Press. Now York IV Edition, 1970.

P.M. Duffieux, "The Fourier Transform and its Application to Optics", Wiley Series in Pure and Applied Optics, John Wiley and Sons Inc., New York, 1983.

J.W. Goodman, "Introduction to Fourier Optics", Mc.Graw Hill, 1968.

A.Sommerfeld, "Optics", Lectures on Theoretical Physics Vol.IV. Academic Press, 1964.

G. Toraldo di Francia, "Electromagnetic Waves", Interscience Publishers Inc. New York, 1953. See also G. Toraldo di Francia "La Diffrazione della Luce", Edizioni Scientifiche Einaudi, 1958, and G. Toraldo di Francia and P. Bruscaglioni "Onde Elettromagnetiche" Zanichelli, 2nd Edition, 1988.

REFERENCE

1. A. Consortini: "How much mathematics should optics students know?" *Education in Optics*, Grigori B. Altshuler, Brian J. Thompson , Editors, Proc. SPIE 1603, 98-108, (1992)