

付立叶光学参考资料

辽宁大学物理系光学教研室

一九八四年七月

傅里叶变换光学基本原理讲座

第一讲 光波衍射与波前变换

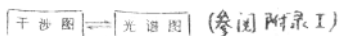
钟 锡 华 (北京大学物理系)

“傅里叶变换光学原理”讲座拟分以下八讲连载:

- 第一讲 光波衍射与波前变换;
- 第二讲 相因子判断法;
- 第三讲 正弦光栅的衍射;
- 第四讲 阿贝成像原理与相衬显微镜;
- 第五讲 夫琅和费衍射实现屏函数的傅里叶变换;
- 第六讲 空间滤波与光信息处理;
- 第七讲 相因子判断法分析全息图的衍射场;
- 第八讲 点扩展函数与光学传递函数。

在物理学的几门基础学科中,光学几乎与力学一样的古老,相对说来电磁学要年轻得多。然而从本世纪四十年代后期开始的三十余年间,光学在理论方法上和实际应用上都有许多重大的突破和进展。因此,“现代光学”的提法目前颇为流行。1948年全息术的提出,1955年作为象质评价的传递函数的兴起,1960年新型光源——激光器——的诞生,它们是现代光学中有重要意义的三件大事。连同六十年代以后由于有了激光的重新装备而迅速发展起来的非线性光学、薄膜光学、纤维光学、集成光学等光学诸方面,使光学这门历史悠久的学科焕发了青春,它正以自身的深刻的变革和日益扩展的应用领域,引人注目地活跃在现代物理学和现代科学技术的广阔舞台上。

现代光学的重大进展之一是引入“傅里叶变换”的概念,由此逐渐发展出光学的一个新分支——傅里叶变换光学,简称变换光学,或傅里叶光学。目前的变换光学大体指两类内容。一是傅里叶光谱仪中存在的那类变换关系:



它从干涉强度的空间频谱中提取光波辐射的时间频谱(即通常说的光谱)。另一类是相干成像系统和不相干成像系统中存在的变换关系:



这一类光学变换的内容相当丰富,它包括光学空间滤波和信息处理,光学系统的脉冲响应和传递函数,波前再现和全息术等等。变换光学的基本思想是用空间频谱的语言分析光信息,用改变频谱的手段处理相干成像系统中的光信息,用频谱谱改变的眼光评价不相干成像系统(光学仪器)的象质。

傅里叶变换光学的基本规律并未超出传统波动光学的范围,它仍然是以波动光学的原理为基础,它是干涉和衍射的综合和提高,它与衍射特别是与夫琅和费衍射息息相关。因此,从衍射问题开始,逐步引出变换光学的概念和方法,并联系光学变换的若干应用,就是这期讲座的主要线索。

讲座内容是在赵凯华教授的直接指导下编写的,并经他修改定稿。

$$U(P, t) = A(P)e^{i(q(P) - \omega t)} \\ = A(P)e^{iq(P)}e^{-i\omega t} \quad (1.2)$$

一、定态光波的复振幅描述

具有如下性质的波场叫定态波场:

(1) 空间各点的扰动是同频率的简谐振荡 (频率与振源相同);

(2) 波场中各点扰动的振幅不随时间变化, 在空间形成一个稳定的振幅分布。

严格的定态光波要求波列无限长, 但任何实际光源的发光过程总是有限的, 特别从微观角度看, 发光过程是断断续续的, 有限波列不可能是严格单色的。不过当波列的持续时间比扰动的周期长得多时, 除了考虑某些特殊问题 (如时间相干性) 外, 我们可把它作无限长单色波列处理, 这样的波在空间传播时形成定态波场。今后, 如果没有特别的必要, 我们一律以定态光波为讨论对象。

普遍的定态标量波的表达式为

$$U(P, t) = A(P)\cos[\omega t - \varphi(P)], \quad (1.1)$$

其中 P 代表场点, 函数 $A(P)$ 反映振幅的空间分布, $\varphi(P)$ 反映位相的空间分布, 二者都与时间 t 无关。波函数 $U(P, t)$ 中唯一与 t 有关的是位相因子中独立的一项 ωt (ω 为圆频率), 这项又是与场点坐标无关的。

用复数来描述简谐振动, 读者是应当逐渐熟练起来的。其办法是用一个复指数函数与余弦 (或正弦) 函数对应。这样做的依据是它们的运算规律 (迭加, 微分和积分) 是对应的, 用复数运算来代替简谐量的运算曾给我们带来极大的方便。

定态波场中各点的扰动是同一频率的简谐振动, 我们同样可将它的表达式用一个对应的复数式代替。例如对于式 (1.1), 有如下对应关系:

$$U(P, t) = A(P)\cos[\omega t - \varphi(P)] \Leftrightarrow$$

$$\tilde{U}(P, t) = A(P)e^{\pm i[\omega t - \varphi(P)]},$$

式中指数上的正负号代表两种不同的选择, 运算时可采用任何一种, 它们实质上完全等效。纯粹由于习惯 (也许还有某些计算上的方便), 我们选用负号, 即

可以看出, 在上式中包含时间和空间变量的两部分完全分离, 成为独立的因子。在讨论单色波场中各点扰动的空间分布时, 时间因子 $e^{-i\omega t}$ 总是相同的, 常可略去不写, 剩下的空间分布因子

$$\tilde{U}(P) = A(P)e^{i\varphi(P)} \quad (1.3)$$

称为复振幅。复振幅 $\tilde{U}(P)$ 由两部分组成, 其模量 $A(P)$ 代表振幅在空间的分布, 其辐角 $\varphi(P)$ 代表位相在空间的分布。复振幅则把定态波场中两项空间分布统一表达在一起, 概念引入的优点表现得很明显了。

下面我们写出平面波和球面波的复振幅分布函数。平面波函数为

$$\tilde{U}(P, t) = A\cos(\omega t - \mathbf{k} \cdot \mathbf{r} - \varphi_0) \\ = A\cos(\omega t - k_x x - k_y y \\ - k_z z - \varphi_0),$$

故平面波的复振幅函数为

$$\tilde{U}(P) = A\exp[i(\mathbf{k} \cdot \mathbf{r} + \varphi_0)] \\ = A\exp[i(k_x x + k_y y + k_z z + \varphi_0)].$$

其特点是振幅 A 是常数, 与场点坐标无关; 位相因子是场点直角坐标的线性函数 (线性相因子)。今后我们遇到的问题, 不单是已知一平面波, 写出其复振幅的表达式, 而往往是反过来, 得到某一复振幅函数, 判断它是否为平面波, 以及是怎样的平面波。上述两个特点将是我们判断的依据之一。平面波波矢 $\mathbf{k}$ 的数值和方向余弦可由线性相因子的系数定出:

$$k = \frac{2\pi}{\lambda} = \sqrt{k_x^2 + k_y^2 + k_z^2}, \\ \cos \alpha = \frac{k_x}{k}, \quad \cos \beta = \frac{k_y}{k}, \\ \cos \gamma = \frac{k_z}{k}.$$

球面波函数为

$$\tilde{U}(P, t) = \frac{a}{r} \cos(\omega t - kr - \varphi_0),$$

故球面波的复振幅函数为

$$\tilde{U}(P) = \frac{a}{r} \exp[i(kr + \varphi_0)],$$

这里 r 是场点 P 到振源的距离。若采用直角坐标系, 设振源在 x_0, y_0, z_0 位置上, 则有

$$\tilde{U}(x, y, z) = \frac{a}{\sqrt{(x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2}} \times \exp[i(k\sqrt{(x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2} + \varphi_0)], \quad (1.4)$$

可见, 除非波源在原点, 球面波复振幅的函数形式是比较复杂的。球面波的特点是振幅与场点到点源的距离 r 的一次方成反比。下面我们将看到, 球面波的相因子在一定的近似条件下可以简化为只含线性项和二次项。

必须指出, 经过光学系统或屏障因衍射而产生的光波场通常是比较复杂的。但是, 在所有多种多样的波场类型中, 球面波和平面波占有特别重要的地位。这一方面是因为它们比较简单, 从而也被研究得比较透彻; 另一方面是因为任何形状的波面可以看作是次波点源的集合, 任何形状的扩展光源同样可以看作是发光点源的集合。点光源之于光学, 正象质点之于力学, 点电荷之于电学一样, 是建筑整个理论体系的基石。现代变换光学的基本思想就是要在复杂的波场中分离出简单的成分——球面波或平面波。从这个意义上看, 定态光波的复振幅描述, 不仅是为了简化同频扰动的迭加运算, 而且更为重要的是便于识别波场和分解波场。

二、惠更斯-菲涅耳原理的实质

与波前函数的意义

“波前”一词, 在现代光学中颇为流行。它过去常指的是一个等相面(波面), 现已泛指波场中任一曲面, 不一定是等相面, 通常用的往往是指一个平面, 如记录介质、感光底片、接收屏幕、透明的黑白画面等所在的平面, 或透镜前后的某个平面。在实际问题中人们往往不必泛泛地讨论三维波场里复振幅的分布, 也无需求复杂波场中波面的形状和波线的轨迹, 而只关

心某一特定波前上复振幅的二维分布。一列波携带着许多信息, 如频率 ω 、波长 λ 和传播方向(二者包含在波矢 k 中), 振幅分布, 位相分布, 传播速度, 等等。对于单色的定态波场, 这些信息全部包含在三维的复振幅分布函数中了。我们要了解波前的复振幅分布, 即波前函数 $\tilde{U}(x, y)$ 。须知, 光波与物质的相互作用正是通过波前来实现的。波前的理论意义可由惠更斯-菲涅耳原理中清楚地看出。

用简短的文字概括起来, 作为衍射理论基础的惠更斯-菲涅耳原理(1818年)可表述如下: 波前 Σ 上每个面元 $d\Sigma$ 都可以看成是新的振动中心, 它们发出次波。在空间某一点 P 的振动是所有这些次波在该点的相干迭加。可参看图 1.1。

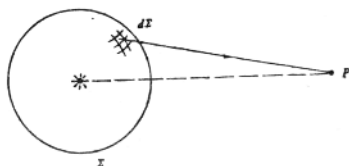


图 1.1 惠更斯-菲涅耳原理

既然是相干迭加, 即可利用复振幅的概念。设 $d\tilde{U}(P)$ 是由波前 Σ 上的面元 $d\Sigma$ 发出的次波在场点 P 产生的复振幅, 则在 P 点的总扰动应为

$$\tilde{U}(P) = \iint_{(\Sigma)} d\tilde{U}(P), \quad (1.5)$$

这可以说是惠更斯-菲涅耳原理的数学表达式。不过要用它来计算, 还需进一步具体化。合理的假设应是:

$$d\tilde{U}(P) \propto d\Sigma; \quad (1.6)$$

$$\propto \tilde{U}_d(Q); \quad (1.7)$$

$$\propto \frac{e^{ikr}}{r}; \quad (1.8)$$

$$\propto F(\theta_0, \theta) \quad (1.9)$$

式(1.6)中 $d\Sigma$ 是面元的面积, 式(1.7)中 $\tilde{U}_d(Q)$

是面元(次波源)上 Q 点的复振幅, 取其等于从波源自由传播到 Q 时的复振幅. 式(1.8)说明次波源发射的是球面波, 其中 r 是面元 $d\Sigma$ 到场点 P 的距离, $k = \frac{2\pi}{\lambda}$. 式(1.9)中的 θ_0 和 θ

分别是源点 S 和场点 P 相对次波面元 $d\Sigma$ 的方位角(见图 1.2), $F(\theta_0, \theta)$ 是 θ_0 和 θ 的某个函数, 它称为倾斜因子, 它表明由面元发射的次波不是各向同性的. 根据以上各条假设, 式(1.5)可写成

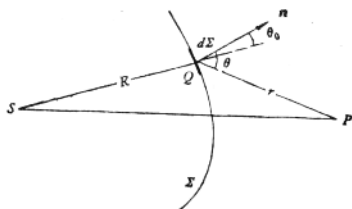


图 1.2 关于菲涅耳衍射公式中各量的说明

$$\tilde{U}(P) = K \iint \tilde{U}_0(Q) F(\theta_0, \theta) \frac{e^{ikr}}{r} d\Sigma, \quad (1.10)$$

式中 K 是个待定的比例常数. 式(1.10)称为菲涅耳衍射积分公式. 最初菲涅耳作上列各假设时只凭朴素的直觉. 六十余年后, 基尔霍夫(1882年)从定态波场的亥姆霍兹方程出发, 利用数学上的格林公式导出了场点的光扰动为

$$\begin{aligned} \tilde{U}(P) = & \frac{-i}{\lambda} \iint \frac{1}{2} (\cos \theta_0 + \cos \theta) \\ & \times \tilde{U}_0(Q) \frac{e^{ikr}}{r} d\Sigma, \end{aligned} \quad (1.11)$$

同时指出, 凡是隔离实际点光源与场点的任意闭合面都可以作为波前, 它不一定是等相面(如图 1.3). 式(1.11)称为菲涅耳-基尔霍夫衍射公式, 它与菲涅耳当初由朴素的物理思想所构造的衍射积分相比较, 两者的主体部分是一致的, 只是倾斜因子有所不同, 比例系数更为明确. 不过在光孔和接收范围满足傍轴的条件下, 倾斜因子的影响可以忽略, 衍射积分公式简化为

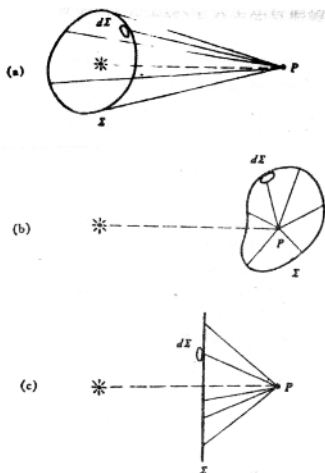


图 1.3 可选作菲涅耳-基尔霍夫衍射公式积分面的各种波前

$$\tilde{U}(P) = \frac{-i}{\lambda r_0} \iint \tilde{U}_0(Q) e^{ikr} d\Sigma, \quad (1.12)$$

其中 r_0 是衍射屏中心到场点的距离. 利用此式计算衍射场的具体分布正是以后几讲要讨论的重点问题之一. 在这里首先强调, 惠更斯-菲涅耳原理的实质是无源空间边值定解. 这是因为该原理告诉人们, 实际光源的作用可以由波前上大量次波源的集体贡献来等效, 而在波前与场点所包围的空间中是无源的. 这类无源空间的边值定解具有物理上的唯一性, 即边界上的波前函数 $\tilde{U}_0(Q)$ 唯一地决定光场. 由此可以直接导出两个重要结论:

(1) 一旦波前函数 $\tilde{U}_0(Q)$ 再现, 即使原物不复存在, 根据边值定解的唯一性, 可以断定无源空间中的光场也必将再现, 这时光场中的一切观测后果就如同实物存在时那样真切. “再现波前必将再现一切观测后果”这个观念, 本来就寓于惠更斯-菲涅耳原理之中, 它是人们领悟奇异的全息术的一个关键概念.

(2) 一旦波前函数 $\tilde{U}_0(Q)$ 发生了变化, 则无源空间中的光场就发生重新分布, 由此给出光波衍射的一个较普遍的定义.

对于“衍射”问题，我们曾有过几种不同深度的认识。最初我们说，当光在传播过程中遇到障碍物时发生偏离直线传播，或更广泛一些，偏离几何光的传播行为，这种现象叫做衍射。在把惠更斯-菲涅耳原理运用到圆孔、圆屏、单缝、多缝等衍射问题后，我们会意识到，衍射的发生，是由于光在传播过程中波面受到某种限制，亦即自由波面发生破损。现在我们要说，当光在传播过程中，由于种种原因而改变了波前的复振幅分布（包括振幅分布和位相分布），后场不再是自由传播时的光波场，这便是衍射。以上各种说法都是可取的，它们反映了人们对衍射现象的认识在逐步深入，其中最后一种说法对衍射现象因果关系的概括更为普遍和本质，它也是从菲涅耳-基尔霍夫衍射积分公式所能作出的最广的直接推论。

三、衍射系统及其屏函数

凡能使波前上复振幅发生改变的物，统称衍射屏。衍射屏可以是反射物，也可以是透射物。透射式衍射屏有圆孔、矩孔、单缝等一类，有小球、细丝、玻璃上的墨点、小颗粒等一类，有黑白光栅、菲涅耳波带片等一类，有一幅景物的底片、一张图象、一页数码字符等一类，也有如透镜这类透明的位相型衍射物。

以衍射屏为界，整个衍射系统被分成前后两部分。前场为照明空间，充满照明光波场；后场为衍射空间，充满衍射光波场。一般说来，照明光波比较简单，它常是球面波或平面波，这两种典型波的等相面和等幅面是重合的，属于均匀波，在其波场中没有因强度起伏而出现的亮暗图样。衍射波则比较复杂，它不是单纯的球面波或平面波，这种复杂波的等相面和等幅面一般不重合，属于非均匀波，波场中常有因强度起伏形成的衍射图样。

在整个衍射系统中，我们特别考虑二个波

前的场分布。如图 1.4 所示，衍射屏之前侧是照明光波前 $\tilde{U}_1(x, y)$ ，它称为入射场，衍射屏之后侧是衍射光波前 $\tilde{U}_2(x, y)$ ，它称为透射场（或反射场）；最后还有一个衍射波向前传播的接收场 $\tilde{U}(x', y')$ 。把波前 $\tilde{U}_1(x, y)$ 转化为波前 $\tilde{U}_2(x, y)$ 是衍射屏的作用，从一个波前 $\tilde{U}_1(x, y)$ 导出另一个波前 $\tilde{U}(x', y')$ 是光的传播的基本问题。两步合起来成为衍射，所以衍射就是改变波前和变换波前。

衍射屏的作用可集中地用如下一个函数来表征：

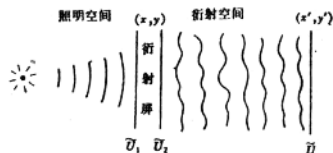


图 1.4 衍射系统中的三个波前

$$\tilde{\tau}(x, y) = \frac{\tilde{U}_2(x, y)}{\tilde{U}_1(x, y)}, \text{ 即 } \tilde{U}_2 = \tilde{U}_1 \tilde{\tau}. \quad (1.13)$$

对于透射屏，它称为复振幅透射率函数；对于反射屏，它称为复振幅反射率函数。二者统称屏函数。显然，直接决定衍射场 $\tilde{U}(x', y')$ 的是波前 $\tilde{U}_2$ ，它也就是各衍射积分公式 (1.10) — (1.12) 中的 $\tilde{U}_0(Q)$ 。

屏函数一般也是复数，它包括模和辐角两部分。 $\tilde{\tau}(x, y)$ 的辐角为常数的衍射屏称为振幅型的； $\tilde{\tau}(x, y)$ 的模为常数的衍射屏称为位相型的。

任何形状的孔或遮光屏是最简单的振幅型衍射屏，其屏函数的形式为

$$\tau(x, y) = \begin{cases} 1, & \text{透光部分,} \\ 0, & \text{遮光部分.} \end{cases}$$

透镜则是最常见的位相型屏函数。关于透镜和棱镜的屏函数的具体形式，留待下一讲（相因子判断法）再讨论。

1.2 Huygens' Principle

A concise treatment of the history of the development of diffraction theory is given by Born and Wolf [1-6] and more detailed treatments of the theory are available for those interested [1-7 through 1-9]. Only a brief introduction is given here. The account begins in 1678 when Christian Huygens postulated that each point on a wavefront could be regarded as the source of a secondary spherical wave. The position and form of the wave at a later time is then described by the envelope of the secondary waves as shown in Fig. 1-3. Huygens' concepts were incorporated into mathematical form by Augustin Fresnel in 1818. Fresnel predicted diffraction patterns, later experimentally verified, which could not be explained by particle theory. This work led to the acceptance of the wave theory of light over the particle theory. The mathematical description was later further refined by Kirchhoff, Rayleigh, and Sommerfeld. The mathematical techniques have

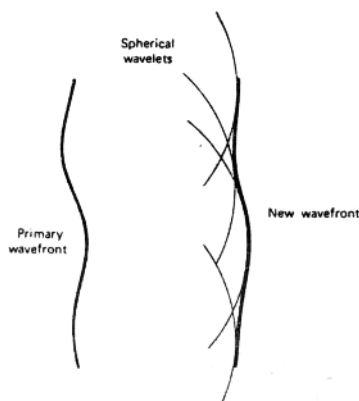


Figure 1-3 The envelope of the secondary waves arising from the primary wave forms the new wave.

not yet been perfected, with most difficulty arising because of the boundary conditions. The boundary condition problems have been studied since 1888 by Maggi [1-10] up to the recent work of Rubinowicz, Wolf, and co-workers [1-11, 1-12].

We have obtained the expression (1-25) by considering the propagation of the plane wave components of the distribution U_1 . The same equation can be developed using Huygens' assumption that each point of the distribution radiates a wavelet uniformly in all directions. Its wavefront is thus spherical and is described by $\exp(ikr)$. The total effect in the second plane is the superposition of the waves from all points in the first plane, being proportional to

$$\iint U_1(x_1, y_1) \exp(ikr) dx_1 dy_1 \quad (1-26)$$

where $U_1(x_1, y_1)$ denotes the amplitude and phase of the wave radiated from the point (x_1, y_1) , and the constant of proportionality has been shown to be $1/i\lambda z$. (See, for example [1-3].)

Reference to Fig. 1-4 shows that r is given as

$$r^2 = z^2 + (x_1 - x_2)^2 + (y_1 - y_2)^2, \quad (1-27)$$

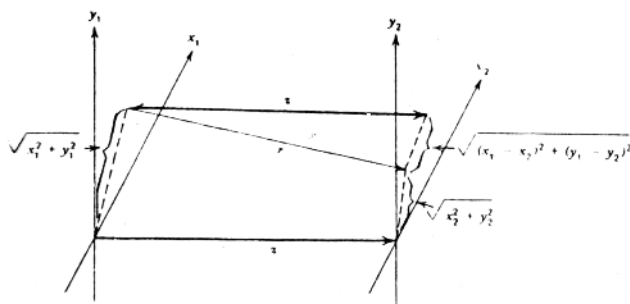


Figure 1-4 The relationship between the coordinates and the distance between a point on the x_1 - y_1 plane and a point on the x_2 - y_2 plane.

or equivalently as

$$r = z \left\{ 1 + \frac{(x_1 - x_2)^2 + (y_1 - y_2)^2}{z^2} \right\}^{1/2} \quad (1-28)$$

The binomial expansion can be used to obtain

$$r \approx z + \frac{1}{2z} [(x_1 - x_2)^2 + (y_1 - y_2)^2] \quad (1-29)$$

where the higher-order terms can be dropped as being negligible if the region of interest is restricted to be such that

$$z \gg |x_1 - x_2|, \quad z \gg |y_1 - y_2|. \quad (1-30)$$

Substituting the approximation (1-29) in (1-26) results in the same expression, (1-25), that was derived by the use of the angular spectrum. In this approach, the physical interpretation of the approximation is that only a portion of the spherical wave from each point is of concern. Consequently, a quadratic approximation to the spherical surface is adequate.

The diffraction formula. Equation (1-25) can be rewritten with the x_2 and y_2 terms outside of the integral as

$$U_2(x_2, y_2) = \frac{1}{i\lambda z} \exp(ikz) \exp\left(ik \frac{x_2^2 + y_2^2}{2z}\right) \times \int \int U_1(x_1, y_1) \exp\left(ik \frac{x_1^2 + y_1^2}{2z}\right) \exp\left(-ik \frac{x_1 x_2 + y_1 y_2}{z}\right) dx_1 dy_1. \quad (1-31)$$

If the value of k is written as $2\pi/\lambda$ in the last exponential of (1-31),

$$U_2(x_2, y_2) = \frac{1}{i\lambda z} \exp(ikz) \exp\left(ik \frac{x_2^2 + y_2^2}{2z}\right) \times \int \int U_1(x_1, y_1) \exp\left(ik \frac{x_1^2 + y_1^2}{2z}\right) \exp\left(-i2\pi \frac{x_1 x_2 + y_1 y_2}{\lambda z}\right) dx_1 dy_1 \quad (1-32)$$

takes the form of a *Fourier transform*. That is, the integral represents the Fourier transform of $U_1(x_1, y_1) \exp[ik(x_1^2 + y_1^2)/2z]$ and the transformation is from the x_1, y_1 domain to the $x_2/\lambda z, y_2/\lambda z$ domain. Alternatively, the integral of (1-32) can be viewed as a *Fresnel transform* of $U_1(x_1, y_1)$. Let us further examine the significance of the term $\exp[ik(x_1^2 + y_1^2)/2z]$.

1.3 Fresnel Region

In the inevitable effort to simplify the mathematics, the term

$$\exp\left(ik \frac{x_1^2 + y_1^2}{2z}\right) \quad (1-33)$$

of (1-32) comes under scrutiny. Obviously if $U_1(x_1, y_1)$ is nonzero over only a small region, then x_1^2 and y_1^2 can be small in comparison to λz , and the term (1-33) can be neglected. If this is not the case, by reason of either a large extent of U_1 or a small value of z , the x_2, y_2 plane is said to be in the *Fresnel region* of the distribution $U(x_1, y_1)$. The Fresnel region is then the region where z is large enough, with respect to the extent of U_1 and U_2 , that the higher-order terms of (1-29) can be neglected but where λz is small enough with respect to the square of the extent of U_1 that the exponential term (1-33) cannot be neglected. This region is of importance in optical imaging and holography.

1.4 Fraunhofer Region

The point at which the term in (1-33) can be neglected is easier to define when U_1 is bounded by an aperture. The limits on the size of the aperture then give a direct means of comparing the maximum values of x_1^2 and y_1^2 with λz . Even so, the exponential term is not suddenly negligible, and different numbers have been given as the values of λz beyond which the exponential (1-33) can be ignored. Consideration of the integral of (1-32) shows that the exponential (1-33) is negligible if its phase change is negligible over the range of nonzero or nonnegligible values of U_1 . It is assumed that this is so if the phase change over the range of integration is

much less than 1 radian. If the maximum value of $x_1^2 + y_1^2$ is $(d/2)^2$, where d is the linear dimension of the maximum extent of U_1 , the restriction on the phase leads to

$$\frac{k(d/2)^2}{2z} = \frac{\pi d^2}{4\lambda z} \ll 1. \quad (1-34)$$

This is frequently written as

$$z \gg \frac{d^2}{\lambda} \quad (1-35)$$

where the ratio of $\pi/4$ is taken to be 1. The criterion is often that the allowable phase shift is much less than $\frac{1}{2}$ radian, leading to

$$z \gg \frac{2d^2}{\lambda}. \quad (1-36)$$

In either case, the region in which (1-33) can be neglected is called the **Fraunhofer region**, or the far field. The Fraunhofer distribution is given by

$$U_2(x_2, y_2) = \frac{1}{i\lambda z} \exp(ikz) \exp\left(ik \frac{x_2^2 + y_2^2}{2z}\right) \times \int \int U_1(x_1, y_1) \exp\left(-ik \frac{x_1 x_2 + y_1 y_2}{z}\right) dx_1 dy_1 \quad (1-37)$$

or

$$U_2(x_2, y_2) = \frac{1}{i\lambda z} \exp(ikz) \exp\left(ik \frac{x_2^2 + y_2^2}{2z}\right) A_1\left(\frac{x_2}{\lambda z}, \frac{y_2}{\lambda z}\right). \quad (1-38)$$

If we recognize that, for small values of x/z and y/z , $\cos\theta = x_2/z$, $\cos\xi = y_2/z$, then (1-7) allows us to write

$$U_2(x_2, y_2) = \frac{1}{i\lambda z} \exp(ikz) \exp[i\pi\lambda z(u' + v')] |A_1(u, v)|. \quad (1-39)$$

Either expression shows that the diffraction pattern in the Fraunhofer region is given by the Fourier transform of the distribution in the first plane. The multipliers show the $1/z$ decrease in amplitude ($1/z^2$ in power), the progressive phase shift due to propagation, and a quadratic phase factor.

It is interesting to put some numbers into (1-35) to determine where (1-33) can be neglected. If λ is 5×10^{-7} meter (in the visible region) and the aperture diameter is 0.5 cm, $z \gg 50$ meters. We shall see in Chapter 5 that there are methods by which we can reduce the value of z and still obtain a Fraunhofer diffraction pattern. These techniques involve the introduction of another phase term into (1-32) to cancel or reduce the effect of the term (1-33).

第二讲 相因子判断法

钟 锡 华

(北京大学物理系)

一、相因子判断法

原则上说,依据菲涅耳-基尔霍夫衍射积分公式,可以由一个已知波前导出另一个波前,譬如由衍射屏的输出场 $\tilde{U}_2(x, y)$ 导出接收屏幕上的衍射场 $\tilde{U}(x', y')$ 。但实际上,这种积分运算通常是复杂的,总需在一定条件下作近似处理,即使如此,能定量给出结果的情况目前也还为数不多。不过,我们注意到波场的主要特征体现在波函数的位相因子之中。如果将复杂波场的复振幅相因子与平面波或球面波的相因子作一对比,那么复杂波场就可看成是一系列平面波成分(或球面波成分)的迭加,因而也就变成人们在概念上容易掌握的一种波场了。其重要的是,复杂波场中各成分(不同方向的平面波和不同聚散程度的球面波)还可用透镜和稜镜等元件进行实际上的分离,这就便于人们对波前作进一步的处理。所谓相因子判断法,简单说来,就是根据波函数的相因子来判断波场的性质和分析衍射场的主要特征。这种方法要比衍射积分运算简便得多。在很多场合下只需要掌握衍射场的主要特征就够用了,在全息术中尤其如此。为此,我们必须熟悉(1)平面波函数的相因子,以及几种典型情况下球面波函数的相因子;(2)透镜和稜镜的位相变换函数。

二、傍轴与远场条件下
球面波的波前相因子

表1 平面波和球面波在波前上的相因子

波的类型 ⁽¹⁾	特 征	相 因 子
(1) 平面波	传播方向 (θ_1, θ_2) ; 当 $\theta_1 = \theta_2 = 0$ 时	$\exp[ik(\sin\theta_1 x + \sin\theta_2 y)]$; (2.1)
(2) 发散球面波	中心在轴上, 坐标 $(0, 0, -z)$	$\exp\left[ik \frac{x^2 + y^2}{2z}\right]$ (2.2)
(3) 会聚球面波	中心在轴上, 坐标 $(0, 0, z)$	$\exp\left[-ik \frac{x^2 + y^2}{2z}\right]$ (2.3)
(4) 发散球面波	中心在轴外, 坐标 $(x_0, y_0, -z)$	$\exp\left[ik \left(\frac{x^2 + y^2}{2z} - \frac{xx_0 + yy_0}{z} \right)\right] \times \exp\left[ik \frac{x_0^2 + y_0^2}{2z}\right]$ (2.4)
(5) 会聚球面波	中心在轴外, 坐标 (x_0, y_0, z)	$\exp\left[-ik \left(\frac{x^2 + y^2}{2z} - \frac{xx_0 + yy_0}{z} \right)\right] \times \exp\left[-ik \frac{x_0^2 + y_0^2}{2z}\right]$ (2.5)

(1) 五类波型分别见图 2.1 (a) — (e)。

下四种球面波的波前相因子,并将平面波的波前相因子列为首位(1)。所谓傍轴条件指的是点源位置 (x_0, y_0) 和场点位置 (x, y) 皆远远小于纵向距离 z , 即

$$\rho(x_0, y_0) \ll z, \rho(x, y) \ll z.$$

现以轴外发散球面波为例谈一谈表中列出的波前函数是如何导出的。

如图 2.2 所示,通过 O 点的 $x_0 y_0$ 平面为物平面,通过 O' 点的 xy 平面为波前,二者间距为 z 。令物平面上任意光点 $Q(x_0, y_0)$ 到场点 $P(x, y)$ 的矢径为 r , 物平面中心 O 到 P 的矢径为 r_0 , Q 到波前中心 O' 的矢径为 r'_0 , 它们的长度分别为

$$r = \sqrt{(x_0 - x)^2 + (y_0 - y)^2 + z^2},$$

表1中(2), (3), (4)和(5)列出傍轴条件

2-1

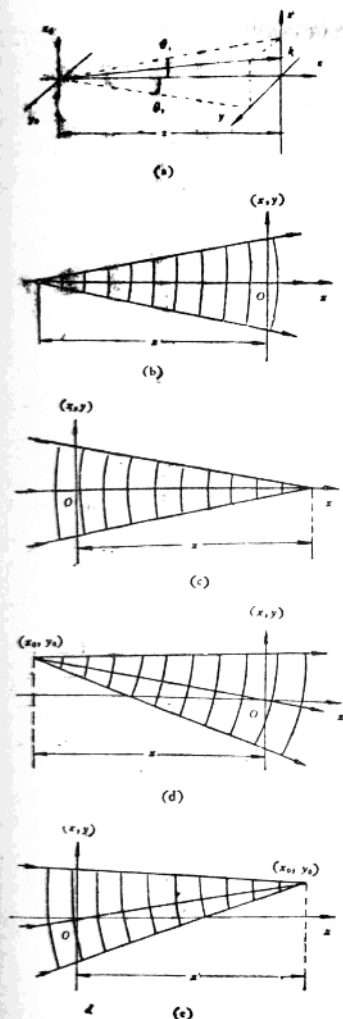


图 2.1

$$r_0 = \sqrt{x_0^2 + y_0^2 + z^2},$$

$$r = \sqrt{x^2 + y^2 + z^2}.$$

源点 Q 在 xy 面上激发的球面波前为

$$\tilde{U}(x, y) = \frac{a}{r} e^{ikr}$$

$$= a / (\sqrt{(x_0 - x)^2 + (y_0 - y)^2 + z^2})$$

$$\times \exp[ik\sqrt{(x_0 - x)^2 + (y_0 - y)^2 + z^2}]$$

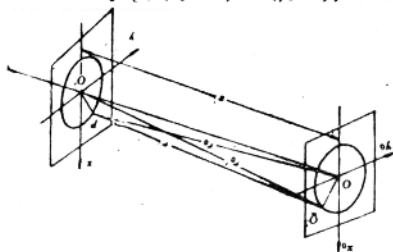


图 2.2 傍轴条件与远场条件(轴外物点)

将 r_0 、 r'_0 和 r 的表达式作泰勒展开:

$$r_0 = z + (x_0^2 + y_0^2)/2z + \dots$$

$$r'_0 = z + (x_0^2 + y_0^2)/2z + \dots$$

$$r = z + (x^2 + y^2)/2z + (x_0^2 + y_0^2)/2z$$

$$- (xx_0 + yy_0)/z + \dots$$

$$= r_0 + (x_0^2 + y_0^2)/2z - (xx_0 + yy_0)/z$$

$$+ \dots = r'_0 + (x^2 + y^2)/2z$$

$$- (xx_0 + yy_0)/z + \dots,$$

在物点与场点都满足傍轴的条件下, 可令振幅中 $r \approx z$, 于是

$$\tilde{U}(x, y) = (a/z) \exp[ik\{r_0 + (x_0^2 + y_0^2)/$$

$$2z\}] \exp[(-ik/z)(xx_0 + yy_0)]$$

$$= (a/z) \exp[ik\{r'_0 + (x^2 + y^2)/$$

$$2z\}] \exp[(-ik/z)(xx_0 + yy_0)],$$

此处只有振幅具有平面波的特点, 即它等于常数。在位相因子中既有 x 和 y 的平方项, 也有 x_0 和 y_0 的平方项, 还有 $xx_0 + yy_0$ 的线性交叉项。为了突出相因子, 将常数 $a e^{ikz}/z$ 简化为 1, 就得到表中所示的函数形式。为要忽略相因子中的二次项, 必须同时满足远场条件。若物点同时满足远场条件:

$$x_0^2, y_0^2 \ll z\lambda,$$

则

$$\tilde{U}(x, y) \propto \exp[ik\{(x^2 + y^2)/2z$$

$$- (xx_0 + yy_0)/z\}], \quad (2.6)$$

此时仍然保留了场点坐标的二次项。若场点同时满足远场条件:

$$x^2, y^2 \ll z\lambda,$$

$$\text{则 } \tilde{U}(x, y) \propto \exp\{ik[(x_0^2 + y_0^2)/2z - (xx_0 + yy_0)/z]\}. \quad (2.7)$$

此时仍然保留了源点坐标的二次项。不论这些被保留下来的二次项对今后讨论的问题有多大的影响,我们必须注意到在(2.4)~(2.7)的相因子中始终存在线性项 $ik(xx_0 + yy_0)/z$, 它的出现预示着平面波成分的存在。

三、透镜和稜镜的位相变换函数

在成像光学系统中,透镜起两方面的作用。一方面它是光瞳,限制着波面,仅提取入射光波中央一部分波面 Σ_1 进入光学系统;另一方面它变换波面,把一种波面变换为另一种。例如,在理想成像的情况下,凸透镜把发散的球面波变换为会聚的球面波 Σ_2 。较为实际的情况是,它把入射波面变换为偏离球面的象差波面 Σ_2' (参见图 2.3)。以往,我们分别用有限孔径引起的衍射和透镜本身的几何象差来描述上述两种作用。其实,从纯粹波动光学的观点出发,引入一个复振幅透过率函数 $\tilde{t}(x, y)$ 即可把透镜的上述两方面的性质全部反映出来。

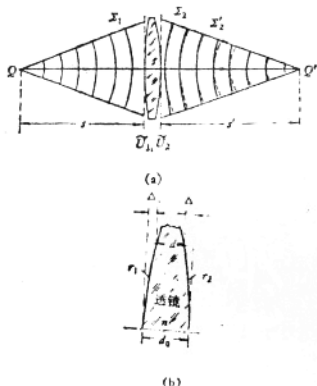


图 2.3 把透镜看作一种位相型透射屏

如图 2.3, 在透镜前后各取一个平面。设它们上边的人射波前和透射波前分别为

$$\begin{cases} \tilde{U}_1(x, y) = A_1(x, y)e^{i\varphi_1(x, y)}, \\ \tilde{U}_2(x, y) = A_2(x, y)e^{i\varphi_2(x, y)}. \end{cases}$$

透镜的透过率函数为

$$\tilde{t}(x, y) = (A_2/A_1)e^{i(\varphi_2 - \varphi_1)} = \begin{cases} a(x, y)e^{i\varphi(x, y)} & r < D/2; \\ 0 & r > D/2. \end{cases} \quad (2.8)$$

这里 $r = \sqrt{x^2 + y^2}$, D 是透镜直径。设透镜材料对入射光波是透明的, 忽略透镜对光的吸收、反射等造成的能量损耗, 则 $A_2 = A_1$, 而

$$a(x, y) = A_2(x, y)/A_1(x, y) = 1,$$

于是在透镜的孔径内,

$$\tilde{t}(x, y) = e^{i\varphi(x, y)},$$

其中

$$\varphi(x, y) = \varphi_2(x, y) - \varphi_1(x, y) \quad (2.9)$$

称为透镜的位相变换函数。

严格求透镜的位相变换函数是困难的, 下面在傍轴条件下计算薄透镜的位相变换函数。由于透镜很薄, 入射点与出射点的坐标相近, 光程可近似地沿平行于光轴方向计算。如图 2.3 (b)。

$$\begin{aligned} \varphi(x, y) &= (2\pi/\lambda)[\Delta_1 + \Delta_2 + nd(x, y)] \\ &= \varphi_0 - (2\pi/\lambda)(n-1)(\Delta_1 + \Delta_2), \end{aligned} \quad (2.10)$$

其中

$$\varphi_0 = (2\pi/\lambda)nd_0,$$

φ_0 是与 x, y 无关的常数, 它不影响波前上位相的相对分布, 常可略去不写。式 (2.10) 中重要的是第二项, 在傍轴条件下 Δ_1 和 Δ_2 可写为

$$\begin{aligned} \Delta_1(x, y) &= r_1 - \sqrt{r_1^2 - (x^2 + y^2)} \\ &\simeq (x^2 + y^2)/2r_1, \end{aligned}$$

$$\begin{aligned} \Delta_2(x, y) &= (-r_2) - \sqrt{(-r_2)^2 - (x^2 + y^2)} \\ &\simeq -(x^2 + y^2)/2r_2, \end{aligned}$$

式中 r_1, r_2 分别是透镜前后两表面的曲率半径 (它们可取正值或负值, 例如对于双凸透镜, $r_1 > 0, r_2 < 0$)。代入式 (2.10) 并略去 φ_0 , 得

$$\begin{aligned} \varphi(x, y) &= -(2\pi/\lambda) \cdot (n-1)/2(1/r_1 \\ &\quad - 1/r_2)(x^2 + y^2) \\ &= -k(x^2 + y^2)/2F, \end{aligned} \quad (2.11)$$

式中

$$F = 1/[(n-1)(1/r_1 - 1/r_2)], \quad (2.12)$$

而 $k = 2\pi/\lambda$ 。于是透镜的透过率函数为

$$\tilde{t}_l(x, y) = e^{-ik(x^2 + y^2)/2F} \quad (2.13)$$

稜镜元件的基本作用不是成像而是偏转，它将一个方向的平行光束变换为另一方向的平行光束，因为平面波具有线性相因子的特点，所以预料稜镜的位相变换函数的指数是线性的。仿照前面的推导，对于楔形薄稜镜(图 2.4)可以

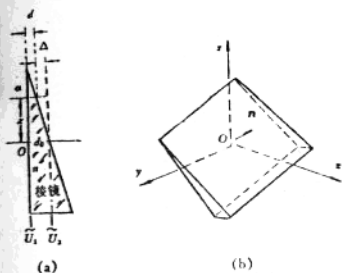


图 2.4 楔形薄稜镜的位相变换函数和透过率函数

近似地认为光线在第二个界面是等高变向出射，位相差为

$$\begin{aligned} \varphi(x, y) &= \varphi_2(x, y) - \varphi_1(x, y) \\ &= (2\pi/\lambda)(\Delta + nd) \\ &= \varphi_0 - (2\pi/\lambda)(n-1)\Delta \\ &= \varphi_0 - (2\pi/\lambda)(n-1)\alpha x, \end{aligned} \quad (2.14)$$

式中 $\varphi_0 = 2\pi/\lambda nd$ ， d_0 为稜镜中心厚度， α 为稜镜的顶角。上述计算是针对稜镜交接平行 y 轴的情形进行的，如果考虑稜镜方位是任意状态，则可取其第二界面法线 n 的前两个方向角的余角 α_1 和 α_2 来标定稜镜的取向，此时位相变换函数成为

$$\begin{aligned} \varphi(x, y) &= \varphi_0 - k(n-1)(\alpha_1 x + \alpha_2 y) \\ (k &= 2\pi/\lambda). \end{aligned} \quad (2.15)$$

略去常数相因子 φ_0 ，并忽略光在稜镜表面和内部的损耗，得楔形薄稜镜的透过率函数：

$$\tilde{t}_l(x, y) = \exp[-ik(n-1)(\alpha_1 x + \alpha_2 y)]. \quad (2.16)$$

总之，成像元件的薄透镜具有二次型的位相变换函数 $\tilde{t}_l$ ，偏转元件的薄稜镜具有线性型的位相变换函数 $\tilde{t}_l$ ，这一点要引起我们足够的

重视。有实际意义的问题往往是倒过来提的，即如果在某种波前变换的场合中出现了如同 $\tilde{t}_l$ 或 $\tilde{t}_l$ 的作用函数，作用在一种波前函数 $\tilde{U}$ 身上，那么，实际效果就象是 $\tilde{U}$ 波经历一个透镜的聚、散或稜镜的偏转，不管在那种场合下是否真有透镜或稜镜元件存在。下面举几个例子，说明如何根据薄透镜和稜镜的位相变换函数，运用相因子判断法求后场中波的类型和特征。

例 1 平行于光轴的单色平行光束入射在透镜上，试判断后场中波的类型和特征。

解 对于正入射的平面波， $\tilde{U}_l(x, y) = A_l$ (常数)，故

$$\begin{aligned} \tilde{U}_2(x, y) &= \tilde{t}_l(x, y)\tilde{U}_l(x, y) \\ &= A_l \exp[-ik(x^2 + y^2)/2F], \end{aligned} \quad (2.17)$$

从相因子看，这是会聚到透镜后距离为 F 处的球面波。以上正是几何光学理论所预期的。决定 F 值的公式 (2.12) 正是几何光学中用光线追迹方法得到的磨镜者(焦距)公式。

例 2 试导出透镜的物象距公式 (见图 2.5)。

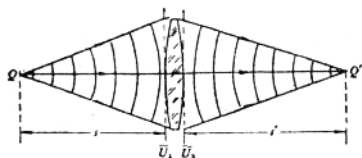


图 2.5 利用透镜屏函数推导物象公式

解 设想在透镜前距离为 s 处置一发光物点 Q ，它发出的球面波在透镜上造成的波前为

$$\tilde{U}_l(x, y) = A_l \exp[ik(x^2 + y^2)/2s].$$

故从透镜输出的波前为

$$\begin{aligned} \tilde{U}_2(x, y) &= \tilde{U}_l(x, y)\tilde{t}_l(x, y) \\ &= A_l \exp[ik(x^2 + y^2)/2s] \exp[-ik(x^2 + y^2)/2F] \\ &= A_l \exp\{-ik[(x^2 + y^2)/2](1/s - 1/F)\}, \end{aligned}$$

从相因子看，这是会聚的球面波，会聚中心(象点) Q' 在透镜后，距离为

$$s' = 1/(1/F - 1/s),$$

或

$$1/s + 1/s' = 1/F.$$

这正是几何光学给出的透镜物象距公式。

例 3 轴上一物点 Q 与稜镜的距离为 s , 求 Q 点发射的傍轴球面波经稜镜后出射波的特征 (见图 2.6)。

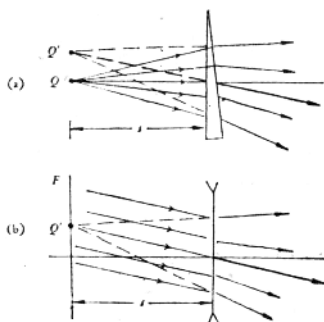


图 2.6

解 出射波前函数为

$$\begin{aligned}\tilde{U}_2(x, y) &= \tilde{U}_1(x, y) \cdot \tilde{t}_p(x, y) \\ &= A_1 \exp[ik(x^2 + y^2)/2s] \\ &\quad \cdot \exp[-ik(n-1)(\alpha_1 x + \alpha_2 y)] \\ &= A_1 \exp\{ik[(x^2 + y^2)/2s \\ &\quad - (n-1)(\alpha_1 x + \alpha_2 y)s \cdot 1/s]\}.\end{aligned}$$

对照表 1, 它正是轴外点源发散球面波 (2.4) 形式, 点源位置可由线性相因子系数定出。

$$\begin{cases} x_0 = -(n-1)\alpha_1 s, \\ y_0 = -(n-1)\alpha_2 s, \\ z_0 = s. \end{cases}$$

有趣的是, 上述球面波经稜镜的偏转, 也可以看成斜入射的平面波经过一个假想透镜的聚焦。这只要将 $\tilde{U}_2$ 改写为以下形式就明白了:

$$\begin{aligned}\tilde{U}_2(x, y) &= \tilde{t}_1 \{ A_1 \exp[-ik(n-1)(\alpha_1 x + \alpha_2 y)] \} \\ \tilde{t}_1 &= \exp[ik(x^2 + y^2)/2s] \\ &\quad - \exp[-ik(x^2 + y^2)/2F],\end{aligned}$$

亦即把 $\tilde{U}_2$ 函数中的二次型相因子看为一个等效透镜的屏函数 $\tilde{t}_1$, 而将后面的线性因子连同振幅系数 A_1 看作一系列平面波。等效薄透镜的焦距 $F = -s$, 是一个发散透镜。几何光学知识告诉人们, 平行光经发散透镜以后变为发散的同心光束, 光束中心位于后焦面上一点 Q' 。后面括号中平面波的传播方向为

$$\begin{aligned}\sin \theta_1 &= -(n-1)\alpha_1 \text{ 和} \\ \sin \theta_2 &= -(n-1)\alpha_2,\end{aligned}$$

所以它在后焦面上的交点位置坐标就与前面的结果一致。

相因子判断法的优越性, 将在下面几讲中得以充分的体现。

Propagation of plane waves. A uniform plane wave traveling in an arbitrary direction can be represented by

$$|U| \operatorname{Re} \exp [-i(\omega t - \alpha x - \beta y - \gamma z + \epsilon)] \quad (1-1)$$

where $|U|$ represents the magnitude of the wave, ϵ denotes the phase, and Re denotes real part. We will normally allow a complex U to include both amplitude and phase. The power density of the wave is given by UU^* . The variable U therefore represents neither the electric nor the magnetic field but could be associated with either, depending on the definition of U . That is, $U = E/\sqrt{\eta}$ or $U = \sqrt{\eta} H$ where η is the intrinsic impedance of the medium, E represents the electric field, and H represents the magnetic field. In our scalar work, a choice need not be made. The intensity of a wave is a measure of the energy per unit time per unit area *normal* to the direction of propagation. The variables α , β , and γ determine the direction in which the wave propagates, and are restricted by

$$\alpha^2 + \beta^2 + \gamma^2 = k^2 = \left(\frac{2\pi}{\lambda} \right)^2 \quad (1-2)$$

where λ is the wavelength. The values of α , β , and γ are related to the direction of propagation; that is

$$\alpha = k \cos \theta, \quad (1-3)$$

$$\beta = k \cos \xi, \quad (1-4)$$

$$\gamma = k \cos \chi \quad (1-5)$$

where θ , ξ , and χ , are respectively, the angles between the x , y , and z axes and the direction of propagation (normal to the wave). Note that the dimensions of α , β , and γ must be radians per unit length.

In writing a description of a propagating wave, we shall normally omit the temporal variable and the propagation constant for the z direction. The temporal variation is understood and the value of γ can be found from (1-2). The expression (1-1) can then be written as

$$U \exp \{ i2\pi \{ ux + vy \} \} \quad (1-6)$$

INTRODUCTION TO SCALAR DIFFRACTION

where u and v are spatial frequencies having the dimensions of cycles per unit length:

$$u = \frac{\alpha}{2\pi} = \frac{\cos \theta}{\lambda}, \quad v = \frac{\beta}{2\pi} = \frac{\cos \xi}{\lambda}. \quad (1-7)$$

Figure 1-1 illustrates the parameters discussed for a plane wave propagating in the x and z directions. The solid lines are maxima and the dotted lines are minima of the tilted wave. Along the x axis the distance between maxima is given by $\lambda/\cos \theta$. That is, the spatial period in the x direction is $\lambda/\cos \theta$, and the spatial frequency is given by (1-7). A wave propagating partially in the negative x direction, such that the spatial period is again $\lambda/\cos \theta$, has the same magnitude spatial frequency, but the phase regresses $(\cos \theta)\lambda$ cycles per unit length rather than advancing by the same amount. This could be considered as a *negative* progression of spatial phase. Notice, however, that the real part of (1-6) is the same in either case. When two waves represented by $U \exp(i2\pi ux)$ and $U \exp(-i2\pi ux)$ are combined, the resulting amplitude distribution along the x axis is $2U \cos(2\pi ux)$, giving a fringe distribution having a spatial frequency of u cycles per unit length. Consequently, we see an intensity pattern described by $4U^2 \cos^2(2\pi ux)$.

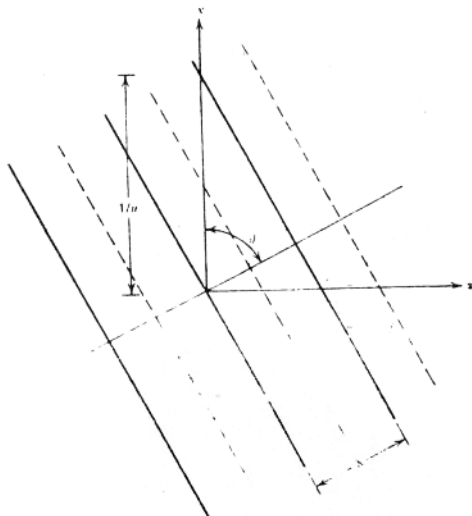


Figure 1-1 A plane wave propagating in the positive x and z directions.