

第 1 章

光学非球面测量概述

非球面是对偏离球面的曲面的总称。在光学系统中使用非球面光学元件,不仅能增加光学设计的自由度,有利于像差校正、改善像质、提高光学系统性能,而且能够减少光学元件的数量和质量,简化仪器结构,大大减少系统的尺寸和质量,降低成本。因此,非球面的研究与应用长期以来一直备受人们的关注。但由于加工与检测困难,非球面的应用最初受到很大的限制。随着科学技术的迅猛发展,满足人们特殊需要的非球面镜片的设计方法和加工手段有了长足进步,检测技术成为制约非球面加工和应用的关键技术之一。如何客观、准确地评测非球面是高精度非球面加工的基础,研究精度高、稳定实用的非球面检测技术日益成为非球面推广应用亟待解决的问题。

本章首先介绍光学非球面的应用,之后介绍相关的数学基础,在此基础上定义光学非球面检测及评价的指标,并概述现有的主流检测方法。

1.1 光学非球面的应用

基于非球面光学元件的优点,光学系统中已经广泛使用非球面替代球面光学元件,小到普通的眼镜镜片,大到照相透镜、平版印刷系统、天文光学系统、军用光电系统。

在军事领域中,所使用的雷达测距仪、光学望远镜、红外夜视仪等装置均不同程度地使用了精密的非球面光学元件^[1-3]。例如,AN/AVS-6 飞行员使用的微光夜视眼镜,采用 5 个非球面和 1 个球面光学元件构成物镜系统,采用 4 个非球面和 1 个球面光学元件构成目镜系统。再如,我国研究的某光电侦察跟踪系统,对于其中一激光接收镜,若全部采用球面光学元件设计该系统需要 21 片光学镜片,而采

用非球面系统则只需 5 片光学镜片。从设计周期上看,由于球面系统的复杂性,其设计需花费 2 个月的时间,而非球面系统的设计只需要 10 天。从系统的体积和质量上看,若采用球面系统,激光接收镜的质量高达 2.5 kg,而采用非球面系统时,激光接收镜的质量只有 1 kg^[4]。由此,非球面光学元件的优势可见一斑。

在民用领域,诸如天文望远镜^[5]、卫星红外望远镜^[6-7]、光纤耦合装置^[8]、太阳能接收镜^[9]、激光打印机、高品质照相机、摄像机、激光光斑整形装置^[10]、高能激光扩束装置、大屏幕投影电视镜头^[11-12]等现代仪器均不同程度地运用了非球面光学元件及其技术。例如,代替哈勃望远镜的詹姆斯·韦伯太空望远镜采用反射式成像光路,主镜是直径为 6.5 m 的抛物面,次镜是双曲面,三镜是椭球面。再如,在医用领域中,非球面光学元件也得到了广泛的应用^[13]。研究表明,白内障患者植入球面人工晶状体后,部分患者存在视力模糊、夜间视力差、眩光、光晕、重影等问题。研究者根据波前像差原理,设计出在光学上与人眼自然晶状体更为接近的非球面人工晶状体,用以提高患者术后视觉功能。非球面人工晶状体与球面人工晶状体相比,高阶像差尤其是球差大大降低,更接近于年轻人的自然晶状体;对比敏感度好,尤其是在大瞳孔、高空间频率和眩光状态下差异更明显,从而可改善夜间眩光状态下的视力^[14]。可见,非球面光学元件已被应用于人类生活的方方面面。

1.2 光学非球面的分类及数学描述

1.2.1 非球面分类

根据表面类型,非球面通常分为旋转对称型和非旋转对称型,如图 1.1 所示。旋转对称型非球面包括二次旋转对称型非球面和高次旋转对称型非球面,前者的母线方程可以表示为椭圆线、抛物线、双曲线等,以曲线顶点为坐标原点,将其依次绕对称轴线旋转形成椭球面、抛物面、双曲面等二次曲面;后者主要是指母线沿着经过其顶点的轴线也具有对称分布特点,但是母线方程只能表示成高阶多项式的解析形式,以母线顶点为坐标原点,将其绕对称轴线旋转形成高次旋转型非球面。

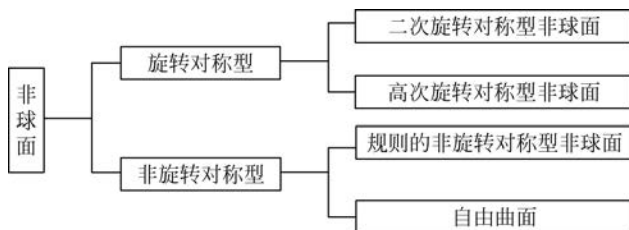


图 1.1 非球面光学零件面形分类

非旋转对称型非球面中的一类是规则的非旋转对称型非球面,其面形有一定的规律,包括连续非球面,如两轴对称式的非球面(非球面片断,也称为离轴非球面)、复曲面、柱面和多轴对称式的非球面;以及非连续非球面,如微透镜阵列、光栅的表面、菲涅尔透镜、二元光学元件等。另一类是面形无规则的非旋转对称型非球面,这一类非球面的面形基本上没有任何限制,只能用离散数据点的形式表示,可以称为复杂的自由曲面。这类自由曲面没有确定的方程描述,是基于点集或曲线生成、由空间离散数据点通过拟合而成,无旋转轴或对称中心的复杂表面。自由曲面在光学系统中的应用越来越多。参数向量可以表示任何形状的自由曲面,包括多项式、贝塞尔曲面、B样条曲面、非均匀有理B样条曲线(NURBS)曲面等。

1.2.2 非球面方程

对于轴对称非球面,用它的子午截线方程表示曲面方程。在实际应用中,经常用三种形式的方程式来表达。设光轴为 z 轴,即非球面的对称轴,坐标原点取在顶点。

第一种非球面子午截线方程式为

$$z^2 = a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + \cdots \quad (1.1)$$

式中, $a_1, a_2, a_3, a_4, \cdots$ 为方程系数。如果非球面是二次曲面,则式(1.1)为

$$z^2 = 2R_0 x + (e^2 - 1)x^2 \quad (1.2)$$

这是二次曲面的通用方程式。式中, R_0 为曲面近轴曲率半径, e 为曲面的偏心率。

第二种非球面子午截线方程式为

$$z^2 = A_1 x^2 + A_2 x^4 + A_3 x^6 + A_4 x^8 + \cdots \quad (1.3)$$

式中, $A_1, A_2, A_3, A_4, \cdots$ 为方程系数。第一项系数和非球面近轴曲率半径有关,即 $A_1 = 0.5R_0$ 。

第三种非球面子午截线方程式为

$$z = \frac{cx^2}{1 + \sqrt{1 - (K + 1)c^2 x^2}} + B_1 x^4 + B_2 x^6 + B_3 x^8 + \cdots \quad (1.4)$$

式中: c 为近轴曲率, $c = 1/R_0$; K 为二次曲面常数, $K = -e^2$; $B_1, B_2, B_3, \cdots$ 为高次系数。

在光学设计和工程运用中,式(1.1)、式(1.3)和式(1.4)往往互相交叉使用,它们之间存在一定的系数关系。

在实际使用中,二次曲面应用最为广泛。通常取式(1.4)右边的第一项表示二次曲面。通过转换可得到二次曲面求 z 的另一个有用的表达式为

$$z = \frac{R_0 - \sqrt{R_0^2 - (1 + K)x^2}}{1 + K} \quad (1.5)$$

当 K 取不同的值时,代表着不同的曲面,如图 1.2 所示。

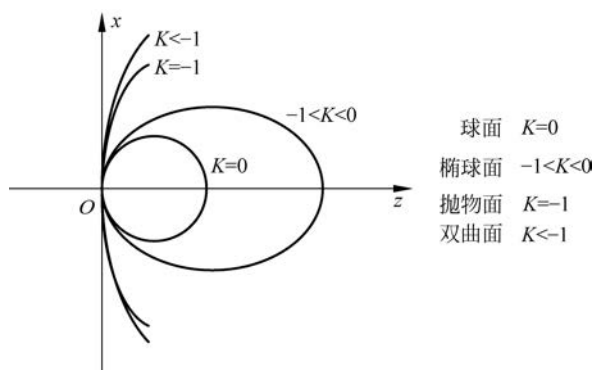


图 1.2 二次曲面的分类及其二次曲面常数

1.3 光学非球面检测及评价指标

本节首先定义非球面检测的常用指标,并介绍与非球面检测相关的光学概念,主要包括几何像差和泽尼克多项式。

1.3.1 面形误差和参数误差

由于加工误差、机械载荷或热载荷的作用,光学元件表面与理想表面之间将产生偏差,或光学元件参数发生变化,影响非球面基本性质,进而影响光学系统的成像质量。因此,在光学元件投入使用前,通常需要分析面形偏差,用以验证加工质量或系统结构设计是否满足光学成像要求。面形误差和参数误差是典型的描述面形偏差的指标。

1. 面形误差

面形误差是指实际非球面表面形状与理想非球面表面形状之间的逐点差别,通常由峰谷值(PV)和均方根值(RMS)表征。

峰谷值 光学加工表面相对理想参考表面各点偏差的峰值 $E_{\max}$ 和谷值 $E_{\min}$ 之差

$$PV = E_{\max} - E_{\min} \quad (1.6)$$

峰谷值一般描述空间周期大于 2 mm 的误差,即对 2 mm 以下的信号通过低通滤波器滤掉。如果采用低通滤波器与不采用低通滤波器时的 PV 有较大的变化,说明光学加工表面存在着明显的中高频残差。

均方根值 光学加工表面相对于理想参考表面各点偏差 E_i 的均方根值,用来

描述被测波前偏离期望波前的程度

$$\text{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N E_i^2} \quad (1.7)$$

波前均方根值能系统全面地反映光学加工表面和理想面形之间的综合偏差情况,就是说可以描述光学加工表面的中高频残差。在高分辨率条件下,一般要求 $\text{PV} \leq \lambda/4, \text{RMS} \leq \lambda/15$ 。

PV 与 RMS 之间的关系 对于高精度光学加工表面, PV 一般是 RMS 的 4~7 倍。随着 PV 的减小, PV 和 RMS 之比也会变小。

2. 参数误差

参数误差是指非球面表达式中的参数存在误差。如式(1.5)中 R_0 是标称的曲率半径,它影响非球面的基本性质,如非球面的焦距; K 是标称的非球面系数,是非球面分类的基础。式(1.4)中 B_i 是高阶非球面系数。这些统称为表面参数,其偏差称为参数误差。

因此,面形误差和参数误差对非球面的光学特性有明显的影响。尽管这些参数之间存在着一定的耦合关系,但也可以进行单独测量。

1.3.2 几何像差与泽尼克多项式

非球面在光学系统中得到广泛应用的主要原因是其优异的消像差能力,此处的像差可以从几何光学中的几何像差考虑,也可以从物理光学中的波像差考虑。非球面的测量大多也利用其对像差的调制作用。因此,像差的概念对理解非球面的功能以及测试原理至关重要。本节简要介绍几何像差的概念和分类,以及波像差的数学表达形式之一泽尼克多项式。

1. 几何像差

像差理论是光学设计的基础,光学设计在很大程度上就是像差平衡设计。设计一个光学系统要通过选型、初始结构的计算和选择、像差的校正平衡与像质评价几个过程;而利用光学设计辅助软件进行设计,也是一个反复优化变量使像差得到最佳校正与平衡的过程,因此设计者应该对像差理论有深入的理解。

像差的实质是由光学系统中透镜材料的特性或折射、反射表面的几何形状而引起的实际像与理想像的偏差。理想像就是由理想光学系统所成的像,然而对于实际的光学系统,只有在近轴区域很小孔径角的光束范围所成的像才是完善的,在实际的应用中,由于成像空间、光束孔径以及不同颜色的光波折射率的不同,使实际光学系统所成的像存在一系列缺陷,即像差。像差大小反映了光学系统的成像质量,主要有六种几何像差,其中对单色光而言有球差、彗差、像散、场曲和畸变五

种像差,复色光除了具有以上五种像差外还存在色差。关于各类像差的具体定义和性质可以参阅应用光学或光学设计相关书籍,此处不再赘述。

2. 泽尼克多项式

通常人们使用幂级数展开式的形式描述光学系统的波像差。其中,泽尼克多项式和光学检测中观测到的像差多项式的形式是一致的,因而它常常被用来描述波前特性。

泽尼克多项式是由无穷数量的多项式完全集组成的,有两个变量 ρ 和 θ ,其在单位圆内部是连续正交的。需要注意的是,泽尼克多项式仅在单位圆的内部连续区域是正交的,通常在单位圆内部的离散的坐标上是不具备正交性质的。泽尼克多项式具有三个和其他正交多项式集不一样的性质。

(1) 泽尼克多项式 $Z(\rho, \theta)$ 可以化解为径向坐标 ρ 和角度坐标 θ 的函数,其形式如下:

$$Z(\rho, \theta) = R(\rho)G(\theta) \quad (1.8)$$

式中,关于角度的函数 $G(\theta)$ 是一个以 2π 弧度为周期的连续函数,并且满足当坐标系旋转 α 角度之后,其形式不发生改变,也就是旋转不变性,

$$G(\theta + \alpha) = G(\theta)G(\alpha) \quad (1.9)$$

其三角函数集形式如下:

$$G(\theta) = e^{\pm im\theta} \quad (1.10)$$

式中, m 是任意正整数或零。

(2) 泽尼克多项式的第二个性质是径向函数(radial function) $R(\rho)$ 必须是 ρ 的 n 次多项式,并且不包含幂次低于 m 次的 ρ 方项。

(3) 第三个性质是当 m 为偶数时, $R(\rho)$ 也为偶函数;当 m 为奇数时, $R(\rho)$ 也为奇函数。

径向多项式 $R(\rho)$ 可以看作雅可比多项式(Jacobi polynomials)的特例,记作 $R_n^m(\rho)$ 。它们的正交和归一化性质可由式(1.11)表示:

$$\int_0^1 R_n^m(\rho) R_{n'}^m(\rho) \rho d\rho = \frac{1}{2(n+1)} \delta_{nn'} \quad (1.11)$$

式中的 $\delta_{nn'}$ 是克罗内克符号(Kronecker delta),即当 $n = n'$ 时, $\delta_{nn'} = 1$; 当 $n \neq n'$ 时, $\delta_{nn'} = 0$ 。并且它具有归一化的性质

$$R_n^m(1) = 1 \quad (1.12)$$

在计算径向多项式时,为了方便起见,我们通常会将其分解成如下形式:

$$R_{2n-m}^m(\rho) = Q_n^m(\rho) \rho^m \quad (1.13)$$

式中, $Q_n^m(\rho)$ 的次数为 $2(n-m)$,由下式给出:

$$Q_n^m(\rho) = \sum_{s=0}^{n-m} (-1)^s \frac{(2n-m-s)!}{s!(n-s)!(n-m-s)!} \rho^{2(n-m-s)} \quad (1.14)$$

通常我们会用实数形式的多项式(正弦和余弦函数)来代替复指数多项式,这样的话,波前像差函数 $W(\rho, \theta)$ 的泽尼克展开式就有如下形式:

$$W(\rho, \theta) = \bar{W} + \sum_{n=1}^{\infty} \left[A_n Q_n^0(\rho) + \sum_{m=1}^n Q_n^m(\rho) \rho^m (B_{nm} \cos m\theta + C_{nm} \sin m\theta) \right] \quad (1.15)$$

式中, $\bar{W}$ 是平均波前像差, A_n 、 B_{nm} 、 C_{nm} 是多项式展开系数。由于“0级”项是个常数(或者叫作平移项)1,并且所有其他的泽尼克项在单位圆区域上的平均值均为零,波前像差函数 W 的平均值就是这个“0级”项的系数 A_0 ,这样式(1.15)就等价于

$$W(\rho, \theta) = A_0 + \sum_{n=1}^{\infty} \left[A_n Q_n^0(\rho) + \sum_{m=1}^n Q_n^m(\rho) \rho^m (B_{nm} \cos m\theta + C_{nm} \sin m\theta) \right] \quad (1.16)$$

对于一个旋转对称的光学系统来说,物体位于子午面内,因而波前像差相对于 yz 面对称的,也就是只有 θ 的偶函数(余弦项)项是非零项。对于一般情况,波前是不对称的,因而也就是同时包含两种三角函数形式。

需要注意的是,泽尼克多项式不一定是用来拟合检测数据的最佳多项式形式。在某些情况下,用泽尼克多项式来描述波前数据具有很大的局限性。比如说,当需要考虑空气扰动的时候,泽尼克多项式几乎没有什么价值,因为扰动本身是随机的,并不具备特定的波像差模式;同样也很难找到一组合适的泽尼克多项式来描述单点金刚石车削中的制造误差。

1.4 光学非球面检测方法

按照是否接触被测面,非球面面形的主要检测方法可以分为两类:接触式测量法和非接触式测量法。

1.4.1 接触式测量法

对于接触式测量,非球面面形的测量方法与球面曲率半径的测量方法类似,可以利用轮廓仪^[15-17]和三坐标测量机^[18-21]进行测量,如图1.3所示。首先,对被测非球面进行坐标逐点测量,然后对测量结果进行曲面拟合,获得顶点曲率半径和二次曲面常数的测量值。计算标称值和测量值的差值,即非球面参数误差。将拟合曲面和实测曲面点对点相减得到随机偏差分布,即非球面面形误差。

接触式测量法一般适用于测量精度在数微米量级的非球面元件和非球面加工



图 1.3 接触式测量仪器

(a) Panasonic 公司的 UA3P 系列三维立体测量仪；(b) Bruker 公司的鹅颈式坐标测量机

过程检验。它只能测量被检表面有限点的矢高,一般来说,检测精度较低。但接触式测量法一般不要求被测表面为光学表面,因而可作为非球面成型、研磨、粗抛光阶段的检验手段。典型的接触式轮廓仪产品有 Taylor Hobson 公司的 Talysurf PGI 系列接触式非球面测量系统, Talysurf1250A 的横向测量范围为 200 mm,纵向测量范围为 12.5 mm,分辨率为 0.8 nm,可完成非球面参数和面形误差测量,并声称对模压非球面测量没有斜率限制,但不能测量材质较软的非球面镜。

接触式测量法容易受到扫描非正交性的影响,引入较大的系统误差。测量仪器接触被测量面,一般情况下要求对被测量面进行抛光处理;在测量过程中,测量仪器和被测非球面均会因挤压而产生磨损和变形,影响测量结果;对环境要求高,静电、灰尘和震动等都会使测量精度降低^[22-23]。此外,由于接触式测量法采用逐点测量的形式,如果想提高空间分辨率,则会导致测量过程中数据的采集量增加、时间成本提高、效率降低,而在测量大口径非球面时,也将产生测量成本增加、仪器精度降低等多方面的问题,大大限制了接触式测量法的应用范围^[24]。

针对接触式测量法的不足,研究者们提出了非接触式的检测方法。非接触式测量法的测量结果精度较高,并且不会造成被测量面的损伤,受到越来越多的关注。非球面的几何关系和测量光路是非常复杂的,在不使用任何辅助元件的情况下,无法直接测量非球面的面形误差。因此,非接触式非球面测量的难度更大。进一步对非接触式测量法进行细分,包含几何检测法与干涉检测法两大类。

1.4.2 几何检测法

几何检测法测量非球面主要包括刀口检测法、朗奇检测法、哈特曼波前检测法,以及激光束偏转法和逆向哈特曼波前检测法。其中针对最后两种方法作者进行了较为深入的研究,将在第 2 章详细阐述。

1. 刀口检测法

刀口检测法由傅科(Foucault)于 1856 年提出,使得非球面镜面的制造精度有

了本质上的提高。刀口检测法检验透镜的原理如图 1.4 所示^[25],其原理是基于被检表面的变形使光线偏离原有的轨迹,通过遮拦这些偏离的光线来测定光线的横向位移,所以刀口检测法可以看作一种横向像差的检验方法。在零位检测中引入刀口检测法,具有设备简单、操作方便等特点,而且刀口检测法对外界环境要求不高,适合加工现场检测,其检测灵敏度可以达到 $\lambda/20 \sim \lambda/10$ 。因此,在干涉仪出现之前,刀口检测法是非球面检测的重要手段。刀口检测法最大的不足是不能定量检测,通过刀口检测法只能知道哪些区域高,哪些区域低,但没有高低的量值,被检面是否合格很大程度上依赖于工人的经验。

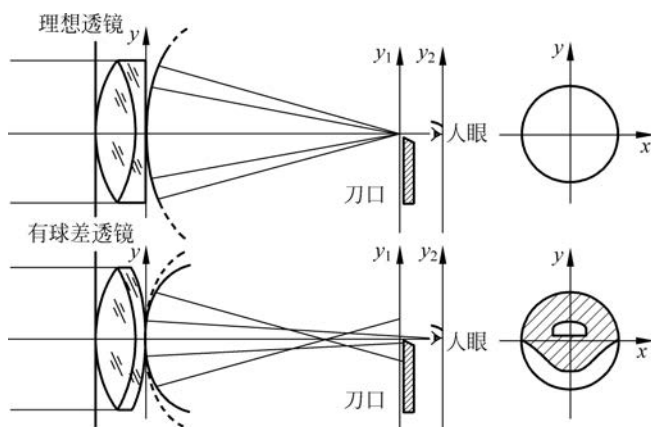


图 1.4 刀口检测法检测透镜原理图

2. 朗奇检测法

朗奇检测法是由意大利物理学家 Vasco Ronchi 于 1922 年发明的,如图 1.5 所示,其原理是将低频光栅置于被检非球面镜曲率中心时,光栅的像又落回光栅,产生莫尔条纹,这些条纹的形状携带着被检反射镜的面形误差的信息。朗奇检测法实际上是直接测量横向像差。由于朗奇检测法是共光路设置,所以条纹信号非常稳定且系统简单、检测效率高。最早提出将朗奇检测法应用于非球面检测的是 Waland 和 Schulz^[25],由于非球面各点曲率半径不一致,得到的朗奇条纹是弯曲的,根据条纹对理想形状的偏离,利用计算机处理可以计算出被检非球面的面形误差。但要注意的是,由衍射效应引起条纹的扩散会导致朗奇检测法的灵敏度降低,因此可将朗奇光栅的刻线制作成带有一定曲率,利用刻线的曲率补偿被检非球面的非球面度,可有效提高朗奇检测非球面的灵敏度。朗奇检测法是评价和测量光学表面质量最简单、有效的手段之一。但是与刀口检测法一样,朗奇检测法对面形的小误差很敏感,故检测的动态范围小,而且它们的定量检测技术都不成熟,在半定量的应用中最有优势。

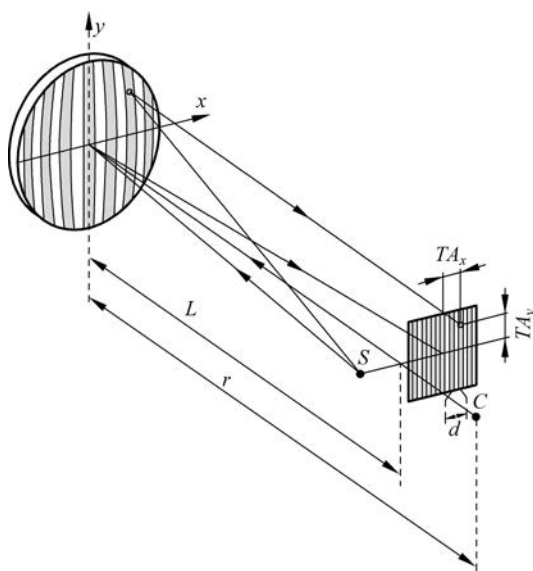


图 1.5 朗奇检测法

3. 哈特曼波前检测法

哈特曼波前检测法由哈特曼首次提出^[26]，后来在光学检测中许多人也采用了类似的原理和方法，可以用来测量光学元件的几何像差，也可以测量面形。与干涉检测法相比，哈特曼波前检测法具有系统简单、稳定，成本较低，无需针对待测面型参数进行逐一设计系统等优点。

传统哈特曼波前检测法测量面形时，通过瞳面附近的哈特曼光阑对光线采样，在焦平面附近用探测器对光线形成的光斑进行探测，由光线在哈特曼光阑和探测面上的位置得到所探测光线的斜率矩阵，复原出被测镜的面形，光路原理图如图 1.6 所示。该方法的优点是对测量光没有相干性要求，无需参考光，结构相对简单。

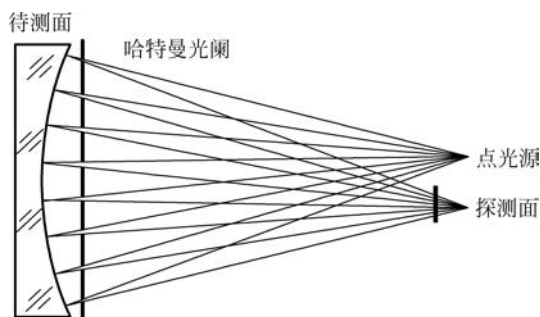


图 1.6 传统哈特曼波前检测法测量原理图