1 导 论

1.1 计算——第三种科学方法

科学计算的兴起是20世纪最重要的科学进步之一.近年来,在各种科学与工程领域中都逐步形成了计算性学科分支,如计算力学、计算物理、计算化学、计算生物学、计算地震学,等等.计算在生命科学、天文学、医学、系统科学、经济学、社会科学以及其他软科学中所起的作用也日益增大.在气象、核技术、石油勘探、航空航天、金融、交通运输、密码破译等国民经济与国防建设的许多重要领域中,计算已经成为必不可少的手段.

著名计算物理学家、诺贝尔奖获得者Wilson教授在80年代就指出:"当今,科学活动可分为三种:理论、实验和计算.定义计算科学最好是通过比较它的核心活动和实验及理论的核心活动.实验科学家从事于测量和设计科学设备及利用这些设备去进行

测量，致力于可控、可重复实验的设计以及分析这些实验的误差；理论科学家研究实验数据之间的关系、这些关系满足的原理（如牛顿定律、对称性原理等）及把这些原理运用到具体特殊情形所需的数学概念和技术．计算科学家构造求解科学问题的计算方法，把这些方法软件化，设计和进行试验，分析这些数值试验的误差．他们研究计算方法的数学特征，通过计算揭露所求解科学问题的基本性质和规律．”

从第一台电子计算机 ENIAC 诞生到今天的半个世纪里，计算速度已经提高了亿倍以上．从 60 年代到现在，计算机的发展更加迅速，计算方法的研究也受到了前所未有的重视，计算方法的效率不到 10 年就提高 10 倍．计算机和计算方法的进步极大地提高了人类的计算能力，从而引起了科学方法论的巨大变革：如果说伽利略和牛顿在科学发展史上奠定了实验和理论这两大科学方法支柱，那么由冯·诺伊曼研制的现代电子计算机则把计算推上了人类科学活动的前沿，使计算成为第三种方法，它与实验、理论共同成为科学方法论的基本环节，它们互相补充，互相依赖，而又相对独立，不可缺少．著名数学家冯康先生生前在多种场合反复强调与论述科学计算的重要性以及计算已成为第三种方法．他的观点影响日趋扩大，得到了越来越多的科学家的赞同．

在全世界，特别是在发达国家，计算机已无处不见，计算的影响已无处不有．计算在许多行业的应用中取得了巨大的经济效益．例如，飞机设计传统的办法是设计、风洞实验、修改设计、再做风洞实验，这样反复进行．风洞产生巨大的风速，需要消耗非常大的能量．而且这样做，使得设计飞机的周期非常长，往往需要几年甚至十多年，耗资是巨大的．现在，有些风洞实验可以用数值模拟来代

替，这样既节约开发经费也大大缩短了研究周期.不经过风洞实验而完全依赖计算机模拟设计出来的飞机已能上天.波音公司的"波音777型"飞机依赖计算机设计的成功在世界上引起了广泛的重视.在其他许多应用方面，科学计算的作用也不可替代.例如，没有计算机就不可能做出准确的天气预报；没有计算机就不可能研制核武器.

在基础研究领域，计算的重要性不仅在于它对实验的简单替代，还在于通过计算对研究的问题得到深入的了解和启发，发现问题的内在规律和特征，验证目前的自然原理或规律(如量子电动力学的适用范围就是一个例子).在其他科学研究领域，计算所起的作用也越来越大.比如，在天文学上，Henyey算法为较大质量恒星的演变过程研究提供了有力的武器；在生物学方面，最近美国加州大学的Doolittle教授提出的关于进化的分子模型的建立也是依赖计算的.

1.2 剧烈的国际竞争

由于科学计算的重要性，世界各国都十分重视这一新领域.自从第一台电子计算机ENIAC 1946年在美国问世以来，美国一直在科学计算领域处于领先地位.即使这样，美国不少著名科学家还时常向政府呼吁，要充分重视科学计算领域的国际竞争.科学家的不断呼吁以及一些学术组织的报告，已使美国政府认识到科学计算对美国的重要性.1991年美国参议院提出了"高性能计算与通讯"的议案(简称HPC&C议案)，其主要内容是研制万亿次超级计算机、建设计算机高速通信网络和培养开发中的能力和提高工业

生产率，确保美国在高科技的优势地位和竞争能力.最近，美国又推出了 ASCI 计划，将完全用计算机模拟代替核试验.

在日本，计算科学和技术受到了极大的重视，近年来在计算机器件方面已经赶上了美国.看到美国的"HPC&C 计划"后，日本的科学界大为震动，提出了"超级计算——日本的生存之路"的论点.日本政府也很快制定了一个与"HPC&C 计划"相类似的计划，要在本世纪末之前建立 10 个超级计算中心和全日本计算机高速网络，大力开展科学计算的研究和应用工作.欧洲的科学计算研究一直处于国际前列.欧洲科学家在计算方法、计算机科学理论、网络设计技术等方面都有很大的贡献.在计算机硬件方面，最近有 16 家欧洲公司和科研机构联合制定了"欧洲工业进取计划"，旨在扩大欧洲地区制造和使用超级计算机的规模，结束欧洲工业依赖外来超级计算机的历史.

计算在国防上有举足轻重的作用，如武器研制、火箭设计与飞行轨道计算等都和计算密切相关.美国军方一直是计算机的大用户以及科学计算方法研究的支持者.在经济上，商用软件往往比计算机硬件贵得多.在相当长的时间内，发达国家将继续对我国封锁最先进的计算机技术和先进软件(包括系统软件和应用软件).我国一定要真正重视科学计算的作用.力争在下世纪 20 或 30 年代能自行设计及成批生产那时最先进的计算机，为我国科学计算跨入国际先进行列提供条件.

1.3 计算数学是科学计算的核心

科学计算离不开计算机，但它更离不开计算方法.美国著名的

计算数学家 Babuska 曾说过:“没有好的计算方法,超级计算机就是超级废铁.”人类的计算能力等于计算工具的性能与计算方法的效能的乘积.这一形象化公式表达了硬件与计算方法对于计算能力的同等重要性.美国计算数学家 Keller 和 Rice 曾提出例证,来纠正那种认为提高计算能力全靠硬件的错误观点.他们指出,从本世纪 50 年代到 70 年代末,计算机的运算速度提高了 5 个量级,与此同时,求解工程上大量出现的椭圆型偏微分方程算法的速度提高了 8 个量级.这种算法工效的巨大提高并不是罕见的事例.

要解决一个具体实际问题,首先要对它进行分析,用数学的语言描述它,得到它的数学模型.然后对该数学模型研究求解方法,以及应用这些求解方法求出模型的解,才能得到结果.对数学模型问题研究求解方法以及分析方法的性质就是计算数学的主要任务.由于一个计算问题的解决必须依赖于某一方法,由于它的解决的好坏以及解决的快慢取决于所用到的计算方法的优劣,所以完全可以说,计算数学是科学计算的核心.

计算数学属于应用基础研究范畴,它研究数值计算方法,包括计算方法的构造、方法的理论性质分析,利用数值方法求解实际问题以及通过计算研究问题的内在性质.在研究内容上,计算数学可分为数值代数、数值积分、数值逼近、微分方程数值解、固体力学计算、流体力学计算、最优化计算等方向.由于计算数学在全部计算性学科中所起的重要作用,它受到了越来越广泛的重视.

数值计算方法的研究虽然依赖于计算机,但更重要的还是靠科学家的脑力资源.所以,在我国计算机硬件相对落后的条件下,我们应扬长避短,力争在计算方法创新上取得优异成果,从而扩大我国科学计算在世界上的影响,为我国国民经济建设多做贡献.

2 牛顿法与分形

2.1 解方程是计算数学的最基本问题

一位有影响的数学家 J. B. Rosser 这样教导说："应用数学家的职责就是生产出有用的算法，使科学家能用以得到数字."的确，就应用的目的而言，数字是重要的.而得到数字的算法里头，通过解方程的方式提供的算法占了很大比重.这是很自然的事：我们所面对的系统中有很多数量，这些数量之间有一定的规则可以用于计算，这些规则就把这些数量分成两大类：作为计算出发点的数量和作为计算结果的数量.如果作为计算出发点的数量都是运用观察、测量、统计等直接手段所能得到的，而作为计算结果的数量又正是我们想得到而运用直接手段所不能或难以得到的，那么那些规则就直接提供了算法.如果情况不正好是这样，有一些(或全部)我们想得到的数量不在结果类里头而是在出发点类里头，那么它们

就不能直接通过这些规则计算得到.但是这些规则还是有用的,我们把想得到而又不在出发点类里的量一个个都用字母 x,y,…加以表示,我们的古人是用天元,地元,……表示,然后依照规则照样得出算式,这时,这些算式就叫做方程式;x,y,……等,或者天元、地元,……等,就叫做未知数或未知元.从方程式中把未知数算出来,就叫做解方程.

说它是未知数,其实可以不只是简单的数而是各种复杂得多的数学对象.计算规则中的运算,也可以不只是加、减、乘、除,而可以是各种复杂得多的数学运算,比方求微分,求偏微分,求积分等等.因此有各种各样的方程.研究各种各样方程解的性质:有没有?有多少个?它的性态是我们常见的还是奇怪的?等等,这是基础数学中好多个分支学科的主要任务.而把它们实实在在地算出来的方法拿出来,并且说清楚你这个方法有多少误差,那些作为计算出发点数据的偏差对你的计算有多少影响?计算中间的四舍五入对你的计算有多少影响?等等,这就是计算数学中好多个分支学科的主要任务.甚至可以说,在给出数字解意义下的解方程,这是计算数学的最基本的问题.

2.2 牛顿法是用处很大的解方程方法

没有万能的解方程的方法,否则这里就不需要数学家了.因为数学家工作的本质是创造而不是套用.他们要创造方法让人家套用.

但是有一个方法,就是牛顿法,那是用处很大的,差不多各种方程都能解.为什么差不多各种方程都能解?现在先把最简单的

情形，也就是最原始意义下的牛顿法说清楚.

牛顿名气很大.大家都知道牛顿发明了微积分.实际上，与牛顿同时独立发明微积分的还有一个德国数学家莱布尼茨.微积分的本质立足于对世界的这种认识：很多规律在微观上是线性的.所谓线性的就是成比例的.因此在每个局部可以用简单的比例关系来建立它的规律.但是并没有限定这是在哪一个局部，所以这样建立的规律是整体通用的.这种线性科学的观点和方法取得了辉煌的成功，大到比方说行星运动轨道的被认识，小到机械部件的设计，等等，都用这种方法.可以这样说，线性科学是迄今为止自然科学最成功的主体部分.当然，如果由于世界的多样性和复杂性触发了你非线性科学应当成为未来科学的主体的直觉，那你就坚定地投入罢！

把线性方法用在解方程上是这样的.设有一个实变量的实函数 $y=f(x)$，要求方程

$$f(x)=0$$

的解.假定估计解在 x_0 附近，那么 $f(x)-f(x_0)$ 的主要部分就和 $x-x_0$ 成比例，比例常数就是函数 f 在 x_0 的导数：

$$\lim_{x\to x_0}\frac{f(x)-f(x_0)}{x-x_0}=f'(x_0),$$

比例关系是：

$$(f(x)-f(x_0))\text{的主要部分}=f'(x_0)(x-x_0).$$

令 $f(x)=0$，就有

$$f(x_0)+f'(x_0)(x-x_0)\doteq 0\ .$$

所以

$$x\doteq x_0-\frac{f(x_0)}{f'(x_0)}.$$

虽然还不能说这样算得的

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

就是方程 $f(x)=0$ 的解，但它很可能比 x_0 更接近方程的解．然后，可以用 x_1 取代 x_0 重复同一过程而得出 x_2．如此反复，如果已经得出一个近似值 x_n，用 x_n 取代 x_0 而得出更好的近似值

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

这个式子中，n 从 0 开始，而 1 而 2，一直下去，直算到 $f(x_n)$ 绝对值小到可以忽略，就认为已经得到方程 $f(x)=0$ 令人满意的解的数值 x_n．

这就是牛顿法，也叫牛顿迭代法．现在这个方法还只能求解一个由实函数确定的方程．至于为什么差不多各种类型的方程都能解，我们下节再说．

总之，用牛顿法解方程是这样的：根据方程 $f(x)=0$，构造一个函数，叫做迭代函数，用 N_f 来表示迭代函数

$$y = N_f(x) = x - \frac{f(x)}{f'(x)}.$$

对根的已有近似 x_n，用迭代函数去修正成新的近似值 x_{n+1}：

$$x_{n+1} = N_f(x_n),\ n = 0, 1, \cdots.$$

因此从一个初值近似值 x_0 出发，得到了一个数列 $\{x_n\}$．如果这个数列收敛，那么它的极限 x^* 就满足

$$x^* = N_f(x^*).$$

我们说 x^* 是函数 N_f 的不动点．方程 $f(x)=0$ 的根 ζ 当然就是 N_f 的不动点：

$$\because \quad f(\zeta) = 0,$$

$$\therefore \quad \zeta = \zeta \quad \frac{f(\zeta)}{f'(\zeta)} = N_f(\zeta).$$

反之，在很多情况下，N_f 的不动点也必定是方程 $f(x)=0$ 的根．例如，当 f 是多项式时就是这样．

2.3 数学抽象的作用

数学的抽象是“乏味”的．比如，听到人家在唱美妙的歌曲：

“你来到我身边，/带着微笑，/带来了我的烦恼！/…她比你先到．”

一个有抽象癖的数学家把歌词抽象为一个定理：

“对随机拓扑空间 X 中的一个随机函数 $\zeta(t)$ 可以定义微笑 $\varphi(\zeta(t))=0$ 和烦恼 $\varphi(\zeta(t))=1$ 两种状态．关于随机函数 $i(t)$ 的状态有如下的判定定理：当随机函数 $y(t)$ 属于随机函数 $i(t)$ 的邻域且 $\varphi(y(t))=0$ 时必有 $\varphi(i(t))=1$，只要存在另一个随机函数 $s(\cdot)$ 和时刻 $t'<t$，使在时刻 t' 时 $s(t')$ 属于 $i(t')$ 的邻域且 $\varphi(s(t'))=0$．”

你看有多恶心！多不协调！

但是数学的抽象并非全是这些无聊的玩意儿．尽管由于同行竞争的加剧，量化管理的盲目，学术评论的缺乏等说得清楚和说不清楚的因素以及一些固有的因素造成了学术界的一些不健康现象，出现了一些毫无意义的论文，但数学研究的主流是很有生机和富有创造性的．“就其本质而言，数学是抽象的；实际上它的抽象比逻辑的抽象更高一阶”(G. Chrystal)，因为数学的抽象追求的是和谐与统一，而和谐与统一的威力是巨大的．

现在就来说一说上节留下来的问题：为什么牛顿法差不多各种方程都能解？这就是数学抽象的作用．大约在十个世纪末，数学家们发展了一门新学科，叫做泛函分析．这门学科首先把某些同类的数学对象，比方一些数组，一些函数，等等看成一些"点"，然后在这些"点"之间定义一个非负的数作为"距离"，当这种"距离"符合一定的规则时（这些规则就是从生活现实的空间抽象出来的），这些同类的点的全体就构成一个"距离空间"．一个完备的距离空间叫做巴拿赫空间，所谓完备就是当一个无穷点列互相之间愈来愈靠近时，它们一定是向一个点靠近而不是向一个空位置靠近．然后再用一定的规则从加于点上的各种运算中分离出"算子"和"可微算子"，就像普通的函数和可微函数那样．

可以用 f 表示一个算子，用 f' 表示 f 的导数．经过这样的抽象，现在形式上写出的方程

$$f(x)=0$$

已经可以包含很多类型的方程了．奇怪的是，现在牛顿法对很多类型的方程居然也照样能用．当然，这时的牛顿迭代公式应当写成

$$x_{n+1}=x_n-f'(x_n)^{-1}f(x_n),$$

这里 $f'(x_n)^{-1}$ 表示 $f'(x_n)$ 这个算子的逆算子．

为了让大家看清楚这一飞跃有多大，我们来仔细研究一下怎样用牛顿法解二元方程组

$$\begin{cases} g(u,v)=0, \\ h(u,v)=0. \end{cases}$$

把二元数组 (u,v) 看成一个点 x，就是平面上的一个点吧！两个点之间的距离就是普通测量的距离．这时的平面就是一个巴拿赫空间．算子 f 作用在 x 上，产生一个点 $f(x)$，它的两个坐标是 g

$(u,v),h(u,v)$.这样一来,这个二元方程组就成了规定的巴拿赫空间中的一个方程 $f(x)=0$,f 的导数是矩阵:

$$f'(x)=\begin{bmatrix}\dfrac{\partial g}{\partial u} & \dfrac{\partial g}{\partial v}\\ \dfrac{\partial h}{\partial u} & \dfrac{\partial h}{\partial v}\end{bmatrix},$$

设这个矩阵当 $x=x_n=(u_n,v_n)$ 时的逆矩阵为

$$f'(x_n)^{-1}=\begin{bmatrix}a_n & b_n\\ c_n & d_n\end{bmatrix},$$

且设 $g(u_n,v_n)=g_n$,$h(u_n,v_n)=h_n$,则这时牛顿迭代公式就是

$$\begin{cases}u_{n+1}=u_n-a_ng_n-b_nh_n,\\ v_{n+1}=v_n-c_ng_n-d_nh_n.\end{cases}$$

一种形式的过渡能使解一元方程的方法去解二元方程,多么令人惊奇!其实还远不止此.也完全是通过形式的过渡,可以让它去解 n 元方程、常微分方程、偏微分方程、积分方程,等等,你看,抽象的威力大吧!

发现这件事并建立起关于它的完整理论的是前苏联的世界著名数学家康托洛维奇.他和荷兰数学家 Koopmans 分享 1975 年诺贝尔经济学奖,因为他们发明的一种叫做线性规划的数学方法有很大很大的经济效益.

2.4 代数方程

代数方程是大家最熟悉,但也是误会最多的一个数学领域.

"高次代数方程无解",许多工程师都这样说.不对!牛顿以后最大的数学家高斯年轻时候在他的博士论文中就已证明,在复数

范围内，每一个代数方程都至少存在一个解，这是在1799年，几乎两个世纪前．这个结论很著名，叫做代数学基本定理．由这个定理我们立即可以得出结论：在复数范围内，一个d次的代数方程正好有d个根．当然，这时重根要计算与其重数相同的次数．

代数基本定理是数学上确定一个研究对象的存在性的一种典范．从高斯的证明中看不出根在哪里，以及如何找到它，但确信它的存在．这是一种单纯的存在性．与之相对，另一种确定存在的典范是给出构造它的方法．这后一种典范称为构造性的．

存在性和构造性体现了数学研究方法和主题的两种截然不同的风格．在数学发展的历史上，可以看到这两种风格交替地成为各个时期的时尚，正是这种交替充分发挥了这两种风格的互补性，有效地推动着数学的发展．因此，任何扬此抑彼的主张都是片面的，不健康的，有害的．

回过头来说代数方程．二次方程的求根公式是大家所熟悉的．一些二次方程的解法在古巴比伦时代就已为人们所知．因此三次和四次方程的解法在很长的历史时期都是人们探讨的对象，到了16世纪，终于找到了它们的解法．此后的3个世纪，人们又把努力的目标指向五次和五次以上的方程．但是现在情况已经起变化．因为对五次和五次以上的方程，并不存在可以用由加、减、乘、除、乘方和开方运算构成的公式，来表示它的解．

这个事实是19世纪初天才的法国数学家伽罗华（1811～1832）用严格的理论证明了的．他为此发明了群论这门新学科．初看起来，群论和方程的解毫无关系，所以他的理论当时的人包括大数学家在内都不理会，几十年后才获得知音，可是他早已饮恨九泉（他是一个激进的学生，死于决斗）．群论现已证明不仅在方程式

论、微分几何、数论和调和分析等数学分支中很有用处，而且在光谱学、结晶学、原子物理和粒子物理中也大有用武之地．这是数学威力的一个很典型的实例，充分显示了数学真理的客观性．

没有公式解，但高次方程不能不解，于是就求数值解，牛顿法又派上用场了．

大不了，把复数看成平面上的一个点．我们已经知道怎样用牛顿法求解由平面上的点到点的函数的方程．其实，复变函数论中的一个叫做柯西－黎曼条件的公式解释了上节中作为二元函数导数的矩阵在复数运算时是非常自然而和谐的．所以更有理由直接把牛顿法用于求解高次代数方程．也就是说，用它最原始的形式，只是这时的自变量 x 是一个复数 $x=u+iv$，导数也用它自然的形式，只要进行复数四则运算就行．

复数 $x=u+iv$ 在 uOv 平面上可以用一个点 (u,v) 表示．这个平面现在大家习惯上叫它作高斯平面，它对于帮助大家理解复数的客观性是大有功劳的．

2.5 初始近似值

在牛顿法的公式中，x_0 称为迭代的初始近似值．怎样选取初始近似值 x_0，这在任何形式的牛顿迭代中都是一个困难的问题．下面我们就专门研究初始近似值对整个迭代过程的影响．

用实验的办法，挑选一些根都已知的方程，然后在整个高斯平面上做实验，也就是把平面上的每一点都依次取作初始近似值，看看它的牛顿迭代有怎样的行为．当然，马上就碰到有限与无限的矛盾．这个矛盾表现在两个方面：①平面是无限的，只能在有限的一

块(称为"窗口")上做实验;②即使在有限的"窗口"上,点是没有面积的,不可能取遍窗口中所有的点,只能相隔一定距离地取.

对于后一点,大家可能会觉得没有问题,只要把相隔距离取得足够小,比方监视器视屏上的光栅之间的距离就可以,就认为是这样吧.对前一点怎么办?这在数学上有个克服的办法,那就是在高斯平面上添上一个无穷远点,然后把它像用皮子包包子那样捏成一个球面("捏"的动作在数学中用叫做麦比乌斯变换的运算来完成),用皮子包包子,上面总留下一个口子,肉包子就是这样的;但如果是糖包,糖会从这个口子流出来,非得用特殊的办法封上这个口子,"添上一个点",这才成了一个球面.这个球面,叫做黎曼球面.无穷远点在黎曼球面上,就像普通点一样了.你看糖包子,处处都光溜溜的.

例 1 方程

(1) $x^2-1=0$;

(2) $(x-1)(x+1)^2=0$.

第一个方程有两个根

$$\zeta_1=1,\qquad \zeta_2=-1.$$

两个根都是单根,第二个方程也有同样的两个根,不过 ζ_2 是二重根.对于第一个方程,我们将看到,当初始近似值 x_0 取在右半平面时,牛顿法所产生的数例 $\{x_n\}$ 都向 ζ_1 收敛,当初始近似值 x_0 取在左半平面时,牛顿法所产生的数例 $\{x_n\}$ 都向 ζ_2 收敛,两个区域的分界线很清楚,是虚轴 $\mathrm{Re}\,z=0$.对于第二个方程,情况基本相同,只是分界线皱了一点,并向右半实轴弯了一点,这使得向 $\zeta_1=1$ 收敛的初始近似值 x_0 的区域变小一点,相反向 $\zeta_2=-1$ 收敛的初始近似值 x_0 的区域变大一点,这很公平合理:ζ_2 是二重根么.

势力大一点.

例 2 方程

(1) $x^3-1=0$;

(2) $x(x^2-1)=0$.

第一个方程有三个根

$$\zeta_1=1,\quad \zeta_2=-\frac{1}{2}+i\frac{\sqrt{3}}{2},\quad \zeta_3=-\frac{1}{2}-i\frac{\sqrt{3}}{2}.$$

这三个根在单位圆周上,成等边三角形.

第二个方程也有三个根

$$\zeta_1=1,\quad \zeta_2=0,\quad \zeta_3=-1.$$

这三个根都在实轴上,等距离分布.

请大家先猜一猜:这时各自向三个根收敛的初始近似值 x_0 是怎样分布的?分割它们的边界是怎样的?

如果一条边界线把平面分成三个或三个以上的区域,那么边界上大多数的点都有这样的性质:它的小邻域内只有两个不同区域的点;它的小邻域内有多于两个不同区域的点的边界点只是少数派.这是大家熟悉的情况.

现在请大家看彩图 2-1(a)和彩图 2-1(b),两个图上各有三个真正的圆,它们各自以一个根为圆心.这些圆中的点,是用数学的方法能够预先确定牛顿法必定向这个根收敛的初始近似值 x_0.圆的外面有一个接近于圆的圈,如果取其中的点为 x_0,牛顿法只用一步就跑进里头的圆,这个圈的外头是比它更偏离圆的圈,如果取其中的点为 x_0,那么牛顿法第一步跑向里头的圈,再一步跑向里头的圆,等等.因此,每个图都有三个大的联成一片的区域,如果取其中的点为 x_0,牛顿法向位于区域内的根收敛.这些联

成一片的区域我们称之为根的直接吸引域.

界于根的直接吸引域之间的地带非常复杂,它们将依次跑向颜色组合相同的直接吸引域.因此,对一个根来说,如果用 D_0 来表示它的直接吸引域,那么必有

$$N_f(D_0)=D_0.$$

也就是说,在牛顿迭代函数 N_f 的作用下,直接吸引域是不变的,而其余的块,在牛顿迭代函数 N_f 的作用下,必定会逐步成为直接吸引域 D_0,即最终不变的.把所有这些向同一个直接吸引域跑的块放在一起,虽然并不联成一片,但正好构成了向一个根收敛的初始近似值 x_0 的全部,称为相应根的吸引域.如果用 D 来表示一个根的吸引域,那么

$$D=N_f(D)=N_f^{-1}(D).$$

$N_f^{-1}(D)$的意思是,使 $N_f(x)$在 D 里的所有 x.上面这个式子意思是,一个根的吸引域 D 在牛顿迭代函数 N_f 作用下是完全不变的.

因此,我们已经在每个图上都发现了三个完全不变的区域.

现在来看三个吸引域的边界,这就是彩图 2-2(a)和彩图 2-2(b)两个图.图中的曲线很复杂,理解其复杂性的关键在于,曲线上的每个点,它的小邻域内都有三个吸引域中的点(对照彩图 2-1(a)和彩图 2-1(b)来看).如果用 J 来表示这些边界点的全体,那么 J 也有这样的性质:

$$J=N_f(J)=N_f^{-1}(J).$$

即,J 在牛顿迭代函数作用下也是完全不变的.但是 J 里面已经没有方程的根,所以 J 里面的点在牛顿迭代的作用下永远不会跑到根的吸引域中去,这表明从 J 中取初始近似值 x_0,牛顿法必然会

失败.幸好,J 的点是占地不多的,用数学上的术语来说,J 的测度为零.

一门崭新的科学正在兴起,这就是混沌、分形和非线性科学.在牛顿迭代函数作用下,J 中的点是乱跑的,毫无规则,这就是混沌.高斯平面上,除了 J 以外的点,在牛顿迭代函数作用下,跑得很有规则,它们统称稳定集.在例子中,稳定集有三个完全不变分支.每一个完全不变分支中,每个联成一片的部分,在牛顿迭代下,要么是不变的,即根的直接吸引域,要么是最终不变的.

在 J 中的每一点,它的小邻域内都有三个完全不变分支的点,这体现了 J 的分形性质.

混沌和分形属于非线性科学,有人认为这是新世纪的科学.但由牛顿法这一线性科学的典型方法竟直接导出非线性科学的一些现象,这说明科学的连贯性.说实在的,对科学的探索需要大家持久的、连续的努力,用赶时髦的办法走捷径,其实是成不了大事的.

必须说明,迭代函数是有理函数.研究在有理迭代下的动力行为,必须放到黎曼球面上而不只在高斯平面上,它们相差一个无穷远点.现在,看彩图 2－1(a)和彩图 2－1(b)两个图,角上都有一个小图,小图反映了黎曼球面上无穷远点附近的情况.

以上两个例子太规则,现在看两个随机的例子.一个是取北半天球上九大星座的 α 星在天球上的视位置作为根,构成一个九次多项式,这九个星座是:

小熊座(α 星即北极星),御夫座,英仙座,仙后座,仙王座,天鹅座,天琴座(α 星即织女星),天龙座(α 星即右枢),大熊座(α 星即天枢).

另一个是以北半球的九大城市的地理位置作为根,构成一个

九次多项式，这九个城市是：

北京，上海，东京，纽约，伦敦，巴黎，柏林，伊斯坦布尔，莫斯科．

例 3 两个随机的九次多项式：

(1) 根是九大星座的 α 星在天球上的位置；

(2) 根是九大城市在北半球的位置．

这时，在牛顿迭代下，九个根的吸引域如彩图 2－3 所示．

以上三个例子，都有一个共同的特点，就是除了占地不多的边界，整个高斯平面上的点都是使牛顿法收敛的初始近似值．

那么，有没有整块的区域存在，在那里取初始近似值牛顿法都会失败呢？

例 4 用牛顿法解方程

$$x^3 - 2x + 2 = 0,$$

这个方程的牛顿法的动态图如彩图 2－4 所示．那些白色的块是由使牛顿法不收敛的初始近似值 x_0 构成的．

这个例子是当代美国大数学家斯梅尔刻意构造的，我们写出它的牛顿迭代函数：

$$N_f(x) = x - \frac{x^3 - 2x + 2}{3x^2 - 2} = \frac{2x^3 - 2}{3x^2 - 2}.$$

容易发现

$$N_f(0) = 1, \qquad N_f(1) = 0 .$$

这是一个死循环，因此，至少在 $x = 0$ 和 $x = 1$ 附近，牛顿迭代是不会收敛的．

对那些白的块，在这个牛顿迭代函数作用下，有两块 A 和 B，这样：

$$N_f(A) = B, \qquad N_f(B) = A .$$

这两块称为 2－周期区域．其余的呢？在 N_f 作用下，最终会变到 A 和 B 上去，所以称为最终 2－周期区域．

所有区域的边界如彩图 2　5 所示．A 是高斯平面上的边界，B 是黎曼球面上无穷远点附近的边界．

2.6　沙列文等的数值实验

沙列文(D. Sullivan)是斯梅尔的学生，他和 J. Curry 等在 1983 年发表的一篇论文阐述了他们三个人的一个数值实验的结果．数值实验的目的是系统地寻找这样的三次多项式，对它们的牛顿法在高斯平面上有大块使之失败的初始近似值．

他们把一般的三次方程化简为

$$f_a(x)=x^3+(a-1)x-a,$$

这里 a 是一个复数．可以证明这种化简是不妨碍一般性的．这个三次方程有实根 $x=1$ 和另外一对共轭复根．因此在牛顿迭代之下，稳定域至少已有三个完全不变分支即三个根的吸引域．如果只有这些，那么牛顿迭代就不存在大块使之失败的初始近似值．因为这些初始近似值也将构成一个稳定域的一个完全不变分支，已经知道，它们是由周期区域(周期大于 1)和最终周期区域构成的．根的吸引域是不变域，也即 1－周期域．

怎样去发现周期域？早在本世纪 20 年代法国数学家法都就有一个定理，这个法都定理说：在有理迭代下，每个周期域，都是从一个临界点跑过去的．什么是一个有理迭代 $r(x)$ 的临界点呢？大致上说，就是 $r(x)$ 导数 $r'(x)$ 的根．

好！现在对牛顿迭代函数

$$N_f(x) = x - \frac{f(x)}{f'(x)}$$

来说，

$$N_f'(x) = 1 - \frac{f'(x)^2 - f(x)f''(x)}{f'(x)^2} = \frac{f(x)f''(x)}{f'(x)^2},$$

所以除了方程的根是临界点外，$f''(x)$的根也是临界点．方程的根作为临界点不必管它，因为从它当然跑向自己的吸引域，即 1－周期域．问题是 $f''(x)$的根将跑向哪里？这是关键．现在

$$f''_a(x) = 6x,$$

它的根是坐标原点．因此，在参数 a 的平面上作图，看 a 对应的方程，在牛顿迭代下坐标原点向哪里跑？请看彩图 2－6．白色、红色和绿色分别表示原点向三个相应根的吸引域跑的参数 a．如果不是这样，就把参数 a 的位置染上紫蓝色．结果，发现到处有紫蓝色的小东西，就像到处有细菌一样．见彩图 2－6(a)．在彩图 2－6(a)最明显的地方放大，这是彩图 2－6(b)．在中轴不明显的地方放大，也有这种形状的东西，这就是彩图 2－6(c)．

在这些蜘蛛状的紫蓝色区域中取参数 a，所对应的三次方程 $x^3 - (a-1)x + a = 0$ 用牛顿法来解时，都有大片使之失败的初始近似值，而不只是占地很少的边界．

参数平面图，即彩图 2－6，是各种三次方程的一个目录，包括没有整片初始近似值使牛顿法失败和有整片初始近似值使牛顿法失败的．

2.7 不好的初始近似值区域的类型

沙列文等的著名实验表明，有很多的三次方程，使它们的牛顿

迭代失败的初始近似值联成一片(形成一个连通区域).对这些方程来说,这些一个个连通区域的总体形成稳定域的一个完全不变分支.在牛顿迭代下,每个连通区域要么是周期的,要么是最终周期的.这些周期区域是有代表性的,因为最终周期区域总是要跑向周期区域的.

周期区域一定和一个周期点有关.所谓周期点就是迭代函数作用若干次之后又回到自身的点.作用次数称为周期.

在有理迭代下有多少种类型的周期区域?沙列文对此作了详尽的研究,表明总共有五种,所以这些区域类型叫做沙列文域.最新的研究表明,多项式的牛顿迭代,作为一种多少有些特殊的有理迭代,它的沙列文域总共只有四种类型.

可仿照上面的参数平面图,把牛顿迭代具有各种类型沙列文域的三次方程列一个目录,即取一个参数 a,但 a 依赖于 λ:

$$a=\frac{1}{2}\sqrt{\frac{36-\lambda}{4-\lambda}}.$$

然后根据 λ 作三次方程

$$f_\lambda(x)=x^3+\left(a-\frac{3}{2}\right)x^2-\left(a+\frac{1}{2}\right)x+\left(a+\frac{1}{2}\right)=0.$$

这个三次方程的牛顿迭代的临界点,除了方程的根之外,是

$$c=\frac{1}{2}-\frac{a}{3}.$$

把在牛顿迭代下临界点 c 不跑向三个根的吸引域的 λ 用紫蓝色染出来,见彩图 2-7.在这个连体蜘蛛中,参数 λ 的各种值反映了这四种类型的沙列文域(大蜘蛛身体的大圆是参数 λ 的绝对值不超过 1 的平面区域),列举如下:

(1) $\lambda=0$,称为超吸性域.此例的超吸性域是 2-周期的.它

包含一个2-周期点，同时是临界点.(彩图2-8(a).彩图2-8(a)L,彩图2-8(a)R)

(2) $\lambda=\frac{2}{3}(1+i)$.称为吸性域.此例的吸性域是2-周期的.它包含一个2-周期点，同时包含一个不是周期点的临界点.(彩图2-8(b)L,彩图2-8(b)R)

(3) $\lambda=\exp(-2\pi\alpha i)$，而 α 是有理数，称为抛物型域.抛物型的边界上有一个周期点，设为 n-周期点，即把迭代函数作用 n 次看成一个函数的话，它的一个临界点被包含在某个以 n-周期点为边界的抛物型域之中.所给出的图中，$n=2$，而 α 的取值有三例：

1) $\alpha=1$.抛物型域是2-周期的.(彩图2-8(c)1L、彩图2-8(c)1R)

2) $\alpha=1/2$.抛物型域是4-周期的.(彩图2-8(c)2L、彩图2-8(c)2R)

3) $\alpha=2/3$.抛物型域是6-周期的.(彩图2-8(c)3L,彩图2-8(c)3R)

(4) $\lambda=\exp(-2\pi\alpha i)$，$\alpha=\frac{\sqrt{5}-1}{2}$.称为齐格尔(Siegel)盘.此例的齐格尔盘是2-周期的，它包含一个2-周期点，而临界点则在边界上.(彩图2-8(d)L,彩图2-8(d)L1,彩图2-8(d)R)

所有的彩图，图号中含有L的是在周期点0附近、图号中含有R的是在周期点1附近.例如，容易在彩图2-8(a)中找到彩图2-8(a)L和彩图2-8(a)R的位置.

3 动力系统的几何算法

动力系统是现代化力学、物理学、化学、生物学乃至社会科学中许多发展规律所服从的数学方程，对于其行为状态的研究除解析方法外，数值方法是不可缺少的工具．科学与工程中的动力系统多属于非线性，它与线性系统有本质的区别，其性态非常复杂，正如 Arnold 所说的，"非现代数学工具所及"，现代计算机提供了一个强有力的工具，使得动力系统研究有了新的发展．

3.1 从正确计算牛顿运动方程谈起

1991 年春天，冯康在北京中国物理学年会上曾作了一个"怎样正确计算牛顿力学方程"的精彩报告．

我们知道，当代科学计算的主题是数值解算这样或那样的数学物理方程．在众多的数理方程中列于首位的自然是牛顿运动方程，即表达 $f = ma$ 的

二阶常微分方程组.微分方程的计算方法,过去由于历史条件的限制发展曾较慢,但自欧拉开始,历经亚当斯、Runge、Kutta 以及斯笃默等的贡献,特别在进入计算机时代后,有了很大进步,积累发展了丰富多样的算法和软件包.有人说,"三体问题"已经变得不重要了,计算机可以解决.在这一情况下,我们还是提出下面两个问题:

问题一:现有算法对计算牛顿运动方程究竟是否合适?

问题二:怎样才能正确计算牛顿运动方程?

关于问题一,看来从没有人认真提出过;从而对于问题二看来也从没有人系统探讨过.我们主要研究物理上更为根本而数学上也较难的守恒型牛顿运动方程.首先,守恒牛顿方程还有另外两种等价的数学形式,即拉格朗日变分形式和哈密顿形式,后者把位置空间内的二阶运动方程组化为相空间内的一阶正则方程组.不同的数学等价形式表达同一物理规律,但因外形的差异,对于"解问题"自然启发不同的技术途径,从而在实践上是不等效的.因此,从不同的等价数学形式中作出合理明智的选择对于解题的难易成败是至关重要的.

选定哈密顿形式为基本形式,动机是:①哈氏方程具有非常对称利落的形式,运动的规律性在哈氏形式下表现得最明显.②哈氏形式有远比牛顿形式为大的遍在性和普适性,它覆盖了经典性的、相对论的、量子性的、有限或无限自由度的一切真实的、耗散效应可忽略的物理过程.因此如果对经典哈氏方程的算法研究取得成功,则可望有广阔发展应用的前景.为此人们曾对浩繁的文献进行查寻,却无所获,有关哈氏方程算法的研究几乎是空白,令人费解.这促使我们认真思考并力求取得前述问题的解答.

我们采取的技术途径是辛几何，即相空间的几何．它的基点是反对称的面积度量，与基点为对称的距离度量的欧氏和黎曼几何并立相对．经典力学基本定理用辛几何的语言就表为“一切哈氏体系的动力演化都使辛度量保持不变，即都是辛（正则）变换”．因此解哈氏方程的“正确”的离散算法就应是辛变换，这样的方法叫做辛（正则）算法或哈密顿算法．我们有意识地在辛几何框架内对哈氏算法的构思、推导进行分析、评估．事实证明这条途径是成功的、是卓有成效的．推出了多种多样的辛算法系列．摸清了它们的性质，奠定了它们的理论基础，通过了严峻的实算考验．

为了比较非辛与辛算法，我们提出了八个“考题”：谐振子、非线性振子、惠更斯振子、卡西尼振子、二维多晶格与准晶格定常流、利萨日图形、椭球面测地线流、开普勒运动．计算机实验无可置疑地肯定了辛算法的高质量和优越性．特别在有关整体性、结构性和长期跟踪能力等方面，一切传统非辛算法、不论精度高低，都无例外地全然失效，而一切辛算法，不论精度高低，则全部无例外地过关，均拥有长期稳健的跟踪能力．对比强烈鲜明，显示了压倒的优越性，参看下述的计算对比图例．

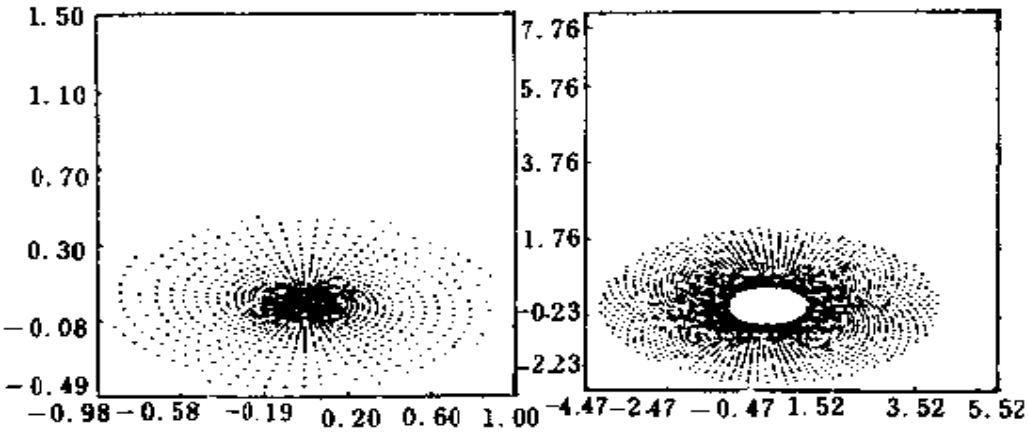
传统算法除极个别例外均非辛，大都是面向渐近稳定系统设计的，都含有耗散机制以保证计算稳定性，哈氏系统不具有渐近稳定性，这些算法不可避免地带进人为耗散性、虚假吸引子以及其他种种非哈氏系统本有的寄生效应，最终导致严重歪曲失真．因此用于短期的、瞬态的模拟尚可，用于长期跟踪和整体的结构性研究则不行，会引向错误的结论．既然牛顿方程等价于哈氏方程，那么问题一就得到颇出人意料的否定性答案．

辛算法不含人为耗散性，先天性地免于一切非哈污染，是“干

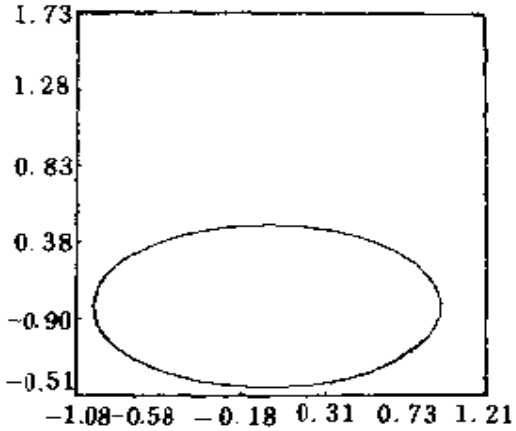
净"的算法.哈氏系统拥有两类守恒律.一类是相空间内偶数面积的不变性,即刘维尔－庞加莱守恒律.在辛算法下都被自动保持.另一类是包括能量在内的运动不变量如动量、角动量守恒律等.我们证明了一切辛算法都拥有自己的形式不变量,它们对于原有不变量的逼近阶与算法本身的逼近阶相当.我们也初步证明了在辛算法下近似可积系统的不变环大部分得到保持.这是著名的KAM定理的一种新模式.这一切都表明离散型哈氏算法的体系结构与守恒律完整并行,高度靠近于哈氏原形,而且拥有理论上无限长期的跟踪能力,这就从根本上保证并说明了哈氏算法的独到性能和优越性.因此一条正确计算牛顿运动方程的途径就是先把方程哈密顿化,然后运用哈氏算法,这就是问题二的答案.

1. 谐振子、椭圆轨道

图3－1(a)Runge－Kutta法,步长0.4,0.3万步,人为耗散,轨道收缩,图3－1(b)亚当斯法,步长0.2,人为反耗散,轨道发散,图3－1(c)二步中心差法,对线性方程为辛,步长0.1,1000万步,初、中、末三个0.1万步轨段重迭打印,完全吻合.



(a) (b)

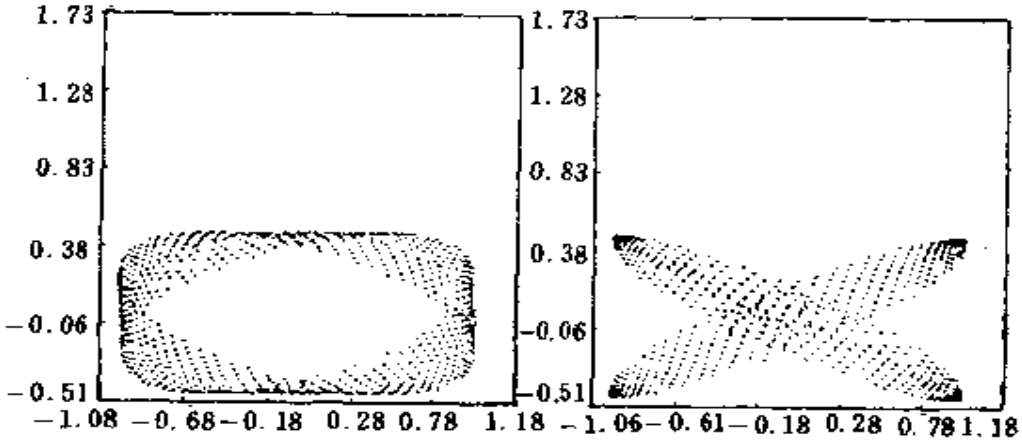


(c)

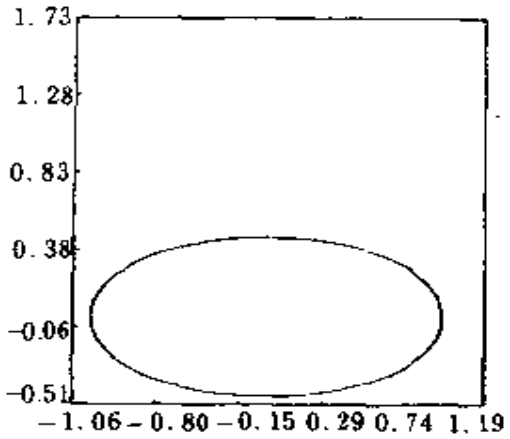
图 3－1　谐振子,椭圆轨道

2. 非线性振子,拟椭圆轨道

图 3－2(a)及 3－2(b)二步中心差法,对非线性方程为非辛,步长 0.2,1 万步,图 3－2(a)为最初 0.1 万步轨段,图 3－2(b)为 0.9～1 万步轨段,都显示了轨道的失真,图 3－2(c)二阶辛算法,步长 0.1,0.1 万步,初、中、末三个轨段重迭打印,完全吻合.



(a)　　(b)

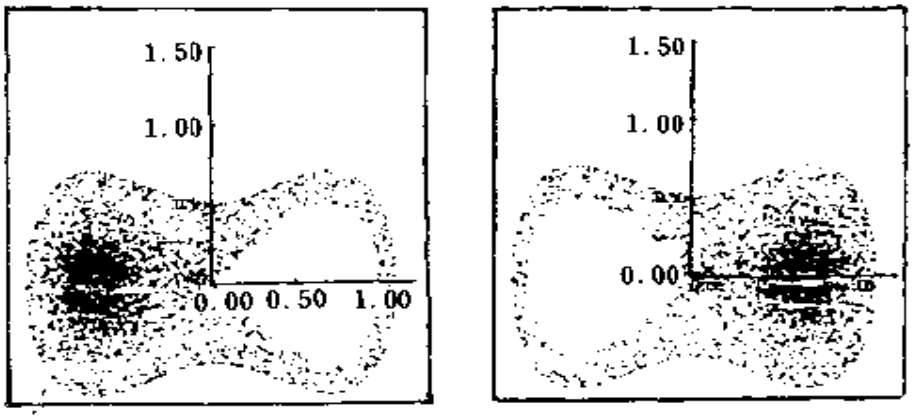


(c)

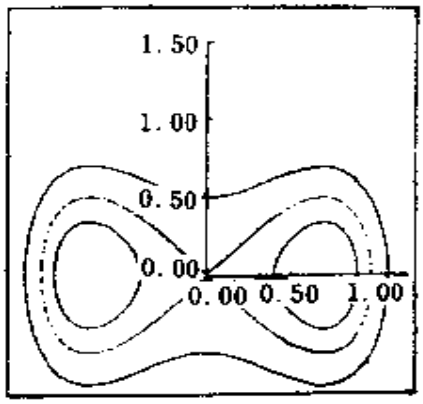
图 3-2　非线性振子拟椭圆轨道

3. 惠更斯振子，轨道为惠更斯卵线，分界线为惠更斯双纽线

图 3-3(a)Runge-Kutta 法，把横轴上左右两个不动点变为两个假吸引子，位于双纽线外的任意初始相点趋于左右两个吸引子的概率均等，图 3-3(a)步长 0.10000005，90 万步，趋向左，图 3-3(b)步长 0.10000004，90 万步，趋向右，图 3-3(c)二阶辛算法，步长0.1，四根典型轨道，各10000万步，每根轨道的初、中、末



(a)　　(b)



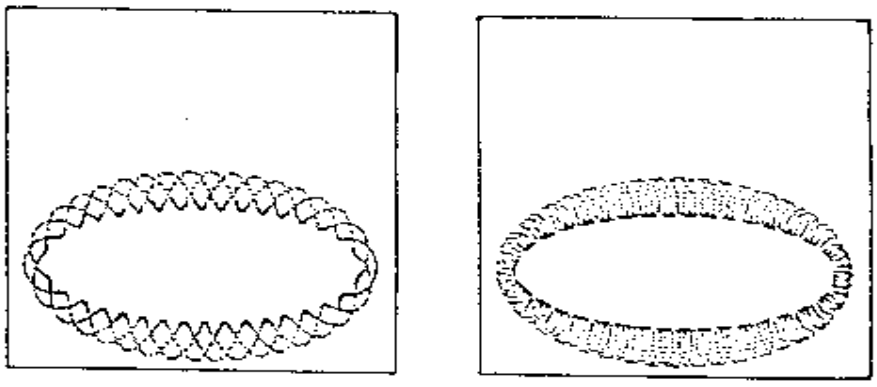
(c)

图 3－3　惠更斯振子，轨道为惠更斯卵线，分界线为惠更斯双纽线

三个 0.05 万步轨段重迭打印，完全吻合，显示了超长期跟踪能力.

4. 椭球面上测地线，无理频率比，稠密轨道.

频率比平方 5/16，步长 0.05658，1 万步，图 3－4(a) Runge－Kutta 法，不趋稠密，图 3－4(b)辛算法，趋于稠密.

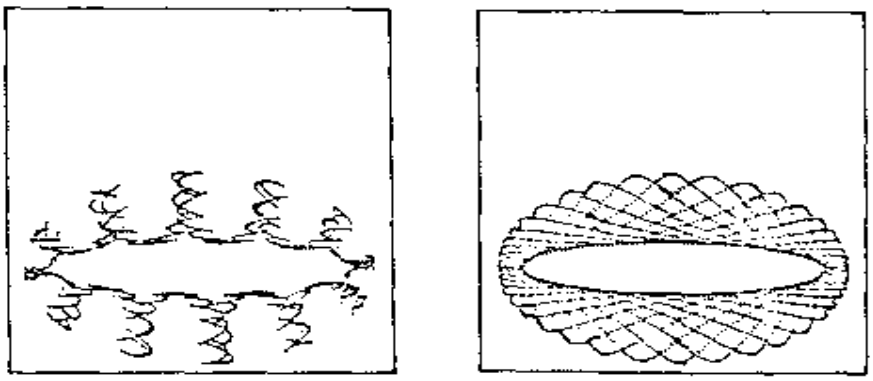


(a)　　(b)

图 3－4　椭球面上测地线，无理频率比，稠密轨道

5. 椭球面上测地线，有理频率比，封闭轨道

频率比 11/16，步长 0.033427，10 万步，25 周期，图 3－5(a) Runge－Kutta 法，不封闭，图 3－5(b)辛算法，封闭.

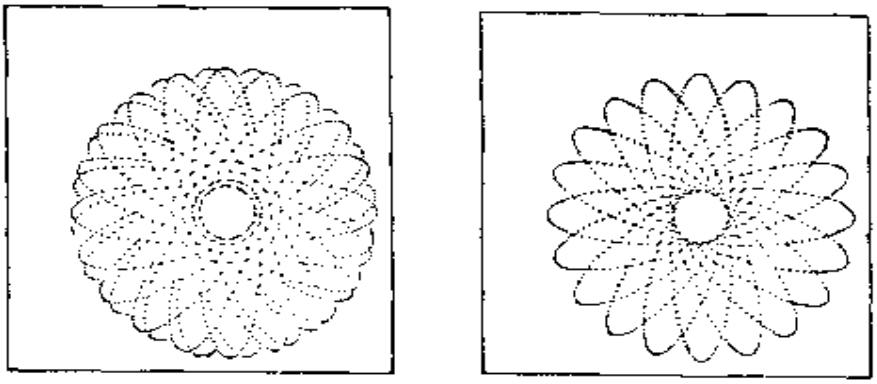


(a)　　(b)

图 3－5　椭球面上测地线，有理频率比，封闭轨道

6. 开普勒运动，有理频率比，封闭轨道

频率比 11/20，步长 0.01605，24 万步，60 周期，图 3－6(a) Runge－Kutta 法，不封闭，图 3－6(b)辛算法，封闭.



(a)　　(b)

图 3－6　开普勒运动，有理频率比，封闭轨道

3.2 哈密顿力学发展的历史

首先，看看经典力学的三种形式．考虑有 n 个自由度的运动，位置向量记为 $q=(q_1,\cdots,q_n)$，势能函数为 $V=V(q)$，则

$$m\frac{\mathrm{d}^2q}{\mathrm{d}t^2}=-\frac{\partial}{\partial q}V$$

是运动方程的标准形式，它是在 n 维位形空间 R^n 中的二阶微分方程组，通称为经典力学的标准形式，亦称牛顿形式．

欧拉及拉格朗日通过引进动能与势能差数的作用量

$$L(q,\dot{q})=T(\dot{q})-V(\dot{q})=\frac{1}{2}(\dot{q},M\dot{q})-V(q),$$

并利用变分原理将运动方程写成如下等价形式

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial\dot{q}}-\frac{\partial L}{\partial q}=0,$$

它被称为经典力学的变分形式即拉格朗日形式．

到了 19 世纪，哈密顿又提出另一种形式．他利用动量 $p=M\dot{q}$ 和总能量 $H=T+V$ 将运动方程描述为

$$\dot{p}=-\frac{\partial H}{\partial q},\qquad \dot{q}=\frac{\partial H}{\partial p},$$

它称为哈密顿正则方程，这是 $2n$ 维相空间或称辛空间中相变量 $(p_1,\cdots,p_n,q_1,\cdots,q_n)$ 的一阶微分方程组，具有极其简单而对称的形式．

经典力学的这三种基本的方程形式在几乎一切理论物理或理论力学教科书中都有介绍．这些不同的数学形式是陈述同一物理规律，由于形式不同，它们对实践上的“解题”(problem solving)自然会提供完全不同的技术途径，因此等价的数学形式，实践上可以

是不等效的，对此我们也有自己的体会.

冯康在早年曾研究了有限元计算方法，这是为解决平衡问题的系统化算法，这类物理问题在数学上有两种等价形式：一种是牛顿形式即解二阶椭圆型方程，另一种是变分形式即能量泛函极值原理.有限元方法在实践上与理论上取得成功的关键在于合理的选取变分形式作为基础.当冯康提出有限元及其基础理论之后，试图把这套方法思想应用到连续介质动态问题，但没有取得而且看来也难望取得相应的成功，故对动态问题的计算方法而言，拉格朗日系统可能不是合理的选择，而且对于牛顿形式同样也难寄期望.因此合理的选择很可能应是哈密顿系统.这当然只是设想，还得通过实践来考验.还得从历史上来考察人们对哈密顿系统的评价.首先应该指出，哈密顿本人是从几何光学着手创建他的理论模式的，而后才转向与光学相距甚远的力学.1834 年哈密顿曾说“这套思想与方法业已应用到光学与力学，看来还有其他方面的应用，通过数学家的努力还将发展成一门独立的学问”.这仅仅是他本人的期望.19 世纪同代人对其反应则很冷淡，认为这套理论“漂亮而无用”.著名数学家 Klein 在对哈密顿形式的理论给予很高评价的同时，对其实用价值亦持怀疑态度，他说“这套理论对于物理学家是难望有用的，而对工程师则根本无用.这种怀疑，至少就物理学的范畴而言，是被随后的历史发展所完全否定了，本世纪 20 年代量子力学却正是在哈密顿形式的框架下发展起来的.量子力学创始人之一 Schrödinger 曾说：“哈密顿原理已经成为现代物理的基石……如果您要用现代理论解决任何物理问题，首先得把它表示为哈密顿形式.”

3.3 哈密顿体系的重要性

哈密顿体系是动力系统的一个重要体系，一切真实的耗散可忽略不计的物理过程都可表示成哈密顿体系.哈氏体系的应用范围很广，它包括结构生物学、药理学、半导体、超导、等离子体、天体力学、材料和偏微分方程，其中前五个方面应用已列为美国研究计划重点“Grand Challenges”.

物理学科发展正是这样，时至今日，几乎无可争辩的是：一切真实的耗散效应可以略去不计的物理过程——不论它是经典的，还是量子性的或是相对性的——都能表达为这样那样的哈密顿形式，不管它有有限多个还是无限多个自由度.

有限自由度：天体与人造星体力学、刚体力学与多刚体力学——包括机器人运动、几何光学和几何渐近方法，包括波动方程射线近似方法与量子力学 WKB 方程、等离子体约束、高能加速器的设计、自动控制，等等.

无限自由度：理想流体力学、弹性力学、电动力学、量子力学与量子场论、广义相对论、孤子与非线性波，等等.

以上说明哈密顿体系是遍在的，普适的，且它具有能将不同物理规律纳为统一数学形式的优点.因此有理由认为，对于哈密顿体系进行计算方法的系统性研究，如果能取得成功，则将会有极其宽广的应用.

现在看一看有关哈密顿方程计算方法的现状.哈密顿体系，包括有限或无限维的都是特定形式的常微分方程或偏微分方程，对于微分方程计算方法的研究从 18 世纪起，至今已有异常丰富的积

累，专著、论文卷帙浩繁.无论是通用的、普适的方法，还是针对特定类型的方法都是这样.但是我们发现针对哈密顿类型方程的计算方法都基本阙如.这一空白贫乏的现状与哈密顿体系的重要性和普适性形成了尖锐的对比，是令人费解的.因此对这片未开垦的处女地进行探索与开辟是值得进行的，是很有吸引力的.

3.4 技术途径——辛几何方法

哈密顿体系基础是辛几何，辛几何理论与应用正在日益发展.

辛几何历史可追溯到19世纪英国天文学家哈密顿.他为了研究牛顿力学，引进广义坐标和广义动量来表示系统的能量，现在通称为哈密顿函数.对于自由度为 n 的系统，n 个广义坐标和 n 个广义动量，张成 $2n$ 维相空间.于是，牛顿力学就成为相空间中的几何学，用现代观点来看，这是一种辛几何学(symplectic geometry).随后雅可比、达布、庞加莱、Cartan、Weyl，从不同角度(代数与几何的)对它进行了研究.但是，现代辛几何的兴起，应该说是从KAM(Kolmogorov，Arnold，Moser)定理的建立(50年代中到60年代初)开始的.在70年代，由于研究Fourier积分算子，几何量子化与群表示论，临界点分类、李代数对偶空间上的哈密顿系统的需要，人们对辛几何作了大量的研究(Arnold，Duistermatt，Guillemin，Weinstein，Marsden)从而推动了这些研究领域的发展.进入80年代后，整体辛几何的研究相继出现，如硬辛几何的研究(Gromov)、辛映射不动点的研究(Arnold猜测)(Conley－Zehnder)、矩映射凸性的研究(Atiyah，Gullemin－Sternberg).看来，不仅辛几何本身的研究是极其丰富而有生命力的，而且它的应用领域极其广

泛，如天体力学、几何光学、等离子体物理、高能加速器的设计、流体力学、弹性力学、最优控制等.

当代计算方法研究的一条不成文的基本法则是，问题原形的基本特征在离散后应该尽可能地得到保持.而为了达到这一效果则离散化应尽可能在问题原型的同一形式框架中进行.例如有限元法正是把离散纳入原体型的解所组成的 Sobolev 函数空间的同一框架中进行，使得对称性、正定性、守恒性等基本特征得到保持，从而从根本上保证了实用上的有效性与可靠性，同时使得理论建立也变得相当容易，还有一些不那么成功的计算方法正好提供了反例.

根据上述指导思想，为了要构造哈密顿体系的算法就必须提出能保持该体系基本特征的算法，不妨称之为哈密顿算法，而为了搞出哈密顿算法就应该在哈密顿体系的同一框架中进行.下面即将说明辛几何就是哈密顿体系的数学框架，所以哈密顿算法应该从辛几何框架内产生，这就是我们研究中的技术途径.

我们将以欧氏几何作为类比，来简单说明辛几何.欧氏空间 R_n 的欧几里得结构取决于双线性对称的、非退化内积：

$$\langle x, y\rangle = \langle x, Iy\rangle, I = I_n \text{ 单位阵},$$

由于非退化，当 $x \neq 0$ 时 $\langle x, y\rangle$ 恒正，从而可以定长度 $\|x\| = \sqrt{\langle x, x\rangle} > 0$.保持内积即长度不变即满足 $A'A = I$ 的线性算子 A 组成一个群 $O(n)$，即正交群，这是一个典型 Lie 群，它的 Lie 代数 $o(n)$ 由满足 $A' + A = AI + IA = 0$ 的条件即反对称变换组成，也就是无穷小正交变换所组成.

辛几何则是相空间中的几何学，辛空间即相空间具有特定的辛结构，取决于一个双线性、反对称的非退化的内积——辛内积

$$[x,y]=(x,Jy),\quad J=J_{2n}=\begin{pmatrix}0 & +I_n\\ -I_n & 0\end{pmatrix}$$

当 $n=1$ 时，

$$[x,y]=\begin{vmatrix}x_1 & y_1\\ x_2 & y_2\end{vmatrix}.$$

这就是以向量 x,y 为边的平行四边形面积，一般地辛内积是面积度量，由于内积的反对称性，对于任意向量 x 恒有 $[x,x]=0$，因此不能由辛结构导出长度的概念，这是辛几何与欧氏几何根本的差别．保持辛内积不变的线性变换满足 $AJA=J$，它们组成一个群 $Sp(2n)$，叫做辛群，也是一个典型的李群．它的李代数则由无穷小辛变换 B 即满足 $BJ+JB=0$ 组成，用 $sp(2n)$ 来表示．由于奇数维中不存在非退化的反对称阵．因此辛空间必定是偶数维的，相空间正是如此．概括说来欧氏几何是研究长度的几何学，而辛几何是研究面积的几何学．

辛几何中一对一的非线性变换称为辛变换也叫正则变换，如果它的雅可比阵处处是辛阵．这类变换自然在辛几何中起主导作用．回到哈密顿动力体系，如果把一对 n 维向量表示为一个 $2n$ 维向量 $z=(p,q)$，则哈密顿正则方程就成为

$$\frac{dz}{dt}=J^{-1}\frac{\partial H}{\partial z}.$$

在辛变换下，哈密顿方程的正则形式不变，哈密顿动力学的基本定理说：对于任意哈密顿体系必定存在依赖 H 及时刻 t_0,t_1 的辛变换族(即相流) $G_H^{t_1,t_0}$．使得

$$z(t_1)=G_H^{t_1,t_0}z(t_0).$$

即 $G_H^{t_1,t_0}$ 把 t_0 时刻状态变为 t_1 时刻的状态．因此哈密顿动力体系

的演化永远是辛变换的演化，这是经典力学体系的普适基本数学特征。当 H 不依赖 t 时，$G_H^{t_1-t_0}=G_H^{t_2-t_1}$，即相流只依赖于参数差 t_2-t_0，可命 $g_H^t=G_H^{t-t_0}$。

哈密顿系统的一个重要问题就是稳定性问题，这类问题在几何上的特点是：它的解在相空间上是保测的，其特征方程的根是纯虚数，所以不能用庞加莱、里亚普诺夫渐近稳定性理论，而必须用KAM定理来加以研究，是一种关于整体稳定性的论断，这是牛顿力学发展史上最重大的突破。辛几何在数值分析中的应用是冯康于1984年在北京召开双微会议上首先提出的。它是基于分析力学中的基本定理：系统的解是一个单参数的保测变换(即辛变换)，从而开创了哈密顿力学计算的新方法。而我们研究哈密顿力学计算方法正是从这个观点出发，使离散化后的方程保持原有系统的辛结构，即恢复离散哈密顿力学的本来面貌。它的离散相流可看成一系列离散辛变换，从而保持一系列相面积和相体积守恒。1988年冯康到西欧各国访问，关于辛算法的研究工作得到了许多著名数学家的承认。他在为庆祝法国著名数学家Lions 60岁生日的报告会上所作的题为"辛几何与计算哈密顿力学"的报告得到与会者一致好评。Lions认为这是"冯康继独立于西方在中国发展了有限元后又开创了哈密顿力学的辛算法"；西德著名数值分析家Stoer说"这是一个长期被人们忽视的领域而又不应忽视的新方法"。

现在我们知道研究哈密顿力学离不开辛几何，而哈密顿力学的计算方法离不开辛差分格式。经典的Runge - Kutta方法不适应解此类问题，它不能保持长期稳定性的计算。例如四阶Runge-Kutta方法当用0.1步长计算20万步以后所得结果面目全非，因

为它不是一个辛算法，而是一个耗散的算法.

3.5 辛几何格式

任何一个格式不论是显式和隐式的，它都能看成从上一时刻到下一时刻的映射，如果这个映射是辛的，我们就说差分格式是辛格式.

首先我们可以从经典差分格式中找，大家所熟知的欧拉中点格式是辛格式：

$$z^{n+1}=z^n+J^{-1}H_z\left(\frac{z^{n+1}+z^n}{2}\right).$$

辛格式往往是隐式的，只有对于可分的哈密顿系统，利用显式隐式交替可得到的实质上是显式的辛差分格式.这时精度仅为一阶，把此格式对称化便得到二阶精度对称格式(或叫可逆格式)，在多级 Runge - Kutta 格式序列里也存在着辛的多级 Runge - Kutta 辛格式，如 2s 阶高斯 多级 Runge - Kutta 格式是辛格式.

另外，冯康从分析力学出发，利用生成函数理论构造了种类繁多的任意阶精度的辛算法.在发展算法同时系统地发展了构造性生成函数理论与哈密顿－雅可比方程.即在线性达布变换框架下，构造了所有类型的生成函数与相应哈密顿－雅可比方程.

3.6 无源系统的保体积格式

在众多的动力系统中有一类动力系统称之为无源动力系统，它的特点是向量场散度为零.

$$\frac{dx}{dt} = f(x), \qquad \mathrm{div} f(x) = 0$$

对这种系统它的相流是保体积的，即 $\det(e_f'(x))_x = 1, \forall x, t$. 这里 e_f 表示方程的相流，$(e_f'(x))_x$ 表示 e_f 在点 x 的雅可比矩阵.

为了正确计算无源系统，我们所构造的算法也应保持 $\det\left(\frac{\partial x_{n+1}}{\partial x_n}\right) = 1$ 即方程的数值解也应保持体积.

我们知道哈密顿系统是偶数维的，而无源系统可以是偶数维，也可以是奇数维的. 此时，欧拉中点格式就不一定保体积格式了. ABC(Arnold - Beltrami - Childress)流是方程的一个例子，它的向量场有如下的形式：

$$\dot{x} = A\sin x + C\cos y;$$
$$\dot{y} = B\sin x + A\cos z;$$
$$\dot{z} = C\sin y + B\cos x,$$

这是一个无源系统，它的相流应保体积，这是一个可分系统，所以可以很容易构造显式保体积格式. 数值试验表明保体积格式可以正确算出此流的拓扑结构. 反之传统的非保体积格式不能正确计算出 ABC 流的拓扑结构.

3.7 切触系统的切触格式

在自然界中存在着这样的奇数维动力系统. 它与偶数维哈密顿系统有类似的辛结构，我们称它为切触系统.

考虑 R^{2n-1} 空间中切触系统：

$2n+1$ 维向量：$\begin{pmatrix} x \\ y \\ z \end{pmatrix}$，其中 $x=\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$，$y=\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$，$z=(z)$；

$2n+1$ 维向量场：$\begin{pmatrix} a(x,y,z) \\ b(x,y,z) \\ c(x,y,z) \end{pmatrix}$，其中 $a=\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$，$b=\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$，

$c=c$.

一个切触系统有一个接触哈密顿函数 $K(x,y,z)$ 生成：

$$\frac{\mathrm{d}x}{\mathrm{d}t}=-K_y-K_z x=a;$$

$$\frac{\mathrm{d}y}{\mathrm{d}t}=K_x-b;$$

$$\frac{\mathrm{d}z}{\mathrm{d}t}=K_z=c,$$

$$K_e(x,y,z)=K(x,y,z)-(x,K_y(x,y,z)).$$

R^{2n+1} 空间接触结构定义为一形

$$\alpha=x\mathrm{d}y+\mathrm{d}z=(o,x^T,1)\begin{pmatrix} \mathrm{d}x \\ \mathrm{d}y \\ \mathrm{d}z \end{pmatrix},$$

一个变换 f 叫接触变换，如果它在相差一个因子 μ_f 下保持切触结构不变. 一个差分格式如果保持上面所说的性质，就叫它切触格式.

这些格式有着潜在的应用：①波阵面的传播，②热力学中的应用，③一阶方程组特征线法.

辛算法、保体积算法、切触算法都是以保持动力系统相空间几何结构为特征算法，统称为动力系统几何算法. 几何算法这个术语系由冯康引入，业已得到国际同行们广泛的认可和采用，例如1996年在英国召开数值方法学科进展情况讨论会，其中有关动力

系统数值方法的讨论中就提到保动力系统几何结构的重要性：对动力系统保几何结构算法，发展了一套关于利用组合格式达到高精度保结构，也就称之为乘积外推方法，即它既提高了格式的精度，又保证系统结构不被破坏。我们把 Yoshida 只对显式格式的情况推广到一般自共轭格式的情况，利用格式与它共轭格式的乘积得到自共轭格式，再用自共轭格式外推达到高精度。

3.8 动力系统几何算法的广泛应用

3.8.1 大时间尺度的动力系统

天体力学与动力天文中所涉及到的力学系统，几乎都是哈密顿系统，或带有小耗散的拟哈密顿系统，相应的动力学问题可由哈密顿正则运动方程来描述，而哈密顿力学问题已成为当今动力系统研究领域中的一个极其重要的方面。但动力天文中所涉及的哈密顿正则运动方程又往往是较复杂的非线性方程，无法给出解析解，尽管在某些情况下可归结为小参数方程，给出相应的小参数幂级数解，可这些级数解只能描述有限时刻真实运动的一种近似，不仅不能满足日益增长的高精度要求，而且更无法给出相应力学系统的长期演化性态，特别是一些带有本质性的非线性现象。故在研究中常借助于数值方法，给出相应数值结果，一方面满足某些定量问题的高精度要求，更重要的是在定性研究中提供力学系统的一些全局图像和重要"信息"，以便进一步进行理论研究，甚至就可由这些数值结果得出某些重要结论。在哈密顿系统的定性研究中，通常有两种途径，一种是用数值方法直接求解对应的哈密顿正则运

动方程，另一种是对运动方程施行更简单的离散化过程，将其变为一简单的映射问题，它可使得计算更为简便，在一般计算机上就可研究动力系统在大时间尺度上的演化性态.

关于数值方法，一般有两类，即单步法(Runge - Kutta 方法为代表)和多步法(常用的有解一阶方程组的亚当斯型方法和解二阶方程组的 Cowell 方法).但是这些传统的数值方法有一个共同点，即均有人为的能量耗散，使得相应哈密顿系统的总能量随时间呈线性变化(即计算中能量误差有线性累积)，这将歪曲哈密顿流的整体特征，从而导致对相应系统长期演化性态的研究失败.从定量结果来看，能量耗损的一个直接后果是导致天体数值轨道沿迹(即沿天体运动方向)误差的严重累积.至少是按积分间隔$(t-t_0)^2$的规律快速增长.

冯康等人 80 年代以来建立的哈密顿算法(即辛算法)不仅是一种新的数值方法，而且从理论上清楚地阐明了传统数值方法导致能量耗损的根本原因，即相应的差分格式是非辛的，其截断是耗散项，辛算法对应的差分格式严格保持哈密顿系统的辛结构，这是哈密顿系统的一个极好的整体结构，有限阶辛算法的截断部分不会导致系统能量发生线性变化，而仅对应周期变化.其后果正是人们期望的，特别对动力天文中的定性研究问题，由于该算法能保持系统的辛结构，将不至歪曲哈密顿系统的整体特征，使所得长期演化性态能较真实地反映天文现象；而能量又是系统运动特征的一个重要参数，它的"保持"将使得相应的数值结果更具实际意义，不至于出现一些非系统本身所具有的"计算机现象".事实上，辛算法除在定性问题的长期跟踪计算中发挥传统方法无法比拟的优势外，在动力天文的一些定量问题中也有它相应的特点，由于能量被

"控制"住,天体数值轨道的沿迹误差不再按$(t-t_0)^2$的规律快速增长,而变为仅随$(t-t_0)$线性增长,这对长弧计算是极为有利的.

鉴于辛算法的上述优点,在动力天文研究中,它已逐渐被广泛采用,特别是定性研究问题.如太阳系动力演化,包括小行星的运动稳定区域,空间分布与轨道共振,大行星(特别是外行星)的长期演化等天体力学与动力天文前沿领域的热点课题.

3.8.2 辛算法在定性研究中的应用

首先可用两个简单的例子来体现辛算法在动力天文定性研究中所起的特殊作用.

例 1 开普勒运动,此即二体问题中的椭圆运动,该问题对应的哈密顿函数为

$$H(p,q)=T(p)+V(q),$$

其中p,q为广义动量和广义坐标矢量,T和V分别为动能和势能.这一问题的解是一不变椭圆,但是,分别采用 Runge - Kutta 方法和辛算法进行数值求解时,前者却明显地使椭圆逐渐缩小,最后完全失真,而辛算法始终保持椭圆的大小和形状(图 3－7,图 3－8,其中,内圈缩的椭圆对应 R - K 算法,外圈的"不变"椭圆对应辛算法,e表示偏心率,150s,1000s 表每圈取 150 步和 1000 步).这就表明非辛算法有人为的能量耗散,而辛算法由于保持了开普勒流的辛结构,使其保持了运动的主要特征.因此在天体运动的长期跟踪计算中,辛算法显然具有其特殊地位,这是所有非辛算法不可代替的.

例 2 轴对称星系中的恒星运动问题,其简化的动力模型所对应的哈密顿函数为

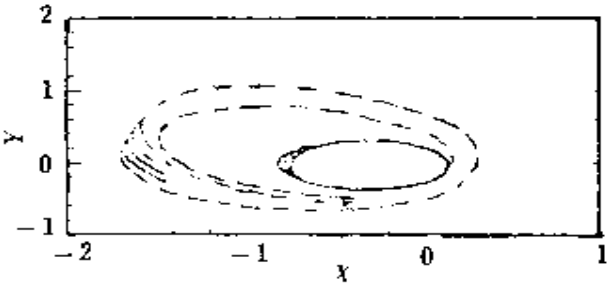


图 3－7　($e=0.7$, 150s) 开普勒运动，R－K 方法和辛算法比较

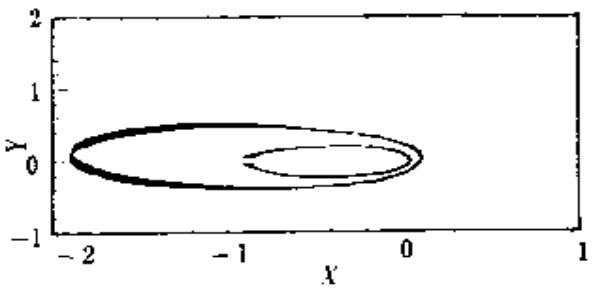


图 3－8　($e=0.9$, 1000s) 开普勒运动，R－K 方法和辛算法比较

$$H(p,q)=\frac{1}{2}(p_1^2+p_2^2)+\frac{1}{2}(q_1^2+q_2^2)+\left(2q_1^2q_2-\frac{2}{3}p_2^3\right).$$

为了了解该系统演化过程中的基本特征，分别采用精度较高的 7,8 阶嵌套的(Runge-Kutta-Fehlberg)(记作 RKF(8))方法和 6 阶显辛算法，计算结果列于图 3－9 至图 3－14，从图 3－9，图 3－11 和图 3－10，图 3－12 可以看出，无论是有序区(LCN＝0)还是

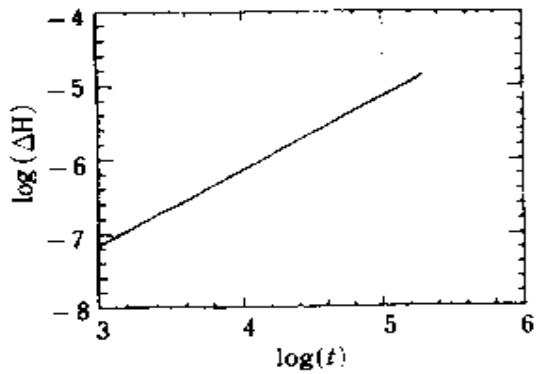


图 3－9　$H_0=0.553$, LCN＝0 时由 RKF7(8)得到的 ΔH 随 t 的变化曲线

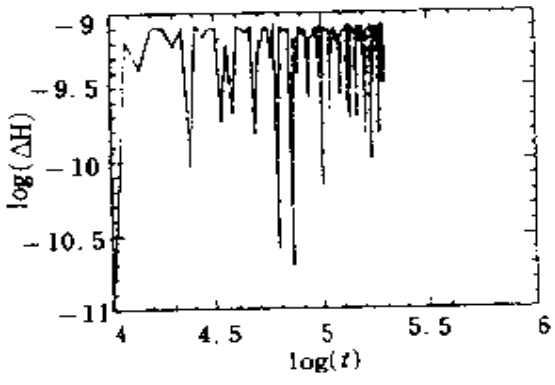


图 3－10　$H_0=0.0553$, LCN＝0 时，由 SY6 得到的 ΔH 随 t 的变化曲线

无序区(LCN＞0)，辛算法的能量变化 ΔH 均无耗散现象，而RKF7(8)却明显地呈现耗散状态，即 ΔH 随 t 线性变化，歪曲了系统的特征；从图 3－13、图 3－14 可更清楚地看出，辛算法在有序区和无序区均可准确地反映该系统的动力学特征，即分别给出不变曲线和混沌性态．

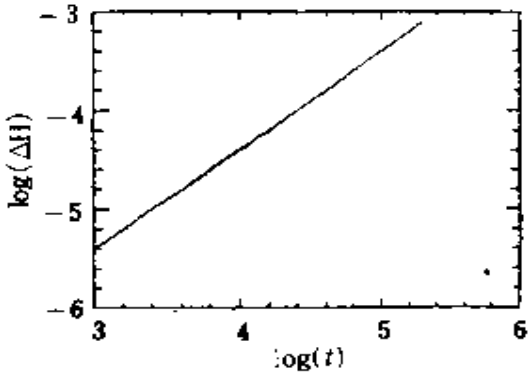


图 3－11　$H_0=0.148$，LCN＞0 时由 RKF7(8)得到的 ΔH 随 t 的变化曲线

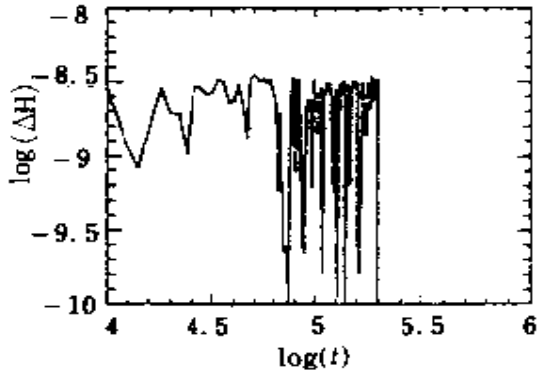


图 3－12　$H_0=0.0148$，LCN＞0 时由 SY6 得到的 ΔH 随 t 的变化曲线

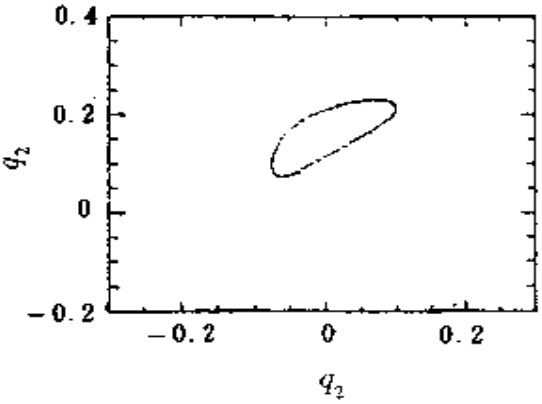


图 3－13　$H_0=0.553$，LCN＝0 时由 SY6 得到的庞加莱截面

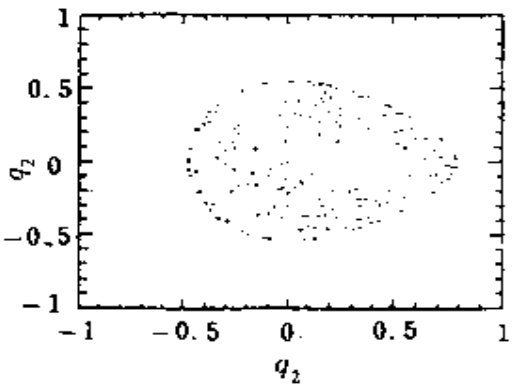


图 3－14　$H_0=0.0148$，LCN＞0 时由 SY6 得到的庞加莱截面

正因为辛算法有保持哈密顿整体结构的特殊作用，从而使计算中系统的演化性态不致失真，这是及其重要的，因此很快被引入到动力天文的研究中.

当前，太阳系动力演化仍是天体力学前沿课题的一大热点，包括大行星(特别是外行星)轨道的长期演化，主带小行星的空间分布特征(Kirkwood 空隙现象)与轨道共振问题，大行星卫星系统的保持，行星环的形成与演化，近地小行星的轨道演化等，所有这些问题，基本上都要借助于数值方法在计算机上进行仿真模拟，但是仿真计算又必须使相应的计算间隔$(t-t_0)$延伸得足够大，如在研究大行星的动力演化中，这一间隔要达到太阳系的年龄10^9年或更长些，在这样长的间隔内要使计算结果能较真实地反映系统本身的运动特征，所有非辛算法都难以满足，而辛算法却能采用低阶和较大积分步长进行计算就可达到要求.

近年来日本、美国等国的天文学家 Kinoshita, Bretit, Wisdom 等人在这方面做了大量工作，特别是 Wisdom 的工作，已在太阳系动力演化问题的研究中得到广泛引用，他针对太阳系的状态，对包含一大质量天体(指太阳)的 N 体系统，采用雅可比坐标系，得到该系统如下形式的哈密顿函数：

$$H(p,q)=\sum_{i=1}^{n-1}H_i(p,q)+\varepsilon\Delta H(p,q),$$

其中$H_i(p,q)$为各对应一个二体问题的哈密顿函数，$\varepsilon\ll 1$是小参数，按这样的分解，可构造各阶显式辛算法，其截断误差比按动能和势能分解对应的哈密顿函数$H(p,q)=T(p)+V(q)$所构造的同阶显式辛算法小ε量级，这样即使采用低阶算法(如二阶算法)亦可成为太阳系动力演化研究中的有效数值方法之一.

自 80 年代末以来，我国天文学家们在动力天文研究中亦成功地引用了辛算法，并作了有益的改进，主要用于以下几个方面：

（1）对于太阳、大行星和小行星构成的限制性三体系统，研究其对应的 1∶1 轨道共振和三角稳动点的动稳定区域，获得了一些新结果，较好地解释了 Trojan 群小行星的分布以及有关几个大行星对应的三角稳动点稳定区域的实际大小.

（2）在研究一些特殊小行星轨道的长期演化中，对辛算法作了类似的改进.在摄动力学系统中，将小行星运动对应的哈密顿系统作了如下分解.

$$H(p,q) = H_0(p,q) + \varepsilon H_1(q),$$

其中 $H_0(p,q)$对应可积系统，是一哈密顿流，$\varepsilon \ll 1$，用算法合成构造了一种改进的显式辛算法，其效果与上述的改进相当，因此用低阶算法对上述小行星轨道的长期演化作计算，其结果中，反映运动实质的能量参数的变化被保持在一定范围内，而无人为的耗散现象，所获结果可靠，有实际意义.

（3）对于星系演化问题，引用了辛算法，在仿真模拟计算中成功地再现了棒(bar)的现象，较好地呈现了星系 NGC4736 中恒星演化的状态.

（4）为了更深刻地刻划天体系统动力演化的特征，对辛积分器作了进一步的探讨，特别是有关形式积分的存在性，各守恒量在演化过程中的状况及其稳定性等问题获得了有益的结果.

除上述对动力天文中哈密顿系统的研究外，对于具有小耗散的情况亦作了探讨与应用.鉴于耗散因素相对微弱的特点，对保守部分(也是相应力学系统的主要部分)和耗散部分，分别采用显式辛差分格式和欧拉中点格式(隐的辛差分格式)构造了一种混合型

的辛算法，其效果亦是显著的，它能保持该系统主要部分哈密顿流的特征.

3.8.3 辛算法及其在定量计算中的应用

由于辛算法能保持系统的辛结构，能量误差无线性累积，因此在积分天体运动方程时，轨道沿迹误差将仅随$(t-t_0)$线性增长，而不像非辛算法那样按$(t-t_0)^2$的规律快速增长.为了表明这一点，下面给出一些算例，读者可清楚地了解这一特征.

以 Lageos 卫星的轨道为背景，只分别考虑地球非球形摄动和同时考虑大气阻力摄动(相对非球形摄动是小量)两种力学系统，前一种受摄运动问题对应哈密顿系统，而后一种受摄运动则对应一小耗散系统，即拟哈密顿系统.对这两种系统分别采用 RKF7(8)和改进的 6 阶辛算法(记作 SY6)，计算 1000 圈的长弧段，将计算结果中最主要的沿迹误差$\Delta(M+\omega)$列于表 3-1、表 3-2.从表中(RKF7(8)和 SY6)的结果可清楚地看出，非辛算法由于存在能量耗散确实使得沿迹误差按$(t-t_0)^2$的规律快速增长，尽管开始精度较高，往下计算效果很快变差；而辛算法则使沿迹误差随$(t-t_0)$线性增长，对于弧段较长的计算，效果明显.从表 3-2 的数据可以了解到，即使将辛算法推广到小耗散系统，亦有同样的

表 3-1　仅考虑非球形摄动的沿迹误差$\Delta(M+\omega)$

方　法	积分步数/圈	100 圈	1000 圈	10000 圈
RKF7(8)	100	$1.5E-10$	$1.4E-08$	$1.3E-06$
SY6	50	$0.5E-09$	$0.6E-08$	$1.0E-07$
RKH	100	$0.9E-11$	$0.9E-10$	$0.9E-09$

表 3－2　同时考虑大气阻力摄动的沿迹误差 $\Delta(M+\omega)$

方　法	积分步数/圈	100 圈	1000 圈	10000 圈
RKF7(8)	100	1.4E－10	1.3E－08	1.3E－06
SY6	50	0.6E－09	0.7E－08	1.0E－07
RKH	100	2.1E－11	3.5E－10	6.2E－09

效果.因此,辛算法在定量问题中,如果计算弧段稍长,它也有一定的应用价值.

这里说明一点,由于辛差分格式受辛条件的限制,本身的计算精度(局部截断误差的反映)与同阶非辛算法相比并不高,因此上面采用的 SY6 是改进的辛格式,即前面 3.8.2 节中提出的将哈密顿函数按另一种分解所构造的一种改进的辛算法.

鉴于辛算法的原理及其本身的定量计算精度并不比同阶非辛算法高,又考虑到定量问题主要关心的是高精度问题,特别是较长弧段的计算,只要能避免沿迹误差按$(t-t_0)^2$快速增长即行,至于是否能保持系统的整体特征并不重要,因此,只要根据辛算法的原理避免能量耗散(相应误差并不要线性累积)即可,这样可控制积分过程中天体轨道沿迹误差的快速增长.由此,完善了一种在非辛算法中采用能量控制的技术,即利用能量关系式,在积分过程中,每积分一步,对能量进行一次调整,以此补偿由于局部截断部分引起的能量耗散,这不仅可以同样达到控制由于局部截断误差导致的沿迹误差的快速增长的目的,而且还可以同时补偿(至少一部分)舍入误差,使效果更加明显.为了表明这一点,仍然计算上面提出的同一问题,对 RFK7(8)采用能量补偿记作 RKH,计算结果中的沿迹误差 $\Delta(M+\omega)$亦列入表 3－1 和表 3－2 中.由结果可看出,效果相当明显,不仅控制了沿迹误差的快速增长,使其与辛算

法的结果相同，亦仅随$(t-t_0)$线性增长，而且本身的计算精度也较高，保持了原相应的非辛算法自身的优点.这种以辛算法原理为依据的能量补偿法，虽然不能解决定性研究中保持系统长期演化应有的特征，但却能在定量精度上达到更好的效果.从解决人为的能量耗散这一角度来看，上述方法所获效果也可以说是辛算法以另一种形式出现所起的作用.

3.8.4 量子系统的计算

量子系统的时间演化由时间相关 Schrödinger 方程

$$i\frac{\partial \Psi}{\partial t}=\hat{H}\Psi,\quad \hat{H}=\hat{H}_0(\boldsymbol{r})+\hat{V}(t,\boldsymbol{r}) \tag{3.1}$$

描述，其中哈密顿算符$\hat{H}$是 Hermite 的.

依据量子力学基本原理，量子系统在一个时刻的状态确定了这个时刻之后的状态，详言之，当t_1时刻的状态$\Psi(t_1,\boldsymbol{r})$给定后，方程(3.1)的解(称为波函数)

$$\Psi(t,\boldsymbol{r})=a(t,\boldsymbol{r})+ib(t,\boldsymbol{r})$$

由一组时间演化算符$\{U_{\hat{H}}^{t_1,t_2}\}$生成：$\Psi(t_2,\boldsymbol{r})=U_{\hat{H}}^{t_1,t_2}\Psi(t_1,\boldsymbol{r})$，每个时间演化算符$U_{\hat{H}}^{t_1,t_2}$是西的并且仅仅依赖于$t_1,t_2$和$\hat{H}$而与初始时刻$t_1$的状态$\Psi(t_1,\boldsymbol{r})$无关.在这种意义下，人们说量子系统的时间演化是西变换的演化.由此，诱导出一组算符$\{S_{\hat{H}}^{t_1,t_2}\}$，其中的每一个算符$S_{\hat{H}}^{t_1,t_2}$对应于$\{U_{\hat{H}}^{t_1,t_2}\}$中的一个算符$\{U_{\hat{H}}^{t_1,t_2}\}$且作用在波函数的虚部和实部作为分量的实函数向量上：

$$\begin{bmatrix}b(t_2,\boldsymbol{r})\\ a(t_2,\boldsymbol{r})\end{bmatrix}=S_{\hat{H}}^{t_1,t_2}\begin{bmatrix}b(t_2,\boldsymbol{r})\\ a(t_2,\boldsymbol{r})\end{bmatrix}.$$

假设 $\phi(t,r)=c(t,r)+id(t,r)$ 是初态为 $\phi(t_1,r)$ 的方程(3.1)的解，于是

$$\phi(t_2,r)=U_H^{t_1,t_2}\phi(t_1,r).$$

算符 $S_H^{t_1,t_2}$ 保持任二实向量 $\begin{bmatrix}b\\a\end{bmatrix}$ 与 $\begin{bmatrix}d\\c\end{bmatrix}$ 的内积和辛积不变；因为任二实向量的内积守恒等价于任一实向量的模方守恒，因此在上述意义下，我们说量子系统的时间演化是模方守恒—辛变换的演化，简称为模方守恒－辛的(norm－preserving symplectic 或NPS)．所以，量子系统是一个(无穷维的)哈密顿系统，时间相关Schrödinger 方程可转化成一个分别以波函数的虚部和实部为广义动量和广义坐标的正则方程，并且波函数模方是它的守恒量(不变积分)；采用模方守恒－辛格式求解这个正则方程是数值研究量子系统时间演化的合理途径．

这样的偏微分方程是无限维的，为了数值求解，应将正则方程离散化成有限维正则方程，并且相应地保持模方守恒．

设时间相关 Schrödinger 方程的算符 $\hat{H}_0(r)$ 在给定边界条件下的本征函数包括分立态和连续态，将方程在规定初态下的演化态(波函数)展开并代入方程，再取充分大的正整数 M、N 和正实数 R 作截断，令 $\Delta=\frac{R}{M}$，将积分离散化便得到微分方程组．

特别地，当哈密顿算符 $\hat{H}=\hat{H}_0(r)+\hat{V}(r)$ 不含时间时，量子系统的能量 $\langle\Psi|\hat{H}_1|\Psi\rangle=Z^THZ$ 是正则方程的守恒量(不变积分)，也是模方守恒－辛格式的守恒量；可用基于 e^{λ} 的一阶对角Padé 逼近的格式来计算，它是四阶模方守恒－辛的．

强场中的量子系统可能激发到高激发态，本征函数算法恰好截掉了高激发态，故一般不适用于强场．下面，我们用一个例子介绍以对称差商代替空间变量偏导数将时间相关 Schrödinger 方程离散化成正则方程的方法．

考虑一维有限宽无限深势阱中的电子在强场 $V(t,x)$ 作用下的时间演化

$$i\frac{\partial\Psi}{\partial t}=\hat{H}\Psi,\qquad \hat{H}=\hat{H}_0+V(t,x),$$

$$\hat{H}_0=-\frac{1}{2}\frac{\partial^2}{\partial x^2}+V_0(x),$$

$$V_0(x)=\begin{cases}0 & 0<x<1;\\ +\infty & x\leqslant 0 \text{ 或 } x\geqslant 1.\end{cases}$$

与本征函数展开法不同，我们将 Schrödinger 方程离散化时未作任何截断，离散成的正则方程的解包容了 $\hat{H}_0$ 的所有本征态．所以正则方程是强场模型的恰当哈密顿形式，采用模方守恒－辛格式求解正则方程是数值研究强场模型的合理途径．

显式辛格式的守恒量随步长的减少而趋于系统的对应的守恒量．依据显式辛格式计算的能量和波函数模方虽不严格守恒，但随步长的减小而趋于系统的能量和波函数模方的精确守恒值．

一维有限宽无限深势阱中的电子在模拟势 $V(x)$ 作用下时间演化的 TDSE 为

$$i\frac{\partial\Psi}{\partial t}=\hat{H}\Psi,\quad \hat{H}=\hat{H}_0+\varepsilon V(x),\quad \hat{H}_0=-\frac{1}{2}\frac{\partial^2}{\partial x^2}+V_0(x),$$

$$V_0(x)=\begin{cases}0 & 0<x<1,\\ +\infty & x\leqslant 0 \text{ 或 } x\geqslant 1,\end{cases}$$

$$V(x)=\begin{cases}2x & 0<x<0.5,\\ 2(1-x) & 0.5<x<1,\\ 0 & x\leqslant 0 \text{ 或 } x\geqslant 1.\end{cases}$$

按照前面的方法将波函数 $\Psi(t,x)$ 按 $\hat{H}_0$ 的本征态 $\{X_n(x)=\sqrt{2}\sin n\pi x, n=1,2,\cdots\}$ 展开，因哈密顿算符是实算符，TDSE 离散化成可分线性哈密顿系统的正则方程，其中

$$S=(S_{mn}),\qquad S_{mn}=\frac{n^2\pi^2}{2}\delta_{mn}+\varepsilon v_{mn},$$

$$v_{mn}=\begin{cases}\dfrac{1}{2}+\dfrac{1-(-1)^n}{n^2\pi^2}, & m=n;\\ 0, & |m-n|=1,3,5,\cdots;\\ \dfrac{-16mn\left(1-(-1)^{\frac{m-n}{2}}\right)}{(m^2-n^2)^2\pi^2}, & |m-n|=2,4,6,\cdots,\\ & n=2,4,6\cdots,\\ \dfrac{-8\{2mn-(-1)^{\frac{m-n}{2}}(m^2+n^2)\}}{(m^2-n^2)\pi^2}, & \\ & |m-n|=2,4,6,\cdots,n=1,3,5\cdots,\end{cases}$$

取初态

$$\Psi(0,x)=\frac{1+i}{2}\{X_1(x)+X_2(x)\},\quad \varepsilon=5\pi^2,$$

因为哈密顿算符不显含时间，系统的能量守恒，$E(b,a)=e^0=42.0110165$，波函数模方保持规一，$N(b,a)=n^0=1$．采用欧拉中点格式、二阶显式辛格式和 R－K 法取步长 $h=10^{-3}$ 数值求解正则方程，结果表明：

（1）R－K 法不能保持能量守恒和波函数规一，见图 3－15 和图 3－16 中的 E_{R-K} 和 N_{R-K}．

(2) 欧拉中点格式理论上严格保持能量守恒和波函数模方规一，计算结果见图 3 - 15 和图 3 - 16 中的 E_E 和 N_E，但图 3 - 15 中的 E_E 显式，能量有微小的阶梯状变化；这是因为格式是隐式的，每前进一步都要自洽迭代多次，必然产生误差，经多步积累显现出了阶梯变化.

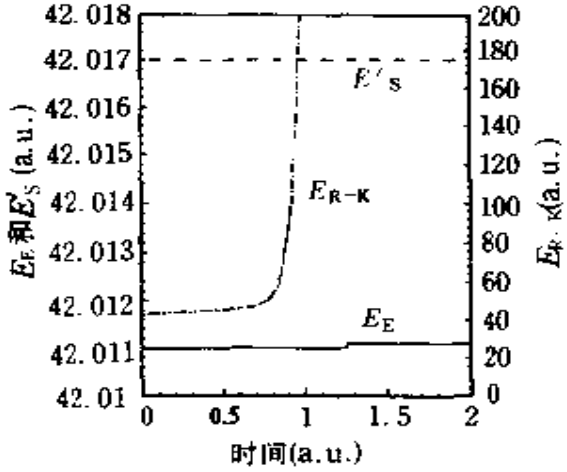


图 3 - 15　三种算法能量图的比较

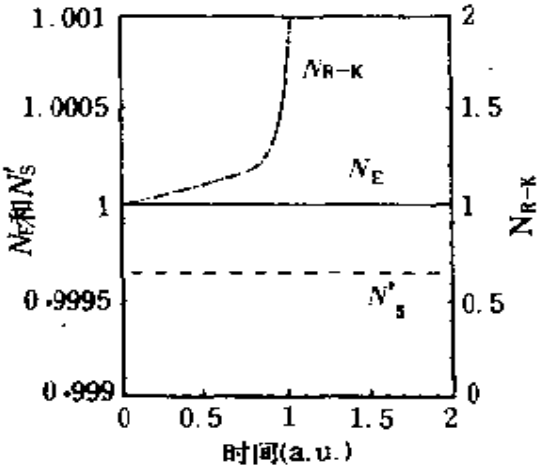


图 3 - 16　三种算法模的比较

(3) 显式辛格式的守恒量 $\widetilde{E}(b^k a^k; h)$ 和 $\widetilde{N}(b^k a^k; h)$ 严格守恒，见图 3 - 15 和图 3 - 16 中的 E'_S 和 N'_S，并且随步长的减小而趋于系统的能量 e^0 和波函数模方 $n^0 = 1$，见表 3 - 3；显式辛格式算得的能量 $E(b^k a^k)$ 和波函数模方 $N(b^k a^k)$ 是振荡的，见图 3 - 17 中的 E_S 和 N_S，并且与 e^0 和 n^0 之差的最大值

$$C_E(h) = \max_k |E_S - e^0|, C_N(h) = \max_k |N_S - n^0|$$

以及振荡的振幅都随步长的减小而趋于 0，见表 3 - 3.

$$e(h) \to e^0, C_E(h) = \max |E(b^k a^k) - e^0| \to 0, \quad n(h) \to n^0,$$

$$C_N(h) = \max |N(b^k a^k) - n^0 \to 0$$

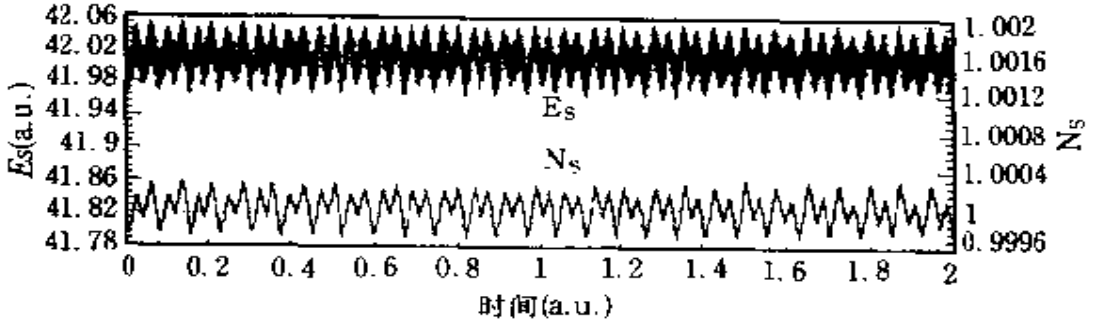


图 3－17　Y算法能量和模图

表 3－3　　步长的变化

h	$e(h)$	$C_S(h)$	$n(h)$	$C_N(h)$
10^{-3}	42.0169964	0.0445060	0.9996509	0.0003106
10^{-4}	42.0110763	0.0004195	0.9999965	0.0000030
10^{-5}	42.0110171	0.0000018	0.999999	0.0000000
10^{-6}	42.0110165	0.0000000	1.0000000	0.0000000
10^{-7}	42.0110165	0.0000000	1.0000000	0.0000000
精确值	42.0110165	0.0000000	1.0000000	0.0000000

由推论和这个例子可见，对不含时间实哈密顿算符的量子系统，在给定精度下，只要步长取得适当小，显式辛格式保持能量和波函数模方守恒．采用显式辛格式有可能克服采用传统数值方法直接求解 TDSE 以研究量子系统时间演化（如强定常场与原子相互作用时）因不能保持波函数模方规一而出现的困难．

设量子系统的实哈密顿算符 $\tilde{H}(\boldsymbol{r};t)$ 显含时间，离散化成的正则方程形式上是一个哈密顿显含时间的 m 维"可分线性哈密顿系统"的正则方程，系统的能量不再是守恒量，但波函数模方仍然

离散化成一个可分二次型守恒量，一维有限宽无限深势阱中的电子在模拟激光场 $V(t,x)=\varepsilon x\sin\omega t$ 作用下的时间演化由 TDSE

$$i\frac{\partial\Psi}{\partial t}=\hat{H}\Psi,\qquad \hat{H}=\hat{H}_0+\varepsilon V(t,x),$$

$$\hat{H}_0=-\frac{1}{2}\frac{\partial^2}{\partial x^2}+V_0(x)$$

描述，按照上面的方法将波函数 $\Psi(t,x)$ 按算符 $\hat{H}_0$ 的本征态 $\{X_n(x)=\sqrt{2}\sin n\pi x, n=1,2,\cdots\}$ 展开，因哈密顿算符是显含时间实算符，TDSE 离散化成显含时间的“可分线性哈密顿系统”的正则方程，其中

$$S(t)=(s(t)_{mn}),\quad s(t)_{mn}=\frac{n^2\pi^2}{2}\delta_{mn}+\varepsilon v(t)_{mn};$$

$$v(t)_{mn}=\begin{cases}\sin\omega t, & m=n;\\ 0, & |m-n|=2,4,6,\cdots;\\ \dfrac{8mn\sin\omega t}{(m^2-n^2)^2\pi^2}, & |m-n|=1,3,5,\cdots.\end{cases}$$

取初态 $\Psi(0,x)=X_1(x)=\sqrt{2}\sin\pi$，因为哈密顿算符显含时间，系统的能量不守恒，但波函数模方保持规一，$N(b,a)=n^0=1$。采用“欧拉中点”格式、一阶显式辛格式和 R-K 法取步长 $h=4\times10^{-3}$ 数值求解正则方程，结果表明：

（1）R-K 法使波函数模方迅速变大，见图 3-18 中的 N_{R-K}；结果不合理，见图 3-19。

（2）“欧拉中点”格式保持波函数模方规一，见图 3-18 中的 N_E，计算结果与理论分析一致；场强 ε 较弱时，如 $\varepsilon=\frac{\pi^2}{2}$，结果与微扰论相同：

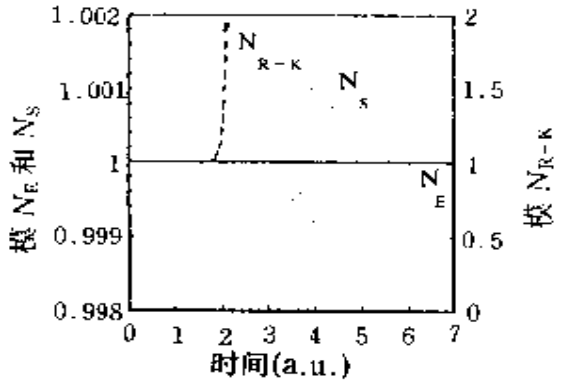


图 3－18　模图

$(\omega=3\pi^2/2,\varepsilon=\pi/2)$

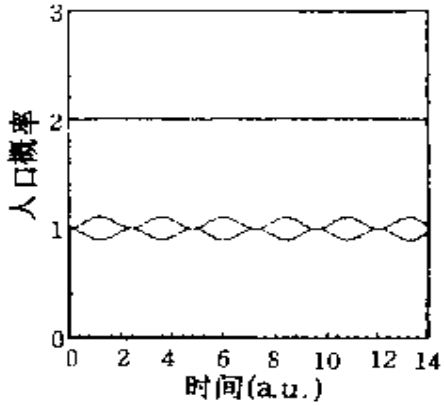
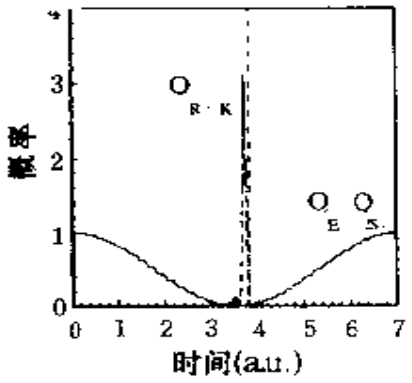


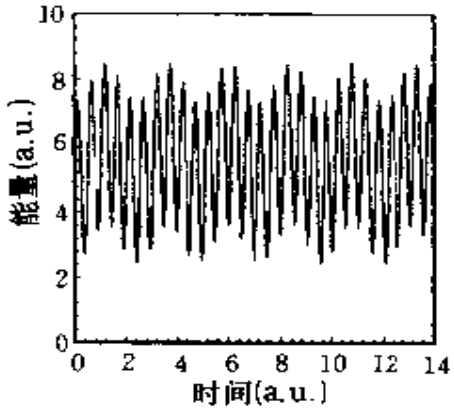
图 3－19　概率图

$(\omega=3\pi^2/2,\varepsilon=\pi^2/2)$

① $\omega\neq\Delta E_{1n}$ 即不共振时，如 $\omega=\dfrac{5\pi^2}{4}<\Delta E_{12}$，不发生态的混杂，见图 3－20(a)和图 3－20(b)；



(a)　概率图$(\varepsilon=\pi^2/2,\omega=5\pi^2/4)$

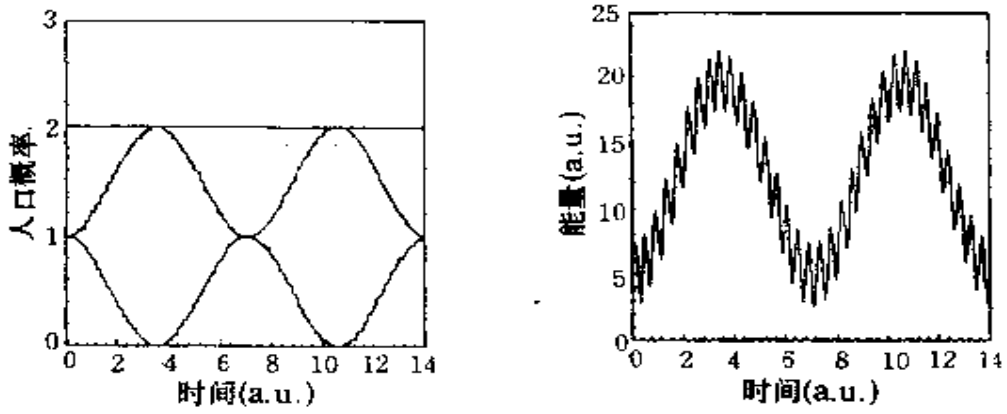


(b)　能量图$(\varepsilon=\pi^2/2,\omega=5\pi^2/4)$

图 3　20　场强较弱且不共振时的概率能量图

② $\omega\approx\Delta E_{1n}$ 即共振时，如 $\omega=\dfrac{5\pi^2}{2}<\Delta E_{12}$，基态与第一激发态混杂，总能量变化周期与态混杂的周期一致，第一激发态几率最

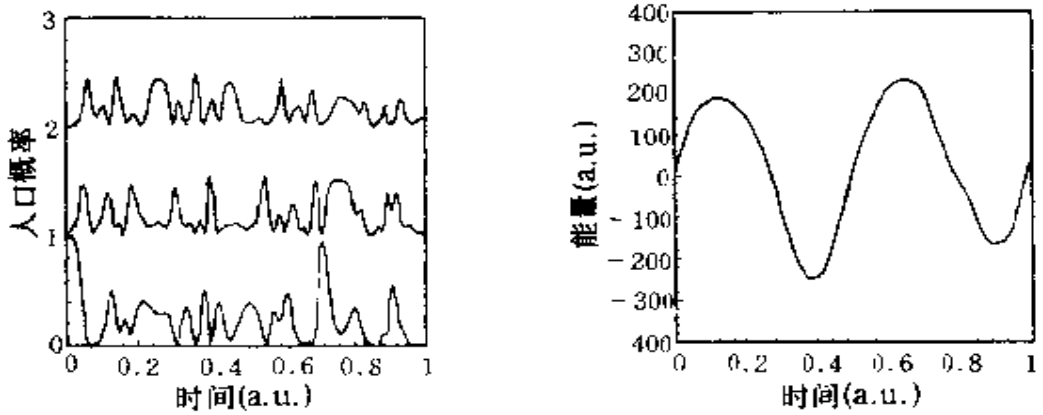
大、最小时总能量也最大、最小，见图 3－21(a)和图 3－21(b).



(a) 概率图($\varepsilon=\pi^2/2,\omega=3\pi^2/2$)　　(b) 能量图($\varepsilon=\pi^2/2,\omega=3\pi^2/2$)

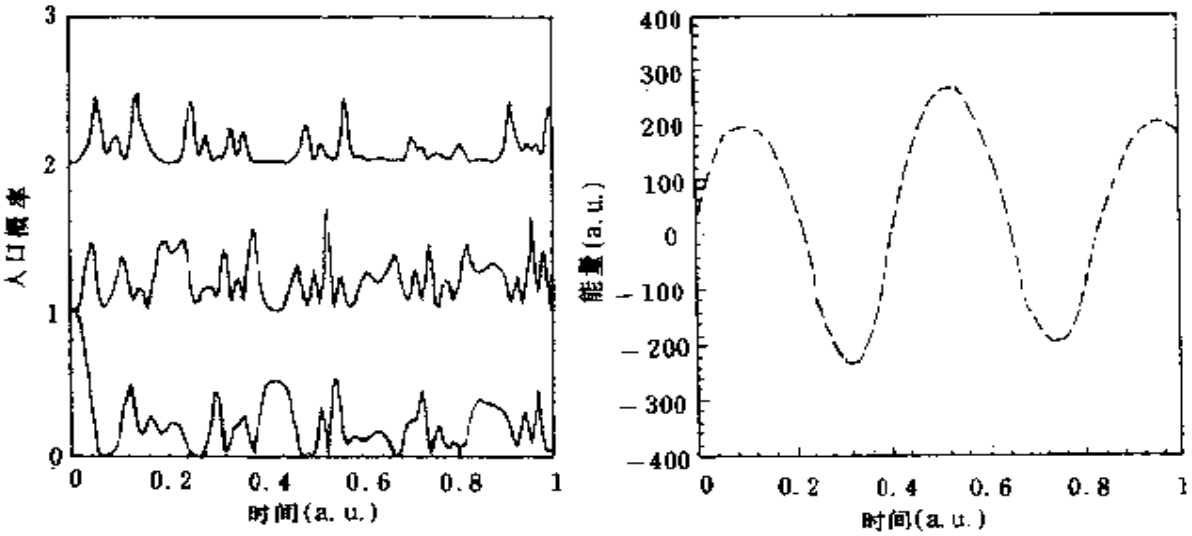
图 3－21　场强较弱且共振时的概率能量图

场强 ε 较强时，选择定则不成立，如 $\omega=\dfrac{5\pi^2}{2}<\Delta E_{12}$，不共振，但基态也与第 1,2,3,…激发态混杂，见图 3－22(a)和图 3－22(b)；又如 $\omega=\dfrac{3\pi^2}{2}=\Delta E_{12}$，共振，基态也与第 1,2,3,…激发态混杂，见图 3－23(a)和图 3－23(b).



(a) 概率图($\varepsilon=50\pi^2,\omega=5\pi^2/4$)　　(b) 能量图($\varepsilon=50\pi^2,\omega=5\pi^2/4$)

图 3－22　场强较强且不共振时的概率能量图



(a) 概率图($\varepsilon=50\pi^2,\omega=3\pi^2/2$) (b) 能量图($\varepsilon=50\pi^2,\omega=3\pi^2/2$)

图 3－23 场强较强且共振时的概率能量图

(3) 2 阶显式辛格式不严格保持波函数模方规一，仅使波函数模方在 1 附近振荡，见图 3－18 中 Q 的 N_2；态的混杂、总能量的变化与(2)中"欧拉中点"格式的结果相同.

由这个例子可见，对显含时间实哈密顿算符量子系统，R－K 法不能保持波函数模方规一，计算结果不合理.我们构造的"欧拉中点"格式保持波函数模方规一，计算结果与理论分析一致.二阶显式辛格式仅使波函数模方在 1 附近振荡，态的混杂、总能量的变化与(2)中"欧拉中点"格式的结果符合得很好.所以，我们构造的"欧拉中点"格式和二阶显式辛格式是求解含时间强场中量子系统时间演化，如强激光场与原子相互作用的合理算法，有可能克服传统算法，如 R－K 法、不能保持波函数模方规一而出现的困难.

3.8.5 A_2B 模型分子反应系统经典轨迹的计算

经典与半经典轨迹是研究微观化学反应动力学的有效理论方

法.经典轨迹方法将原子近似为质点,将反应系统近似为质点系,将反应过程近似为质点系在电子势能面上的经典运动.通过计算大量的经典轨迹计算反应系统的动力学参量,达到研究微观化学反应的目的.Bunker 等最早采用 R - K 法计算了反应系统的经典轨迹,Karplus 等曾对多种数值方法进行了大量的试算,筛选出 R - K - G(Runge-Kutta-Gear)法,并将经典轨迹的计算从 10^{-15}s 推进到 10^{-12}s.虽然 R-K-G 法大大推进了微观化学反应动力学的理论研究,已成为计算经典轨迹普遍采用的方法,但距微观化学反应动力学研究,特别是包含长寿命中间体时所需考虑的时间(10^{-8}s)仍相差几个数量级.已有的某些计算结果表明,采用这些数值方法算得的微观化学反应动力学参量的定性趋势虽然是对的,但数值上常与实验值存在差异,有时甚至存在数量级上的差异.经典轨迹方法将微观反应系统近似为经典哈密顿系统,具有辛结构.因此,采用辛算法代替传统数值方法计算经典轨迹有可能克服目前经典轨迹计算中存在的困难,从根本上改进经典轨迹方法.

在这里,我们取质子质量为单位质量,1 时间单位 $= 4.45 \times 10^{-14}$s.

考虑 A_2B 型分子反应系统在电子势能面上保持 C_{2v} 对称性的经典运动,H_2O、SO_2 等都是这类分子.设 A 与 B 原子的质量分别为 $m_A = 1$ 与 $m_B = 2$,在 A_2B 分子平面上取质心为原点 O、以 C_2 轴为 z 轴取定平面直角坐标系 yOz,记两个 A 原子和 B 原子的坐标分别为 (y_1, z_1),(y_2, z_2) 和 (y_3, z_3).采用 Banerjee 等的有效坐标分离法可求得 A_2B 分子反应系统的广义坐标

$$q_1 = z_1 + z_2 - 2z_3, \qquad q_2 = y_2 - y_1$$

和广义质量 $M_1 = 0.25$,$M_2 = 0.5$,从而可得广义动量

$$p_1 = 0.25\frac{\mathrm{d}q_1}{\mathrm{d}t} \text{和} p_2 = 0.5\frac{\mathrm{d}q_2}{\mathrm{d}t}$$

以及 A_2B 分子反应系统的动能

$$K(p) = 2p_1^2 + p_2^2,$$

电子势能函数采用 Bancrjee 等采用过的，引入 C_{2v} 对称性并记 $D = \sqrt{q_1^2 + q_2^2}$，有

$$V(q) = 5\pi^2(D^2 - 5D + 6.5) + 4D^{-1} + 0.5\pi^2(|q_2| - 1.5)^2 + |q_2|^{-1},$$

这样 A_2B 分子反应系统的哈密顿函数为

$$H(p,q) = K(p) + V(q).$$

经典运动的正则方程

$$\frac{\mathrm{d}p_1}{\mathrm{d}t} = -\frac{\partial V}{\partial q_1} = -f_1(q),\quad \frac{\mathrm{d}q_1}{\mathrm{d}t} = \frac{\partial K}{\partial p_1} = g_1(p),$$

$$\frac{\mathrm{d}p_2}{\mathrm{d}t} = -\frac{\partial V}{\partial q_2} = -f_2(q),\quad \frac{\mathrm{d}q_2}{\mathrm{d}t} = \frac{\partial K}{\partial p_2} = g_2(p),$$

由哈密顿函数 $H(p,q)$ 可见 A_2B 分子系统是一个可分哈密顿系统，可采用显式辛格式求解正则方程组的初值问题而求得数值解

$$t^k = Kh,\quad q_1^k = q_1(t^k),\quad q_2^k = q_2(t^k),$$

$$p_1^k = p_1(t^k),\quad p_2^k = p_2(t^k),$$

继而依据

$$y_3 = 0,\quad z_3 = -\frac{q_1}{4};\quad y_2 = -y_1 = \frac{q_2}{2},\quad z_2 = z_1 = \frac{q_1}{4}$$

求得 A_2B 分子反应系统的经典轨迹和动能、势能、总能量随时间的变化.

我们取初始条件

$$q_1(0) = 3,\quad q_2(0) = \frac{3}{2};\quad p_1 = 0,\quad p_2 = 0,$$

采用4阶显式辛格式和4阶R－K法取步长 $h=0.01$ 求解初值问题，计算 A_2B 分子反应系统在分子平面 yOz 上的经典轨迹以及动能、势能和总能量．图3－24是电子势能函数在相平面上的势能曲面，当 $|q_1|\to+\infty$ 时，$V(q)\to+\infty$；当 $|q_2|\to 0$ 或 $q_2\to+\infty$ 时 $V(q)\to+\infty$．因此从理论上可知 A_2B 分子系统的总能量守恒，B 原子和两个 A 原子作准周期振动，周期性构型反转永无休止．图3－25是总能量随时间的变化图，辛算法至 6.23×10^{9}s 仍守恒；R－K法使总能量迅速减小．图3－26(a)(c)(e)和(b)(d)(f)分别是辛算法和R－K法算得的 A_2B 分子反应系统在分子平面上的运动轨迹图，辛算法的结果与理论分析一致，从势能最大值开始，B 原子沿 z 轴，两个 A 原子沿两侧的对称曲线相向运动，当 B 原子和两个 A 原子运动到 y 轴上时，系统达到势能鞍点，此后，B 原子和两个 A 原子相背运动直至达到势能最大值，这是

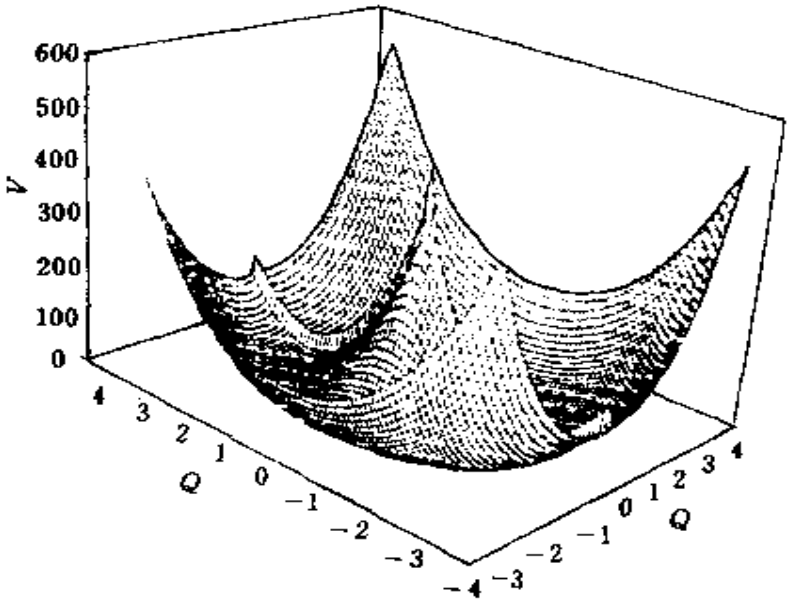


图3－24　在构形空间 Q_1－Q_2 中的势函数

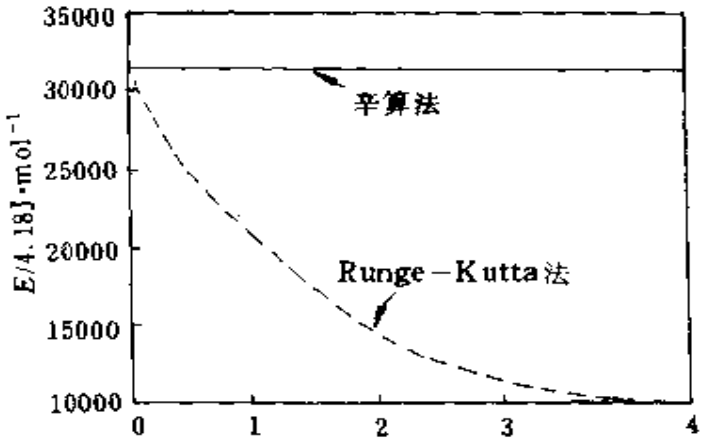


图 3 ·25 能量比较

半个周期，而后再反向运动，完成一个周期．辛算法至6.23×10^{-9}s依然作这种构型反转的准周期运动．R-K 法则不然，总能量和振动范围不断减小，计算到 8.9×10^{-10}s 时总能量已低于势能鞍点值，构型不再反转，以后总能量继续减小，至 6.23×10^{-9}s 时，B 原子和两个 A 原子各自在某个位置附近作微小振动，与理论分析不符．我们还采用一阶、二阶辛算法和欧拉折线法、改进的欧拉折线法作计算，也得到相同的结论．

我们对 A_2B 模型分子反应系统的上述计算表明，传统数值方法，如 R-K 法，亚当斯法、欧拉折线法，不能保持微观反应系统经典运动的辛结构，不可避免地渗入非哈密顿“耗散”，经长时间计算后面目全非；辛算法保持微观反应系统经典运动的辛结构和固有性质，没有非哈密顿“耗散”，适于多步数、长时间的计算，采用辛算法代替传统数值方法有可能克服目前经典轨迹计算中存在的困难，将理论计算推进到微观反应动力学研究，从根本上改进经典轨迹方法．

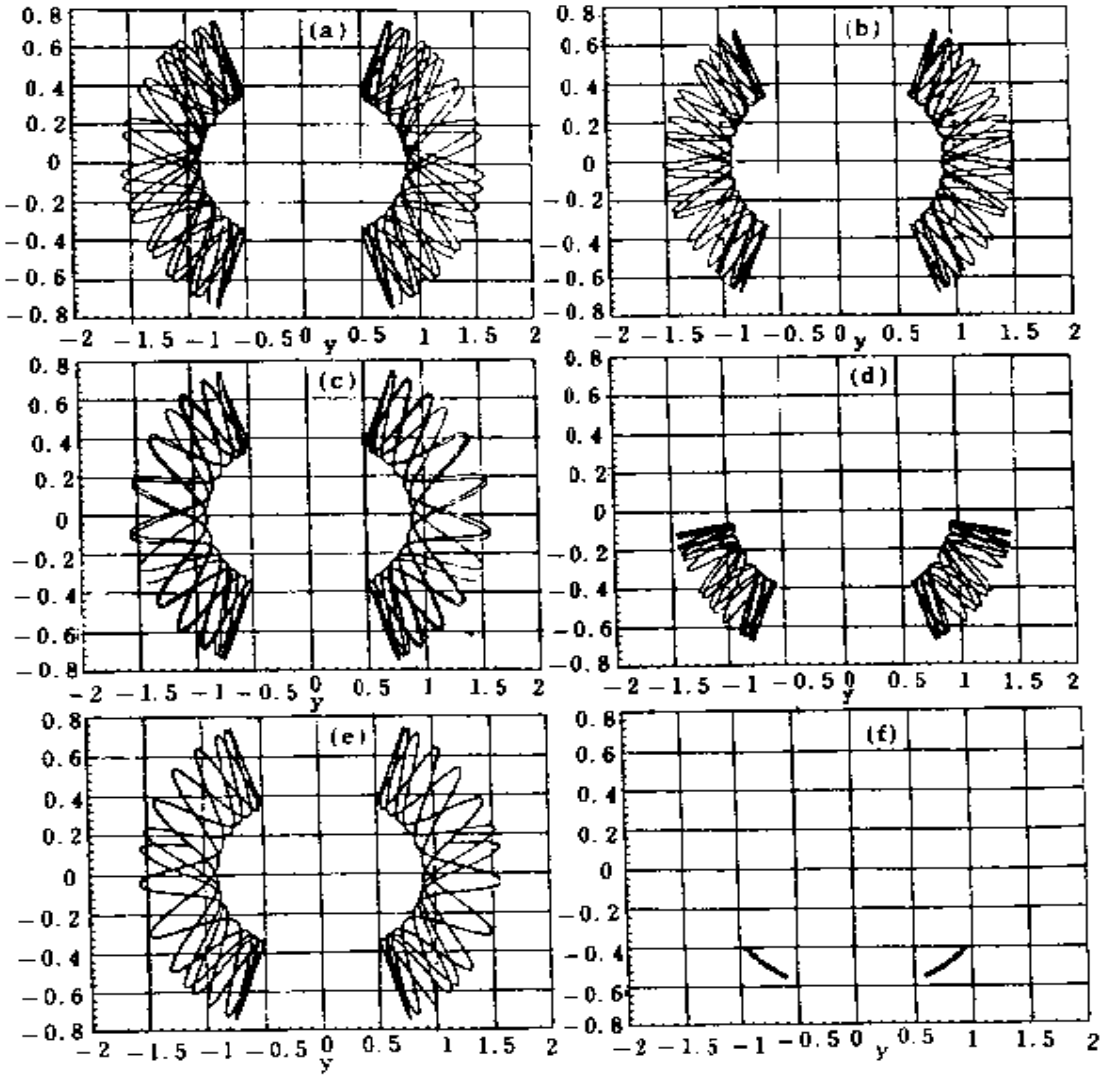


图 3－26　在椅形空间中的轨道(步长 4.45×10^{-16}s)

(a)(b)周期范围从 4.45×10^{10}s 到$(4.45\times10^{10}+4.45\times10^{13})$s. (c)周期范围从 6.23×10^{-9}s 至$(6.23\times10^{-9}+4.45\times10^{13})$s. (e)(f)周期范围从$6.23\times10^{-9}$s 至$(6.23\times10^{-4}+4.45\times10^{13})$s. (a)(c)(e)是辛算法轨道，(b)(d)(f)是 Runge-Kutta 算法轨道

3.8.6 双原子微观反应系统经典轨迹的计算

考虑双原子微观反应系统 AB 在电子势能面上的经典运动. 设 A 和 B 的质量分别为 m_1 和 m_2，在 A，B 连线上以质心为原点 O，取定坐标轴 Ox，A 和 B 原子的坐标分别为 $-x_1$ 和 x_2. 容易求得在 AB 双原子系统的广义坐标 $q = x_2 + x_1$ 和广义质量 $M = \frac{m_1 m_2}{m_1 + m_2}$，从而广义动量 $p = M\frac{\mathrm{d}q}{\mathrm{d}t}$ 和动能 $U(p) = \frac{p^2}{2M}$. 电子势能函数取 Morse 势 $V(q) = D\{e^{-2a(q-q_e)} - 2e^{-a(q-q_e)}\}$，参数 D、a、q_e 采用 E. Ley-Koo 等最近提出的. 于是 AB 反应系统的总能量 $H(p,q) = U(p) + V(q)$，经典运动的正则方程为

$$\frac{\mathrm{d}p}{\mathrm{d}t} = -\frac{\mathrm{d}V(q)}{\mathrm{d}q} = f(q),\quad \frac{\mathrm{d}q}{\mathrm{d}t} = \frac{\mathrm{d}U(p)}{\mathrm{d}p} = g(p),$$

这是一个可分哈密顿系统，可以采用显式格式求解正则方程的初值问题而求得数值解

$$t^k = kh,\quad p^k = p(t^k),\quad q^k = q(t^k),$$

继而求得 AB 反应系统的经典轨迹

$$x_1 = \frac{m_2 q}{m_1 + m_2},\quad x_2 = \frac{m_1 q}{m_1 + m_2}$$

以及系统的动能、势能和总能量随时间的变化.

我们对同核双原子反应 Li_2 和 N_2 的某些态、异核双原子反应系统 CO，CN 的某些态采用一阶，二阶，四阶显式辛格式计算经典轨迹和总能量，并分别与欧拉折线，2 阶、4 阶 R - K 法作比较. 图 3 - 27，图 3 - 28，图 3 - 29 是采用 4 阶显式辛格式和 4 阶 R - K 法取步长 $h = 0.005$ 计算双原子反应系统 Li_2 和 $X^1\Sigma_g^+$ 态（$D = 8541\mathrm{cm}^{-1}$，$q_e = 2.67328\text{Å}$，$a = 0.867\text{Å}^{-1}$），初态为 $q(0) = q_e$，

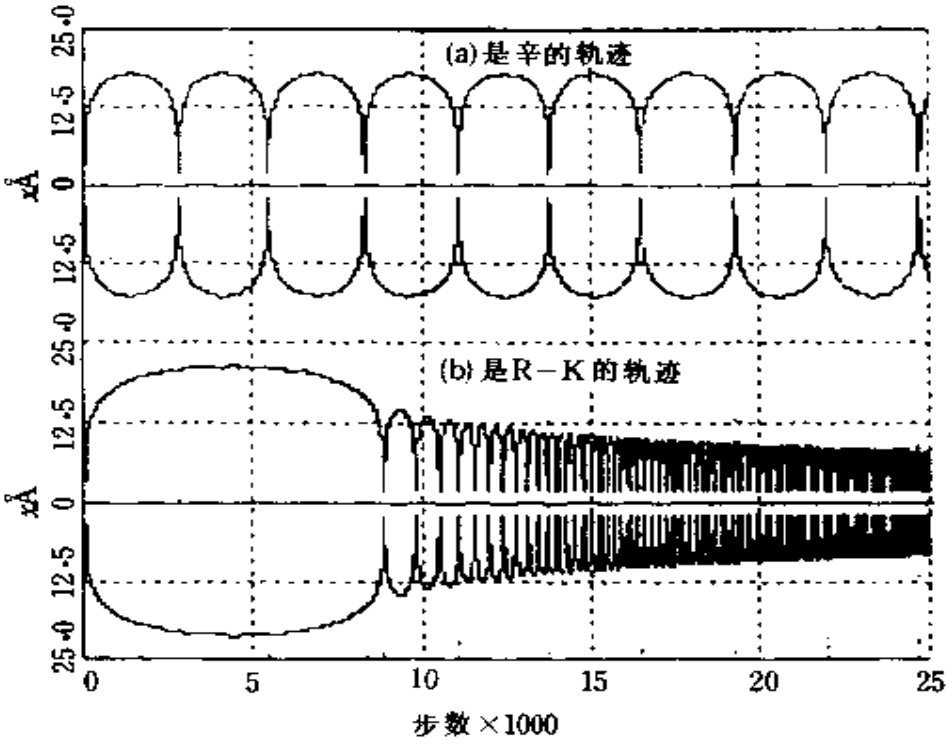


图 3－27　双原子 Li 经典轨迹(其中 Å＝0.1nm)

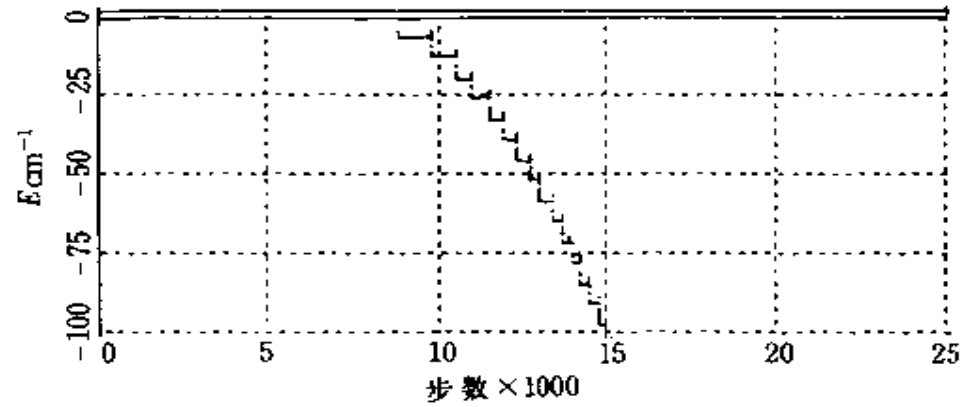


图 3　28　能量比较

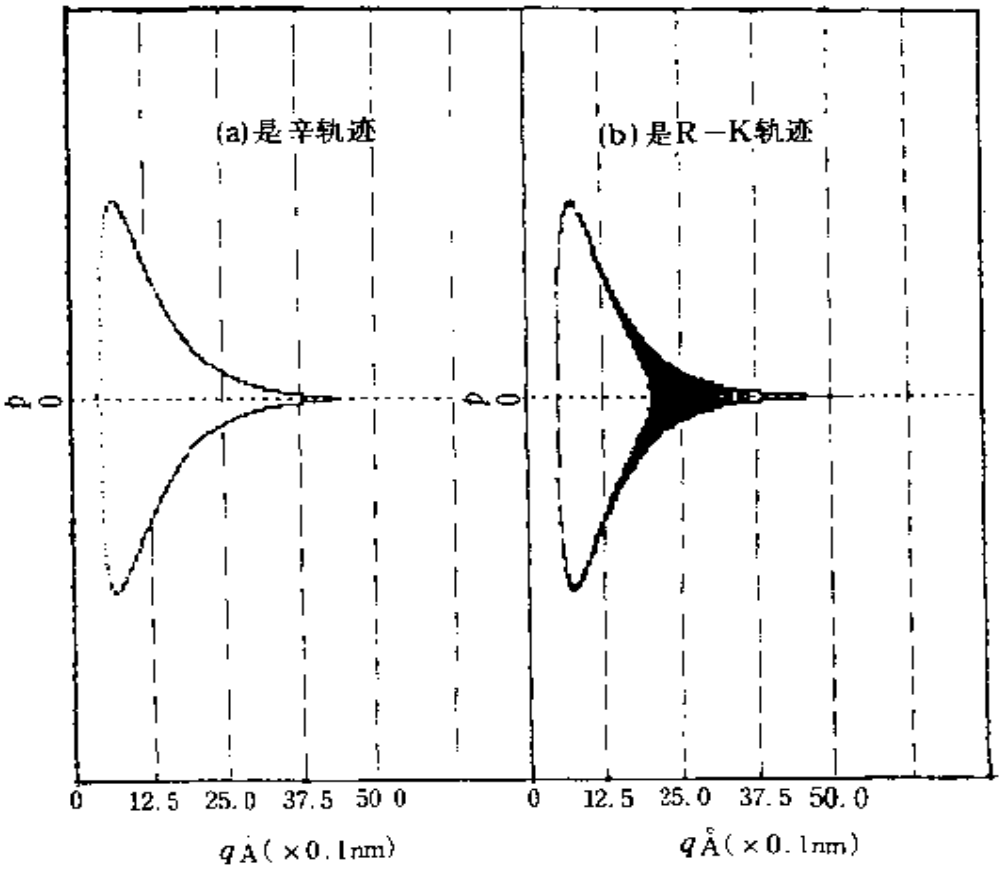


图 3－29　在 $q-p$ 相平面上的轨道

$p(0)=\sqrt{2MD}$　0.0001 的经典轨迹，总能量和 q　p 相平面上的轨道，结果显示，辛算法(我们曾计算到 10^6 步 $=2.23\times10^{-10}$s)仍保持系统的总能量守恒，两个 Li 原子作周期振动，$p-q$ 相平面上的轨道保持不变，与理论和实验一致；R－K 法则不然，至 3000 步时总能量，两个 Li 原子的振动周期和振幅开始变小，$q-p$ 相平面上的轨道沿 q 方向变扁，至 50000 步(1.12×10^{-11}s)时已变得面目全非，两个 Li 原子分别在某位置附近作微小振动，与理论和实验不符，对 Li_2 双原子反应系统采用一阶、二阶格式以及对 N_2，

CO 和 CN 反应系统的计算都得到了类似的结果.

我们对真实的 Li_2, N_2 和 CO, CN 双原子反应系统的辛与非辛算法的计算再一次表明,辛算法保持微观反应系统经典运动的辛结构和固有性质不变,适于多步数、长时间的计算;采用辛算法代替传统数值方法有可能克服目前经典轨迹计算中存在的困难,将理论计算推进到微观反应动力学研究,从根本上改进经典轨迹计算方法.

3.9 不变流形与分歧计算

科学与工程中的动力系统多属非线性并存在一些十分复杂的行为状态,如分歧、混沌现象等,研究动力系统的行为状态除了解析方法之外,数值方法是不可缺少的工具.传统的数值方法往往因破坏或失去原系统固有的结构、性质而出现许多虚假现象和计算过程的不稳定性,缺乏可靠模拟和长时间跟踪动力系统行为状态的能力,本项目(大规模科学与工程计算的方法和理论,攀登计划项目)首席科学家冯康院士发现了传统算法的弊病,于 1984 年在国际上首创了动力系统保结构算法的研究,提出利用动力系统算法研究动力系统,这对于动力系统的研究有着深远的意义,从动力系统的理论中发现,许多分歧现象是由动力系统的"特征结构"所决定的,一些算法带来的虚假行为也正是由于破坏原系统的特征结构所致,因此,将保结构算法在动力系统不变流形与分歧计算中的应用作为课题"动力系统与计算方法"的一项重要内容.不变流形(如中心流形、不变环面、同宿与异宿轨等)对于动力系统的分析、研究是十分重要的,其本身就是对系统行为的很好刻划,同时

也是约化复杂动力系统的重要工具．然而，由于不变流形隐含在微分系统中，很难求出它的明显表达式．因此，如何利用数值方法计算不变流形也是动力系统分析、计算中的重要研究课题．

3.9.1 一个例子的启示

作为一个典型例子，讨论 $x-y$ 平面上的非线性系统

$$\begin{cases}\dfrac{dx}{dt} = sx - y - x(x^2+y^2);\\ \dfrac{dy}{dt} = x + sx - y(x^2+y^2),\end{cases}$$

s—实参数．若令 $x+iy=z(i=\sqrt{-1})$，此系统又可写成

$$\frac{dz}{dt} = (i+s-|z|^2)z.$$

不难看出，对任意参数值 $s>0$，方程存在唯一的一条闭轨 $\Gamma(s)$：$|z|=\sqrt{s}$（周期解）．换句话说，在空间(z,s)中，轨道族$\{\Gamma(s),\ s>0\}$是系统从$(z,s)=(0,0)$（分歧点）始发的一条 Hopf 分歧曲线（空平衡状态 $z=0$ 到周期状态 $|z|=\sqrt{s}$ 的一类分歧现象）．另外，由常微中的线性化方法可知，对于任意给定的 $s>0$，轨道 $\Gamma(s)$ 具有整体稳定性，即方程以 z 平面上除 $z=0$ 外任意一点作为初始点的轨道当 $t\to+\infty$ 时均趋于 $\Gamma(s)$，可见 $\Gamma(s)$ 就是系统的整体吸引子．

当运用数值方法分析微分动力系统行为时，人们总是用某种离散形式的差分方程近似替代系统的微分方程，这样一来，自然有一个颇为重要的问题：所借助的离散系统（差分方程）能否或在何种程度上近似原连续系统（微分方程）的行为、状态？这个问题关系到数值分析、计算方法的可靠性．

针对上述例题，首先考察一下常用的显式欧拉差分格式

$$z_{n+1} = z_n + h(i + s - |z_n|^2)z_n, \quad n = 0,1,\cdots,$$

其中 $h > 0$ 是时间步长，差分方程可视为含双参数：$s \geq 0, h > 0$ 的迭代映射族

$$z_{n+1} = \phi(z_n, s, h), \quad n = 0,1,\cdots,$$

当 h 很小时它被看作是近似方程的一个离散系统，相应于连续系统的平衡状态 $z = 0$，迭代映射族具有不动点 $z = 0$，按照对应关系，与连续系统的闭轨相应的应该是映射族 $\phi(\cdot; s, h)$ 的不变圆环．已知，对固定的 $s > 0$，$\Gamma(s)$ 是系统的唯一闭轨，然而容易看到经过上述差分近似所得离散系统却有两个不变圆环：

$$\Gamma_h^-(s): |z| = \left|s + (1 - \sqrt{1 - h^2})/h\right|^{\frac{1}{2}} = r_h^-;$$

$$\Gamma_h^+(s): |z| = \left|s + (1 + \sqrt{1 - h^2})/h\right|^{\frac{1}{2}} = r_h^+,$$

其中 Γ_h 可视为"真实圆环"$\Gamma(s)$ 的一个近似（当 $h \to 0$ 时，$r_h^- \to \sqrt{s}$），而另一个不变圆环 Γ_h^+ 则是"过剩"的，在连续系统中找不到它的"对应物"，因此被称为"ghost cycle"，由于上述"过剩"不变圆环的出现，离散系统在相空间中的行为出现了一些严重偏离原连续系统行为的情况，甚至产生了虚假的混沌现象，通过分析比较它们在无穷远点性质的差别，揭示出显式欧拉差分格式产生"ghost cycle"的原因．

为了消除显式欧拉格式的弊病，这里我们建议采用下述差分方程作为连续方程的近似

$$z_{n+1} = z_n + h(i + s - |z_n|^2(z_{n-1} + z_n)/2, n = 0,1,\cdots,$$

记 $\{z_n\}$ 为它的解，并令 $q_n = |z_n|^2$，则有

$$\left[1-\frac{h}{2}(s-q_n)\right]^2+\frac{h^2}{4}\,q_{n+1}$$

$$=\left[1+\frac{h}{2}(s-q_n)\right]^2+\frac{h^2}{4}\,q_n.$$

由此可知，离散系统对每个固定 $s>0$ 具有唯一的不变圆环即 $z=\sqrt{s}$（此与连续系统保持一致），不出现"ghost cycle"，进一步，若令 $z_n=\sqrt{s}e^{i\theta_n}$ 即 $z_n\in\Gamma(s)$，通过计算可验证：$z_{n+1}\in\Gamma(s)$ 并且

$$z_{n+1}=\sqrt{s}e^{i\theta(\theta_n+\Delta)},\quad \Delta=h+O(h^3).$$

以上初步分析表明，当选用差分格式来模拟和分析连续系统的分岐行为和结构将能获得可靠的结果．根据分岐理论作进一步的分析将发现，类似于连续系统存在的一些分岐现象是由相应的线性化系统的"特征结构"决定的．前面第二种差分格式之所以能够可靠的模拟连续系统的分岐行为是因为它较好的保持了原系统的特征结构，而显式欧拉格式的弊病正是由于它在一定程度上破坏了原系统的特征结构所致．从上述例子得到如下启示：当用差分逼近构造微分动力系统的离散近似模型时，如常规的显式欧拉差分格式等，可能产生一些虚假又称"过剩"（原连续系统中根本没有）的现象，与此同时亦可能使原系统的行为状态发生根本性的改变（如在例子中所看到的 ghost cycle 和虚假混沌现象）．另外，在模拟原系统行为如分岐行为方面，不同差分格式（即数值方法）的性能具有很大的差别，针对模拟、分析微分动力系统特定行为的需要，尤其是关于分岐问题以及系统的整体、长时间行为的分析，选用或构造保持系统结构（几何的、特征、对称性等）的数值方法是十分重要的．因此，我们曾预言由冯康院士所倡导的保结构算法的思想将会对动力系统的分析和计算起重大的推动作用，并就保结构

算法在动力系统不变流形与分岐问题的数值分析和计算中的应用方面开展了一系列的研究工作.

3.9.2 对称格式与主部对称格式

现在介绍一类具有保持特征结构性质的重要差分格式,考虑一般 m 维自治微分系统

$$\frac{\mathrm{d}y(t)}{\mathrm{d}t} = f(y(t)), \qquad y = (y_1, y_2, \cdots, y_m),$$

函数 $f: R^m \to R^m$ 完全确定这一系统的机制,称 f 为此系统的速度向量场.不妨假定 $f(y)$ 充分光滑并且对任意 $y_0 \in R^m$,初值问题

$$\frac{\mathrm{d}y}{\mathrm{d}t} = f(y), \qquad y(0) = y_0,$$

存在唯一解 $y(t), t \in R$,记

$$y(t) = \Phi(t, y_0) = G^t(y_0).$$

$$G^t : y_0 \to G^t(y_0), \forall\, y_0 \in R^m$$

称作方程的"相流",在前面的假定下,$\{G^t, t \in R\}$ 构成 R^n 上一个单参数变换群,即满足"群条件":

$$G^0 = I$$

(恒同映射);

$$G^t \circ G^s = G^{t+s}, \qquad \forall\, t, s \in R.$$

当 t 固定,称映射: $y_0 \to \Phi(t, y_0) = G^t(y_0)$ 为方程的"t - 流".设 $h \neq 0$ 为时间步长并记 $t_n = nh$.用于逼近微分方程的"单步"差分格式的一般形式是

$$y_{n+1} = \varphi(h, y_n), \qquad n = 0, 1, \cdots,$$

例如,当 $\varphi(h, y) = y + hf(y)$ 时此即显式欧拉格式,这里,称 $\phi(h, \cdot) = g^n(\cdot)$ 为离散的"h- 流",它被当作连续的"h- 流"即

$\Phi(h,\cdot)$ 的一个近似．如果

$$\varphi(x,y)=\Phi(h,y)+O(h^{p+1}),\quad p\geqslant 1,$$

则称为 p 阶精度差分格式，其次，定义它的共轭格式为

$$y_{n+1}=\phi^{*}(h,y_n),\quad n=0,1,\cdots,$$

其中 $\phi^{*}(-h)$ 满足条件

$$\phi^{*}(-h,\phi(h,y))=y,$$

对任意充分小的 h 和 $y\in R^{m}$．据此定义，反过来说也是对的，容易验证，显式欧拉格式

$$y_{n+1}=y_n+hf(y_n)$$

与隐式欧拉格式

$$y_{n+1}=y_n+hf(y_{n+1})$$

是一对互为共轭的格式．

定义 1　给定格式，如果它是自共轭的即满足

$$\varphi(-h,\varphi(h,y))=y,$$

对任意充分小 h 和 $y\in R^{m}$．此时格式称为对称格式[此类格式是 H.J. Stetter(1970 年) 首次提出的]．令 $y=y_n$，由此可知

$$y_{n+1}=\varphi(h,y_n),\quad y_n=\varphi(-h,y_{n+1}).$$

这表明任一对称格式都是"可逆"的，反之，任一可逆的格式也必定是对称格式，可见，对称格式等价于可逆格式，另外，若采用记号 $g^{h}(y)=\varphi(h,y)$，条件又可写成

$$g^{-h}\circ g^{h}=I.$$

将此与 $\{G^{t},t\in R\}$ 的性质比较，便可看到对称格式在一定的意义下保持了原问题中解算子 G^{t} 关于参数 t 的群性质．对称格式的简单例子有：

$$y_{n+1} = y_n + hf\left(\frac{y_{n+1} + y_n}{2}\right)$$

(中心欧拉格式),

$$y_{n+1} = y_n + \frac{h}{2}[f(y_{n+1}) + f(y_n)]$$

(梯形格式).

容易证明:将任一格式 $\varphi(h,\cdot)$ 与其共轭格式 $\varphi^*(h,\cdot)$ 复合所得格式 $\phi^* \circ \phi$ 或 $\phi \circ \phi^*$ 均为对称格式.基于这一事实,利用 Yoshica 和秦孟兆复合的方法可以构造出多种多样的高阶精度对称格式

一个 s- 级 Runge-Kutta 方法 $M(A,b)$:

$$Y_i = y_n + h\sum_{j=1}^{s} a_{ij}f(Y_j), \qquad i = 1,2,\cdots,s,$$

$$y_{n+1} = y_n + h\sum_{j=1}^{s} b_j f(Y_j),$$

当消去过渡值 $\{Y_i\}$ 后所得单步格式属于对称格式时,则称 $M(A,b)$ 为对称型 Runge-Kutta 方法,为了构造此类格式,取

$$0 \leq c_1 \leq c_2 \leq \cdots \leq c_s \leq 1,$$

利用在区间 $[t_n, t_{n+1}]$ 上的点:$t_n + c_i h$, $i = 1,2,\cdots,s$ 的"配置"即方程

$$y(t_n + c_i h) = y(t_n) + h\int_0^{c_i} f(y(t_n + h))\mathrm{d}h,$$

$$i = 1,2,\cdots,s$$

和拉格朗日插值近似

$$f(y(t_n + \eta h)) \approx \sum_{j=1}^{s} l_j(\eta) f(y(t_n + c_j h))$$

可构造出一个 s- 级 Runge-Kutta 方法,其中系数

$$a_{ij} = \int_0^{c_i} l_j(\eta)\mathrm{d}\eta, \qquad b_j = \int_0^1 l_j(\eta)\mathrm{d}\eta.$$

定理1 设 $M(A,b)$ 是由上面所定义的 s- 级 Runge-Kutta 方法，则

(1) $M(A,b)$ 为对称型的充分与必要条件是 $\{c_j\}_{j=1}^{s}$ 在 $[0,1]$ 上的分布关于中点 $h=\frac{1}{2}$ 对称；

(2) 当且仅当 $\{c_j\}_{j=1}^{s}$ 为高斯 - Legendre 节点时，$M(A,b)$ 具有 $2s$ 阶精度；

(3) 采用高斯 - Legendre 节点所定义的方法 $M(A,b)$ 具有如下性质：$b_j \geq 0, j=1,2,\cdots,s$ 并且，$Q=(q_{ij})_{s\times s}=\Theta$(零矩阵)，其中 $q_{ij}=b_ia_{ji}-b_ib_j, \quad i,j=1,2,\cdots,s.$

关于性质(3)的一点注释：

考虑如下"test"系统

$$\frac{\mathrm{d}y}{\mathrm{d}t}=f(y),\langle f(y)-f(z),y-z\rangle=0,\forall y,z\in R^m,$$

易知此系统的任一轨道满足 $|y(t)|=|y(0)|$，即 R^m 中以原点为中心的任一球面皆是系统的不变流形．现将格式用于系统的计算，当条件 iii) 成立时，可以证明：

$$|y_n|=|y_0|,n=1,2,\cdots,$$

可见此种格式很好地保持了原连续系统轨道和不变流形的几何性质．

下面，特地分析一下对称格式的保特征结构的性质，这里所谓系统的特征结构说的是其线性化问题

$$\frac{\mathrm{d}y}{\mathrm{d}t}=Df(\bar{y})y,$$

这里 $\bar{y}$ 是方程的某一平衡态之解算子 $G_A=e^A, A=Df(\bar{y})$ 的特征结构，对应于算子 G_A，我们定义函数 $g_\lambda=e^\lambda, \lambda\in C$，并将 A 的

特征值分成三部分：

$$\sigma(A) = \sigma_- \cup \sigma_0 \cup \sigma_+,$$

其中 σ_-，σ_+ 分别由带负实部和带正实部的特征值组成，而 σ_0 则是那些落在虚轴上的特征值之集合．固定 t，将函数 $g_A^t : \lambda \to g_A^t$ 视作复数平面上的一个 1－1 映射，它将原属虚轴的点集 σ_0 变为单位圆周上的点集，而将 σ_- 和 σ_+ 分别变至单位圆内部和外部．连续问题解算子 G_A 的这一个变换特征就是"离散化"或"近似"的过程中希望能够得到保持或需要继承的．

对于微分系统的一个给定差分格式，将它应用到线性化问题可得到解算子 e^{hA} 的近似：

$$\psi(hA) \approx e^{hA} (\text{或 } \psi(h\lambda) \approx e^{h\lambda}),$$

例如，对应于显式欧拉格式和中心欧拉格式（或梯形格式）的函数 $\psi(h\lambda)$ 分别是

$$\psi_E(h\lambda) = 1 + h\lambda - e^{h\lambda} + O(h^2);$$

$$\psi_C(h\lambda) = \frac{1 + \frac{h\lambda}{2}}{1 - \frac{h\lambda}{2}} - e^{h\lambda} + O(h^3).$$

现在，仍将 $\psi(h\lambda)$ 视作复数平面上的映射．

定义 2　系统的一个差分格式，如果在它用于线性化问题时对应的函数 $\psi(h\lambda)$ 具有类似于前面描述的函数 e^{hA} 的变换特征，则称此格式是保结构特征结构的格式．

显然，$\Psi_E(h\lambda) = 1 + h\lambda$ 不具有 e^{hA} 的变换特征，所以显式格式不是一个保特征结构的格式．值得庆幸函数 $\psi_C(h\lambda)$ 是著名的 Cayley 变换，它完全保持 e^{hA} 所述的变换特征．因此按前面定义，中心欧拉格式和梯形格式属于保特征结构的格式．一般地，由于对

称格式所对应的函数 $\psi(h\lambda)$ 满足条件

$$\psi(h\lambda)\psi(-h\lambda)=1,$$

通常地 $\psi(h\lambda)$ 取为实系数有理多项式(Pade 逼近)，由此可证

定理 2 设 $\psi(h\lambda)$ 为实系数的有理多项式，则与此相应的对称格式具有保特征结构的性质.

一般说来，对称格式是隐式的并且是非线性格式(当用于非线性问题)，出于简化计算的目的，我们常用主部对称格式代替对称格式.

定义 3 设 $\bar{y}=\varphi(h,\bar{y})$，即 $\bar{y}$ 是格式的一个不动点，如果下述线性格式

$$y_{n+1}-\bar{y}=D\varphi(h,\bar{y})(y_n-\bar{y}),\qquad n=0,1,\cdots$$

为对称格式，则称格式在 $y=\bar{y}$ 附近还是主部对称的. 如此定义的格式，它在 $y=\bar{y}$ 附近仍然具有保特征结构的性质.

3.9.3 中心流形的计算

在动力系统的理论研究中，中心流形(center manifold) 是一个非常有用的工具，一个 $m\times m$ 的微分方程组若存在中心流形，将此方程组在中心流形上约化可转化为 2×2 方程组，所得简化方程包含原方程组的全部性态行为，然而，如同其他不变流形一样，中心流形是蕴含在微分方程组之中的，所以找这种流形或计算它们是一件困难的工作. 近年来，出现了一些用数值方法计算中心流形的研究工作.

先介绍一下中心流形的概念. 考虑如下典型形式的 $m\times m$ 微分方程组

$$\frac{dy}{dt} = Ay + g(y), \quad y = (y_1, y_2, \cdots, y_m),$$

其中 A 是 $m \times m$ 实矩阵，函数 $g: R^m \to R^m$ 充分光滑并满足

$$g(0) = 0, \quad Dg(0) = 0.$$

假定矩阵 A 存在一对实部为 0 的特征值 $\pm i\alpha$，而它的其余特征值之实部皆为负数，并设 Z 和 X 分别代表特征值 $= i\alpha$ 和其余特征值对应的特征空间，于是存在空间分解

$$R^m = X + Z,$$

其中 X 和 Z 的维数分别为 $(m-2)$ 和 2.

在上述关于 A 和 $g(y)$ 的假定下，动力系统理论已经证明：在 R^m 的原点的某个领域 V 内存在一个光滑函数 $q: z \to x$，它满足

$$q(0) = 0, \quad Dq(0) = 0,$$

并且由函数 $x = q(z)$ 所定义的 2 - 维流形

$$M = \{(x, z) \in V : x = q(z)\} \subset R^m$$

具有如下性质：

(1) M 关于方程的相流 G^t 是不变的，即若 $y \in M$，则

$$G^t(y) \in M, t \geq 0;$$

(2) M 是吸引的，即从 M 附近出发的轨道当 $t \to +\infty$ 时将无限地接近 M.

上述2维流形 M 被称为系统的中心流形，条件表示流形 M 通过原点并存在原点与特征空间 Z 相切. 所以，简单地说中心流形就是系统的一个通过原点并在原点与零实部特征值相应特征空间相切的不变流形.

对任意 $y \in R^m$，记其分解为 $y = x + z, x \in X, z \in Z$. 令 P 代表 $R^m \to Z$ 的投影，A_0 表示 A 在特征子空间 Z 上的限制；利用

中心流形 $x = q(z)$，方程组可化简成如下 2×2 方程组

$$\frac{dz}{dt} = A_0 z - Pg(q(z) + z), \quad z = (z_1, z_2)^T \in Z.$$

此方程组包含着原方程组的所有分歧信息. W.J.Beyn, J. Lorenz(1987) 和马富明(1988) 首先研究了借助离散差分方程去计算微分系统中心流形的问题，他们分别证明了：相应于惯常 Runge - Kutta 格式和欧拉格式皆存在依赖于步长 h 的不变流形 $M^{(h)}$，并且当 $h \to 0$ 时，这族不变流形 $\{M^{(h)} : h > 0\}$ 将收敛到微分系统的中心流形 M，然而，由此得到的近似流形 $M^{(h)}$ 并非中心流形，想利用它化简系统和进一步研究系统分歧行为这一目的是难以实现的，因此，为了能够真实地模拟系统的分歧行为宜采用"保特征结构"的差分方程即对称或主部对称格式.

对于方程组，基于积分关系式

$$y(t_{n+1}) = e^{hA} y(t_n) + h \int_0^1 e^{(1-s)hA} g(y(t_n + sh)) ds,$$

不难构造下列形式的差分方程

$$y_{n+1} = \psi(hA) y_n + G_h(y_n), \qquad n = 0, 1, \cdots,$$

$$G_h(0) = 0, \quad DG_h(0) = 0,$$

其中 $\psi(hA)$ 是算子 e^{hA} 的某个实有理多项式近似并满足

$$\psi(-hA)\psi(hA) = I.$$

按先前定义，上述格式属于"主部对称格式"，例如，格式

$$\frac{y_{n+1} - y_n}{h} = A \frac{y_{n+1} + y_n}{2} + g(y_n)$$

就是上述的一个特例，此时

$$\psi(hA) = (I - \frac{h}{2}A)^{-1}(I + \frac{h}{2}A),$$

$$G_h(y_n) = (I - \frac{h}{2}A)^{-1}g(y_n),$$

并满足上述条件格式.

下面,我们将格式写成"迭代映射"

$$y_{n+1} = \phi(h, y_n), \quad n = 0, 1, \cdots,$$

由条件成立

$$\phi(h, 0) = 0, \qquad D\phi(h, 0) = \psi(hA).$$

另外,根据当初对于 A 的假定以及条件可以断定:

矩阵 $D\phi(h, 0)$ 存在一对位于单位圆上的特征值 $\lambda_{1,2}(h) = \psi(\pm ih\alpha)$,而它的其余特征值 $\lambda_i(h), i = 3, 4, \cdots, m$ 均位于单位圆内部.

X, Z 如前面定义,仍代表 A 的特征子空间并存在空间分解

$$R^m = X + Z.$$

定理 3 设 $\phi(h, \cdot), 0 < h < h_0$ 是 R^m 中满足上述条件的一族光滑映射,则在原点的某个领域 $V_\delta: |y| < \delta$ 内一致地(关于 h)存在函数 $q_h: Z \to X$,满足

$$q_h(0) = 0, \quad Dq_h(0) = 0,$$

并且流形

$$M^{(h)} = \{(x, z) \in V: x = q_h(z)\}$$

关于映射 $\phi(h, \cdot)$ 是不变的和吸引的.

注记 $M^{(h)}$ 关于 $\phi(h, \cdot)$ 不变即对任意 $(x, z) \in M^{(h)}$,如果

$$(x', z') = \phi(h, x + z) \in V,$$

则必定有 $(x', z') \in M^{(h)}$."$M^{(h)}$ 是吸引的"含义为:

对于任意 $(x, z) \in V_\delta$,如果

$$(x_j, z_j) = \phi(h, x - z)) \in V_\varepsilon, j = 1,2,\cdots,$$

那么有 $|x_j - q_h(z_j)| \to 0$，当 $j \to +\infty$.

定理3所给出的关于映射族 $\{\phi(h,\cdot), 0 < h < h_0\}$ 的不变流形 $\{M^{(h)}, 0 < h < h_0\}$ 具有与微分方程中心流形 M 完全一样的特征，故将 $M^{(h)}$ 定义作为差分方程的中心流形，并将 $M^{(h)}$ 当做 M 的近似．上述定理可以证明.

另外，如果格式为 p 阶精度格式，即对于方程的任一精确解 $y(t)$ 成立

$$y(t_{n+1}) - \phi(h, y(t_n)) + O(h^{p+1}),$$

则可证明如下形式的误差估计

$$|q_h(z) - q(z)| \leq ch^p.$$

当 h 充分小，它表明当 $h \to 0$ 时离散系统的中心流形 $M^{(h)}$ 将按 $O(h^p)$ 的速度收敛到连续系统的中心流形 M.

从以上讨论看到，保特征结构的差分格式将是近似计算中心流形的可靠方法，目前，如何实际计算高维微分系统的中心流形仍然是动力系统数值计算中人们十分关心的研究课题.

3.9.4 Hopf 分歧与环面分歧的计算

Hopf 分歧代表从平衡（定常解）到周期振荡（周期解）的分歧现象，例如，在本节开始讨论的例子中，当参数 s 沿实数轴上变化越过原点时，系统除了定常解 $(x,y) = (0,0)$ 之外还会出现另一支振幅为 $\sqrt{s}$ 的周期解．研究这类分歧行为，对于许多科学与工程技术（如蒸汽机自动调速，反应器设计，机翼颤振以及人体神经活动的分析等）具有非常重要的意义.

为说明 Hopf 分歧的特征，现考虑如下含参数 $\lambda \in R$ 的 m 维微分系统

$$\frac{dy}{dt} = f(y,\lambda), \qquad y = (y_1, y_2, \cdots, y_m).$$

关于该系统，E. Hopf 在 1942 年曾证明：

定理 4 假设

(1) $f(\bar{y},\bar{\lambda}) = 0$;

(2) $f_y(\bar{y},\bar{\lambda})$ 具有一对简单纯虚数特征值 $\mu(\bar{\lambda}) = \pm i\beta$，并且它的其余特征值皆具有非零的实部；

(3) $\frac{d}{d\lambda}(\mathrm{Re}\mu_i(\lambda))|_{\lambda=\bar{\lambda}} \neq 0, i = 1,2, \mu_i(\lambda)$

代表 $A(\lambda) = f_y(\bar{y},\lambda)$ 的特征值.

那么在上述条件下，系统在 $(\bar{y},\bar{\lambda})$ 附近存在一族依赖于参数 λ 的周期解 $\{y_\lambda(t)\}$，$y_\lambda(t)$ 具有周期 $T_\lambda \approx \frac{2\pi}{\beta}$ 当 $\lambda \approx \bar{\lambda}$.

这里，$(\bar{y},\bar{\lambda})$ 称作"Hopf 分歧点"，$\{y_\lambda(t)\}$ 代表从此点发出的一条"Hopf 分支"，条件(2)、(3)描述了系统在 Hopf 分歧点的特征结构，通常称(3)为"横截条件"，简单地说，在 Hopf 分歧点 $f_y(y,\lambda)$ 有且仅有一对共轭特征值位于虚轴上，并当 λ 从小于 $\bar{\lambda}$ 变至大于 $\bar{\lambda}$ 时这对特征值从虚轴的一侧"横跨"进入另一侧，从产生、形成 Hopf 分支的上述机理、特征，我们深信具有"保特征结构"的差分格式一定能够成为计算这种分支的最为恰当的离散化方法.

设 $h > 0$ 为时间步长，微分方程被如下差分方程近似地替代

$$y_{n+1} = \phi(y_n,\lambda), \qquad n = 0,1,\cdots,$$

这里，$\phi_h(\cdot,\lambda)$ 被视为 $R^m \to R^m$ 的映射，为了分析离散系统行为与连续系统行为之间的关系，我们首先将 Hopf 分支的概念推广到

映射变换.

定义 4 设 $\phi(\cdot,\lambda),\lambda\in R$ 为 $R^m\to R^m$ 的光滑映射并成立：$\phi(\bar{y},\bar{\lambda})=\bar{y}$，若存在从 R 到 R^m 的周期函数族：

$$y_\lambda(t),|\lambda-\bar{\lambda}|<\delta\quad(y_{\bar{\lambda}}(t)\equiv\bar{y}),$$

使得 $\gamma_\lambda=\{y_\lambda(t):t\in R\}$ 是映射 $\phi(\cdot,\lambda)$ 的不变流形，即 $\phi\cdot\gamma_\lambda\subset\gamma_\lambda$，则称

$$\{\gamma_\lambda:|\lambda-\bar{\lambda}|<\delta\}$$

是映射 $\phi(\cdot,\lambda)$ 从不动点 $(\bar{y},\bar{\lambda})$ 发出的 Hopf 分支.

对于映射变换 $\phi(\cdot,\lambda),\lambda\in R$ 可以证明完全平行于定理 4 的结果，即有

定理 5 假定

(1) $\phi(\bar{y},\bar{\lambda})=\bar{y}$;

(2) $D_y\phi(\bar{y},\bar{\lambda})$ 具有一对共轭特征值：

$\mu_1(\bar{\lambda}),\mu_2(\bar{\lambda}),|\mu_1(\bar{\lambda})|=|\mu_2(\lambda)|=1$(即在单位圆周上)，

而它的其余特征值 $\mu_i(\bar{\lambda}),i=3,4,\cdots,n$ 皆按模不等于 1;

(3) $B(\lambda)=D_y\phi(\bar{y},\lambda)$ 的特征值 $\mu_1(\lambda),\mu_2(\lambda)$ 满足

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}|\mu_i(\lambda)|\right)_{\lambda=\bar{\lambda}}\neq 0,\quad i=1,2.$$

在上述假定下，映射 $\phi(\cdot,\lambda)$ 在 $(\bar{y},\bar{\lambda})$ 附近必定存在一条 Hopf 分支

$$\gamma_\lambda=\{y_\lambda(t):t\in R\},|\lambda-\bar{\lambda}|<\delta\quad(\gamma_{\bar{\lambda}}=\{\bar{y}\}).$$

现在再来讨论离散系统，假定离散系统满足

$$\phi_h(\bar{y},\bar{\lambda})=\bar{y},$$

这里$(\bar{y},\bar{\lambda})$是连续系统的Hopf分歧点，另外，进一步假定，当$0<h<h_0$时

$$B_h(\lambda)=D_y\phi_h(\bar{y},\lambda)$$

为算子$e^{hA(\lambda)}$的一个"保特征结构"的近似，其中$A(\lambda)=f_y(\bar{y},\lambda)$. 由此以及定理5的假定条件(2)和(3)可知，与$A(\bar{\lambda})$的纯虚数特征值$\pm i\beta$相应的矩阵$B_h(\lambda)$存在一对位于单位圆上的特征值，而$B_h(\lambda)$的其余特征值均按模大于或小于1，这样一来，由定理5推知，对任一$0<h<h_0$，映射$\phi_h(\cdot,\lambda)$而格式在$(\bar{y},\bar{\lambda})$附近确实存在一条Hopf分支

$$\gamma_\lambda^{(h)}=\{y_\lambda^{(h)}(t):t\in R\},\ |\lambda-\bar{\lambda}|<\delta\quad(\gamma_\lambda=\{\bar{y}\}).$$

另外，利用它对于微分方程的逼近性质：

$$y(t_{n+1})=\phi_h(y(t_n),\lambda)+O(h^{p+1}),p\geq 1.$$

此处$y(t)$代表连续系统的任一光滑解，可以证明当$h\to 0$时$\gamma_\lambda^{(h)}$将收敛于γ_λ并可获得理论误差估计$\mathrm{dist}(\gamma_\lambda^{(h)},\gamma_\lambda)=O(h^p)$.

当格式为方程的对称或$y=\bar{y}$处的主部对称差分逼近时，通常情形可将$B_h(\lambda)$写成$\psi(hA(\lambda))$的形式，并且$\psi(hA(\lambda))$是线性化问题

$$\frac{\mathrm{d}\tilde{y}}{\mathrm{d}t}=A(\lambda)\tilde{y},\ \tilde{y}=y-\bar{y},\ A(\lambda)=f_y(\bar{y},\lambda)$$

解算子$e^{hA(\lambda)}$的"保特征结构"近似(见定义3)，所以上面的分析结果成立. 另外，根据前一小节的分析，对称格式与主部对称格式具有继承原系统中心流形的性质. 因此，在实际计算中，可先计算出格式的中心流形，然后将它化简成一个二维系统，求此简化系统的Hopf分支便可得到$\{\gamma_\lambda^{(h)}\}$.

下面，我们讨论环面分歧的计算问题.

环面(torus)分歧是在Hopf分支上再发生的新分歧现象，故又称为二次Hopf分歧，对于 m 维系统，假若存在双周期函数

$$r = r(\theta) \in R^m,\ \theta = (\theta_1, \theta_2) \in R^2,$$

使得流形 $M = \{r : r = r(\theta), \theta \in R^2\}$

是所考虑系统的不变流形，则称 M 为此系统(微分方程或映射)的不变环，从圆周轨道到环面的分歧现象称为环面分歧，这是导致混沌现象的途径之一.

W. F. Langford(1979) 研究过如下形式的双参数、自治常微系统

$$\frac{dy}{dt} = f(y, \sigma, \lambda),$$

这里 $y \in R^m, \sigma, \lambda \in R$ 是参数，其中主要的假定条件为：

(1) $f(0, \sigma, \lambda) = 0$，当 $\sigma, \lambda \in (-\delta, \delta)$；

(2) $f_y(0,0,0)$ 具有特征值0和 $\pm i\beta_0 (\beta_0 > 0)$，它们的代数重数是1，其余特征值的实部皆小于零；

(3) $A(\sigma, \lambda) = f_y(0, \sigma, \lambda)$ 存在特征值 $\gamma(\sigma, \lambda)$ 和 $\alpha(\sigma, \lambda) \pm i\beta(\sigma, \lambda)$，满足 $\gamma(0,0) = 0, \alpha(0,0) = 0, \beta(0,0) = \beta_0$；

(4) $\left.\frac{\partial(\alpha, \gamma)}{\partial(\sigma, \lambda)}\right|_{(\sigma,\lambda)=(0,0)} \neq 0$；

(5) $\frac{\partial \alpha}{\partial \lambda}(0,0) \neq 0,\ \frac{\partial \gamma}{\partial \lambda}(0,0) \neq 0.$

根据上述假定条件，首先证明在参数平面 (σ, λ) 原点的一个邻域内存在一条Hopf分歧曲线，进一步又发现：由于静态分支 $y = 0$ 和Hopf分支的相互作用，在初始的Hopf分支曲线上将发生新的分歧(周期加倍或者是环面分歧行为)，W. F. Langford 并对产生

环面分歧的情形作了细微的分析.

为了用数值方法分析、计算系统的 Hopf 和环面分歧行为，很自然地应该采用有利于保持系统特征分布结构即性质(2) – (5)的差分格式，这就是对称或主部对称格式，我们基于展开

$$f(y,\sigma,\lambda) = A(\sigma,\lambda)y + Q(y) + R(y,\sigma,\lambda);$$

$$A(\sigma,\lambda) = f_y(0,\sigma,\lambda);$$

$$Q(y) = \frac{1}{2}f_{yy}(0,0,0)yy$$

构造方程的如下形式的主部对称差分逼近：

$$\frac{y_{n+1}-y_n}{h} = A(\sigma,\lambda)\frac{y_{n-1}+y_n}{2} + Q(y_n) + R(y_n,\sigma,\lambda).$$

通过对它的特征结构的讨论与局部渐近解分析，我们证明了：当原连续系统满足条件(1) – (5)时且 h 充分小时，相应的离散系统在参数平面(σ,λ)中存在一条经过原点的曲线 C_0，使得对任意$(\sigma,\lambda) \in C_0$，在 $y = 0$ 附近一定存在环面分歧，即双周期函数 $g(s,t) \in C^1([0,2\pi],[0,2\pi])$ 使得

$$M = \{g(s,t), 0 \le s,t \le 2\pi\}$$

关于它所定义的映射 $\phi_h(\cdot,\sigma,\lambda)$ 是不变的．可见，运用上述格式模拟，计算系统的 Hopf 与环面分歧行为会是可靠的.

L. Dieci、Lorenz、黄明游和 T. Küpper 曾研究过如下形式微分系统

$$\begin{cases} \dfrac{d\theta}{dt} = f(\theta,\gamma), \theta \in T^p = \{\theta = (\theta_1,\cdots,\theta_p); \theta_j \in R(\mathrm{mod}2\pi)\}, \\ \dfrac{d\gamma}{dt} = g(\theta,\gamma), \ \gamma \in R^q, \ t \in R, \end{cases}$$

$$f: T^p \times R^q \to R^q, \qquad g: T^p \times R^q \to R^p,$$

寻求和计算该系统形如

$$M = \{(\theta, R(\theta)) : \theta \in T^n\} \subset T^n \times T^n$$

的不变环面.

3.9.5 同宿与异宿轨的计算

同宿与异宿轨乃是区分动力系统相空间中性质相异区域之间的分界线(separatrix),它对于刻划动力系统的整体行为具有重要意义.

考虑含参数的微分系统

$$\frac{dy}{dt} = f(y, \lambda), \qquad y \in R^m, \qquad \lambda \in R^l,$$

设 $f(y_\pm, \bar{\lambda}) = 0$,即 y_-, y_+ 为平衡点,并设 $\bar{y}(t)$, $-\infty < t < +\infty$ 为系统的一条轨道,它满足

$$\lim_{t \to -\infty} y(t) = y_-, \lim_{t \to +\infty} \bar{y}(t) = y_+,$$

那么,当 $y_- = y_+$ 时称 $\bar{y}(t)$ 为同宿轨(homoclinic orbit),而当 $y_- \neq y_+$ 时则称之为异宿轨(heterclinic orbit).

计算同宿与异宿轨的问题,首先要解方程 $f(y, \lambda) = 0$,求出平衡点 $(y_+, \bar{\lambda})$,然后求解

$$\begin{cases} \dfrac{dy(t)}{dt} = f(y(t), \bar{\lambda}), & -\infty < t < +\infty; \\ y(-\infty) = y_-, \qquad y(+\infty) = y_+, \end{cases}$$

这是一个无穷区间 $(-\infty, +\infty)$ 上的边值问题.

近几年来出现了一些利用数值方法分析、计算同宿与异宿轨的研究工作. W. J. Beyn(1989 年)首先提出了采用射影边界条件计算"非退化"同宿、异宿轨(y_- 和 y_+ 均属双曲面平衡点,此时

$\bar{y}(t)$ 是系统在 y_- 点的不稳定流形与 y_+ 点的稳定流形的交线)，同时他还对离散系统即映射族的同宿、异宿轨进行了系统的研究，最近，邹永魁和 W. Z. Beyn 研究了"退化"同宿轨($y_- = y_+$ 属于鞍点，此时 $\bar{y}(t)$ 是系统在该点的不稳定中心流形与另一稳定不变流形的交线)，对于这类同宿轨的计算，涉及到中心流形的逼近，所以采用具有对称性的差分方程作原连续系统中微分方程的近似是必要的.

下面，就同宿轨的计算问题作一简单介绍，类似的方法亦可推广到异宿轨的计算. 考虑如下动力系统

$$\frac{dy}{dt} = f(y,\lambda),$$

这里 $y \in R^m$，$\lambda \in R^l$，$l = 1$ 或 2，其中函数 $f: R^m \times R^l \to R^m$ 充分光滑.

假定系统具有平衡点 $\xi(\lambda)$，即 $f(\xi(\lambda),\lambda) = 0$ 在 $l = 1$ 的情形. 假定 $\xi(\lambda)$ 为双曲平衡点(即 $A(t) = f_y(\xi(\lambda),\lambda)$ 没有实部为零的特征值)，当 $l = 2$ 时，假定存在一点 $\lambda^0 = (\lambda_1^0, \lambda_2^0)$ 使得 $\xi^0 = \xi(\lambda^0)$ 是方程 $f(y,\lambda) = 0$ 的简单、二次鞍-结平衡点，更具体地说，就是矩阵 $f_y(\xi^0,\lambda^0)$ 有一简单零特征值且在参数 (λ_1,λ_2) 平面上存在曲线 $\lambda_1 = \alpha(\lambda_2)(\lambda_1^0 = \alpha(\lambda_2^0))$ 使得：对固定的 λ_2，$\xi(\alpha(\lambda_2),\lambda_2)$ 是 $f(y,\lambda) = 0$ 的简单、二次平衡点.

对于 $l = 1$ 情形，可从数值求解方程

$$f(y,\lambda) = 0$$

计算平衡点 $\xi(\lambda)$；

对于 $l = 2$ 情形，则需求解扩展方程

$$\begin{cases} f(y,\lambda_1,\lambda_2)=0, \\ D_y f(y,\lambda_1,\lambda_2)\phi=0, \\ \phi^T\phi-1=0, \end{cases}$$

计算鞍－结点 $\xi(\lambda_1,\lambda_2)$ 和相应的参数曲线 $\alpha(\lambda_2)$.

进一步假定存在 $\lambda=\lambda_0$，使得方程具有唯一的同宿轨

$$\bar{y}(t):\lim_{t\to\pm\infty}\bar{y}(t)=\xi_0=\xi(\lambda^0),$$

当 $l=1$ 时 $\bar{y}(t)$ 是“非退化”同宿轨（又称双曲同宿轨）；当 $l=2$ 时 $\bar{y}(t)$ 则属“退化”同宿轨（又称鞍－结同宿轨）.

为了使用计算机进行系统的数值模拟，必须将其中的微分方程离散化，比如，用如下形式的隐式差分方程

$$F(y_{n+1},y_n,\lambda,\delta)=0, \qquad n=0,\pm1,\cdots$$

（其中 $h>0$ 为时间步长）近似代替原系统，当 h 充分小，上式又可改写成迭代映射

$$y_{n+1}=\phi(y_n,\lambda,\delta),$$

R^m 中一条离散轨道 $\Gamma_h=\{y_n\}$，$n\in z$，如果

$$\lim_{n\to\pm\infty}y_n=\xi,$$

其中 ξ 是 $\phi(\cdot,\lambda,h)$ 的不动点，则称 Γ_h 为它的同宿轨道.

采用离散化方法计算同宿轨，首先要研究的问题是差分方程是否能够定性地继承原方程的解的结构即有无同宿轨？这对于“退化”同宿轨来说，一般的差分逼近是保证不了的，但是，邹永魁和 W.Z.Beyn 证明了：当采用如下中点欧拉差分方程

$$\frac{y_{n+1}-y_n}{h}=f\left(\frac{y_{n+1}+y_n}{2},\lambda\right)$$

作近似时，那么所得相应的差分方程，确实具有继承原方程的包括双曲与鞍－结同宿轨的理想性质．这样，我们可将近似系统的离

散同宿轨 Γ_h 作为系统的同宿轨 $\Gamma = \{\bar{y}(t), t \in R\}$ 的近似.

现在,我们假定平衡点 $\xi(\lambda)$ 和参数平面上函数 $\alpha(\lambda_2)$ 业已求出,介绍进一步计算它的鞍 - 结同宿轨的方法,基本思想是将计算双曲同宿轨的射影边界条件推广到鞍 - 结同宿轨的情形,首先,对 $n \in Z$ 作限制即考虑有限集 $\{n_-, n_- + 1, \cdots, n_+\}$ 上的差分方程

$$F(y_{n+1}, y_n, \lambda, h) = 0, \qquad n = n_-, \cdots, n_+ - 1,$$

并附设边界条件

$$b_1(y_{n_-} - \xi(\lambda)) = 0, \qquad b_2(y_{n_+} - \xi(\lambda)) = 0,$$

这里,$b_i = 0, i = 1,2$ 分别是相应映射 $\phi(\cdot, \lambda, h)$ 在其不动点 ξ 处中心不稳定流形和稳定不变流形的线性近似,例如,可以取 $b_1 \in R^{m_c \times m}, b_2 \in R^{m_s \times m}$,使 b_1 和 b_2 的列分别地构成矩阵 $((D_1F)^{-1} D_2F)^T$ 的稳定不变子空间和中心不变子空间的基底(在计算双曲同宿轨的情形本应是不稳定不变子空间,即对应于正实部特征值的特征空间).

由此可见,采用欧拉中点差分格式,只要恰当地确定近似边界条件,既可以计算双曲同宿轨也可计算鞍 - 结同宿轨.所述计算方法可平行地推广到双曲异宿轨和鞍 - 结异宿轨的计算,我们研究了一个实际应用问题,成功地计算出所需的同宿和异宿轨.

4 数学物理反演问题

4.1 耳朵能“听出”鼓的形状吗

有这样一个有趣的问题：仅仅通过鼓的声音能否判断出鼓的形状？即所谓的“盲人听鼓”问题.据了解，问题最早由丹麦著名物理学家 Lorentz 在 1910 年的一次讲演中提出，它的背景来自于射线理论.以耳代目可能吗？生活经验告诉我们这是有一定道理的.例如，当物体的材料确定后，它的音调的高低和其形状密切相关，有经验的人不难发现它们之间的某些联系.现在这一问题摆到数学家面前，势必得到一番认真的研究.

一个物体的音色可以由一串谱 $\lambda_1 \leq \lambda_2 \leq \cdots$ 来确定，它们在物理上对应着物体的固有频率.“盲人听鼓”即是要求通过已知的谱来确定一个二维鼓面的形状.经过近一个世纪的研究，数学家们使用了许多

深奥而巧妙的数学手段，现在这一问题已经解决，答案是否定的。但是，从鼓声中我们确实能得到相当多的形状信息：能够"听"出鼓的面积有多大、周边有多长甚至鼓的内部是否有洞、有几个洞，而且数学家一一给出了计算公式，例如，鼓的面积可以由下式确定

$$\text{鼓面积} = 2\pi \lim_{\lambda\to\infty} \frac{\text{小于}\lambda\text{的谱的数目}}{\lambda}.$$

问题解决的最后一步是新近才得到的，1992 年 Gordon 等人构造出了两个同声鼓，它们的形状不同，却有着相同的音调，单凭耳朵无法给出鉴别！我们看到，通过数学家的努力，使人们对于这个生活中的普通问题有了透彻的理解。

"盲人听鼓"问题的研究至此告一段落，然而类似的饶有兴味的数学问题还有很多很多，它们来自生产活动的各个领域，向数学家们寻求解答。这些问题在数学上可以归属于一门新兴的分支——反问题。那么，什么是反问题呢？

4.2 什么是反问题

顾名思义，反问题是相对于正问题而言的。以前面所举的"盲人听鼓"反问题为例，它的正问题就是要在已知鼓的形状的条件下，研究其发声规律，这在数学物理历史上已经研究在先，而且比较成熟。此时鼓的所有谱都能通过一套算法利用计算机算出来。如何区分某个问题的"正""反"？这并没有一个严格的标准，但可以粗略地这样理解：世间的事物或现象之间往往存在着一定的自然顺序，如时间顺序、空间顺序、因果顺序，等等。所谓正问题，一般是按照这种自然顺序来研究事物的演化过程或分布形态，起着由因

推果的作用.反问题则是根据事物的演化结果,去可观测的现象来探求事物的内部规律或所受的外部影响,由表及里,索隐探秘,起着倒果求因的作用.可以看出,正、反两方面都是科学研究的重要内容.

尽管一些经典反问题的研究可以追溯很早,反问题这一学科的兴起却是近几十年来的事情.在科学研究中经常要通过间接观测来探求位于不可达、不可触之处的物质的变化规律;生产中经常要根据特定的功能对产品进行设计,或按照某种目的对流程进行控制.这些都可以提出为某种形式的反问题.可见,反问题的产生是科学研究不断深化和工程技术迅猛发展的结果,而计算技术的革命又为它提供了重要的物质基础.

现在,反问题的研究已经遍及现代化生产、生活、研究的各个领域.简单的概括不足以说明问题,下面具体介绍一些常见的反问题类型,希望大家能够对它有一个概括的了解.

4.3 定向设计

工业生产离不开产品设计,如何设计出优质产品使之更好地实现其功能,是关系到厂家信誉和企业生存的大问题.在这方面,从事反问题研究的数学家可以为企业家出谋划策.

事实上,最早的反问题研究就是起源于定向设计问题.我们知道,单摆的等时性只是在小角度的假设下才近似成立.能不能找到一种特殊轨线的摆,使它严格满足等时性?惠更斯于1673年提出并解决了这一问题,这种特殊的轨线就是旋轮线,它的方程为

$$\begin{cases} x = l(\varphi + \sin\varphi), \\ y = l(1 - \cos\varphi). \end{cases}$$

到了19世纪，挪威数学家Abel将惠更斯的问题推广为：测出了物体从不同高处落下的时间，如何反求物体下落的轨道？他于1823年给出了问题的解答.

当代工业产品的极大丰富为反问题的研究提供了广阔的用武之地，许多工业设计问题是相当困难的，需要用到高深的数学手段.例如，国外的光学仪器厂家提出：能否设计一种光栅，利用其非线性衍射效应产生出高能量的单色光射线？这就是一个定向设计问题，它要求数学家利用推导和计算手段构造出所需要的曲面(光栅)形状.

定向设计不限于产品，它的应用相当广泛.比如说：一个城市的某条街道车流量很大，不堪负荷，怎样通过铺设新的路段来进行分流？在军事行动中如何对不同种类的炮火进行分布以达到特定的轰炸效果？这类问题往往涉及各种事物的组合、分配、布局，要求在各种相互制约、相互影响的因素中寻找出最佳方案，为领导的决策提供依据.

4.4 物性探测

给你一只管子，不允许直接进入内部测量，你能算出里面的形状吗？如果管子是轴对称的，这时只需要知道内部的截面半径就可以了.美国贝尔电话实验室的Sondhi和Gophinath提供了一个方法：在管子的一边发出声音，用仪器测量管口的位移速度和压力.通过测量结果就可以推知管内的截面半径.理论计算与实验结

果吻合得很好.

不要小看了这个例子,它实际上暗示了许多不能直接测量的物性探测问题可以通过类似的间接方法来解决.通常说"上天入地"都是很困难的事情,可是在一些情况下似乎必须"入地"才能解决问题,比如说石油勘探.石油通常埋在几千米的地下,无法直接观察油田的位置和储量,靠试打井的办法来探测不但费用昂贵(一口井的代价要几千万元),而且效率极低(只能探测到井附近的局部信息).一个可行的办法是通过地面爆炸向地下发射地震波,同时接收地层的反射波信号.可以想象,地面接收到的反射信号中含有地下的物性结构信息(地层的密度、声速等等),利用数学手段将这些信息提取出来,就可以对地下的油储及其分布作出科学的判断.这很像在夏天人们挑西瓜,把瓜放在耳边拍一拍,有经验的人就知道瓜瓤熟不熟,不需要切开来看,不会破坏西瓜的完整.

类似的探测方法可以应用于许多方面,如:农用土壤分析、地下水勘查,甚至于在考古发现上也有应用.位于三峡库区的四川省云阳县故陵镇有一个大土包,相传为楚国古墓,但是历经3000余年的变迁,已经难以确认了.科技工作者在地表利用地震波法、高精度磁法、电场岩性探测和地化方法四种手段进行探测,不但确认了古墓的存在,而且得到了关于古墓的埋藏深度、形状、大小甚至墓道的准确信息,为抢救和保护文物作出了贡献.

4.5 扫描成像

在前面讲到的Abel反问题中,如果把下落的物体用扫描射线替代,从另一个角度来看,它为我们提供了从射线的走时响应反推

其传播轨迹的方法，将不同轨迹射线的反演结果组合起来就能得到传播介质的内部形态信息．本世纪初，Hebglotz 和 Wiechebt 应用 Abel 型反演方法解决了在一定对称条件下通过地震波的走时曲线来反推地层内部形貌的方法．据此 Mohobovic(1909 年)发现了地壳与地幔之间的断层．现在，利用地震波的接收信号通过成像来考察地层地貌形态已经成为地球物理勘探最为重要的手段．例如，通过走时成像，可以得到地震波在不同深度的传播速度；而在已知速度的前提下，利用声波方程或其单程波方程偏移成像方法，又可以得到反射界面的位置和形状．

扫描成像的另一个重要应用是医学上的计算机层析成像(CT)，这是 X 光射线自 Roentgen 发明(获 1900 年诺贝尔奖)以来在医疗诊断上的重大进展，其发明人 Hounsfield 和 Cormack 因此获得了 1979 年的诺贝尔医学奖．CT 技术是医学、电子技术、计算机技术和反演数学相结合的产物，它利用计算机来对穿越人体的 X 射线信号进行处理，重建体内的结构信息，生成透视图像供医疗诊断参考，其核心算法的数学基础是二维 Radon 变换．继之而起的是基于三维 Radon 变换的核磁共振成像，在诊断效果和无伤害性方面更为优越．事实上，类似的方法也可以借助于声波、光波、电磁波，在无损探伤、雷达侦察、射电望远镜探测、环境监测等多方面进行广泛的应用．

4.6 逆时反演及其他

在科学研究中，经常遇到这样的问题：知道了某个事物的现在状态，希望了解它的过去，即通常所说的“恢复历史的本来面目”．

这往往可以提为逆时反问题.当然,反问题研究不是历史学,它所研究的对象一般要满足某种类型的演化方程或数学模式.例如,通过远程测得的某次爆炸产生的辐射波,如何确定爆炸的位置和初始能量?这是波动方程的逆时反问题;又如,根据近来的温度变化能否确定过去某个时间的温度状态?这就成为热传导方程的逆时反问题.

前面介绍了反问题的几种类型,它们在研究和应用上经常是相互联系的,分门别类只是为了叙述方便.另外,反问题与其他数学学科之间并没有一个严格的界限,而是互为补充,互相促进.反问题的研究起源于数理方程,其反演算法中包含了微分方程数值解法、最优化方法和概率统计等方面的许多思想和技巧.另一方面,反问题的研究也促进了人们对世界的认识,使得研究更全面、深化.一个著名的例子是反散射方法在孤立子发现中的作用:反散射问题是量子物理学研究中的一个问题,通过谱和谱函数在无穷远处的散射性态反推一维 Schordinger 方程的位势函数.它由前苏联数学家 Gelfand 和 Levitan(1955 年)一举解决.在此基础上引发了一系列突破性进展,最为著名的是利用这个结果 Lax(1968 年)得到了关于 KDV 方程的巧妙解法,从而发现了非线性方程中的孤立子现象.这是近代非线性科学研究的重要事件.

4.7 反问题研究的难点及对策

与正问题相比,反问题的研究起步较晚,发展还远不成熟.从本质上来说,反问题的研究的难度一般比相应的正问题要大.这是因为反问题的求解往往违背了物理过程的自然顺序,从而使正问

题中的许多良好性质不再满足.这种现象在许多学科的研究中都是普遍存在的.比如说:曹雪芹创作了古典名著《红楼梦》,这是人所共知的,但是要从现存的史料和文物"碎片"来恢复这位伟大作家的人生经历和创作历程则是一件万分艰辛的事情,更何况这些"碎片"信息真伪交杂,且时有含混.反问题的研究也经常遇到类似的困难,这些困难体现在:

1.存在性

要求的反问题的解很可能不存在!无解的原因多种多样,可能是在定向设计中问题的提法不合理,也可能是探测时接收到的响应中含有假信息(噪音),将求解引入歧途.

2.唯一性

有的反问题的解虽然存在,却不唯一,有几个甚至无限多个.这是因为收集到的信息不够,不足以确定解的性态.对大多数反问题(比如探测问题)来说,真正的解只有一个,这就要从许多解当中进行挑选,去伪存真,颇费周折.

3.稳定性

利用计算手段,由接收信息来反演物质的结构和特性是反问题研究的重要内容.可是实际的接收响应中不可避免地含有噪音,计算过程也有误差累积.微小的误差会不会导致反演结果面目全非?研究表明:相当多的反问题正是具有这样的病态性质.热传导方程的逆时反问题就是一个例子.热力学第二定律告诉我们,热传导是一个不可逆过程,它的反问题求解是高度病态的.为了解决温度的逆时反演,就不得不冒这种"差之毫厘,谬以千里"的危险.

存在性、唯一性和稳定性,三者之一不满足就称为不适定性问题.用传统的眼光来看,这样的问题是不值得研究的.正是反问题

的研究开阔了人们的视野，认识到这样的问题是大量存在的，而且有着重要的研究和应用价值.

如果一个问题的解不存在、不唯一、不稳定，那么求解得到的结果可信吗？这是反演工作者必须面对的问题.解决的办法是有的！奠基性工作是由前苏联学者 Tikonov 等提出的解决线性不适定问题的正则化方法.方法的主要思想是：利用对解和数据误差的先验估计可以将问题的求解限定在某个较小范围内，对问题的提法进行适当的改造后，原本不适定的问题就可以转化为适定的最优化问题求解，而且先验估计表明在一定精度下用正则化方法求得的解是合理的.这比如猜谜语："后，打一人名"，无从猜起.如果限定"打《红楼梦》中一人名"，范围缩小了，可以用书中 601 个人物（有的书中没有交代姓名）逐一比较，最后选出最优的答案——"王夫人".

充分利用各种合理的先验信息对问题作适当形式的转换，是反问题求解的重要方法，在解决实际生产问题中经常要用到.拿地震波勘探为例，限于技术原因，地面接收的信号噪音很大，信息残缺不全，完全的反演是很困难的.为了满足生产的要求，必须尽最大可能恢复出地下的结构形态.这时，多种反演方法并用是一个可行的办法：如果在目的地有一口油井，那么可以把井下的信息作为局部约束来校正反演结果；为了计算的稳定性还必须使用一些特殊的数学技巧.这样得到的反演结果与资料解释人员的经验结合起来，可以对油田的决策与发展提供参考依据.

除了前面提到的不适定性以外，反问题的研究与应用还经常面临非线性的困扰.即使正问题是线性的，它的反问题也往往表现为非线性，这为反演的研究和计算带来了很多麻烦.为了求解非线

性反问题，通常要线性化后反复进行正、反演迭代，在高维情况下将导致十分巨大的计算量．我们知道，一个效率低下的算法在生产应用中将导致时间和人力、物力的极大浪费．所以反问题的计算效率也是一个非常重要的课题．它要求计算数学工作者从实际应用出发，充分研究问题的性质和特点，构造出精巧、快速的算法以适应生产的需要．

反演问题有着特殊的困难，它向我们提出了许多在认识论、方法论中富有挑战性的课题，深化了对客观现象的理解．反问题的研究确有它独立的价值．

4.8 发展和展望

反问题研究的兴起不过是近几十年的事情，它主要的研究对象是与探测、识别和设计有关的应用问题．实际生产的迫切需求是推动这一学科迅速发展的原动力．1987 年，以“反问题、反演方法和数据反演计算”为主要内容的专题杂志《Inverse Problems》创刊，标志着反问题的研究走向独立和成熟．目前国际 Internet 网上开设了有关反问题的专栏 IPNET；美国建立了工程反问题组织（AGIPE）；世界上每年都举行各种形式的反问题研讨会，得到了数学、物理、工程技术等多方面专家的响应．需要指出的是，在国外对反问题研究的资助不仅来自于科研和工业部门，还来自于国防部门．

我国的反问题研究自 80 年代初由冯康先生首倡，在实际问题的推动下，先后在中国科学院、哈尔滨工业大学、山东大学、中山大学、南京大学以及石油工业部门等许多单位取得相当丰富的理论

和实际应用成果.1991 年,中国科学院计算数学与科学工程计算研究所(原计算中心)承担了国家攀登项目"大规模科学与工程计算的方法和理论"中的反演课题——"波动方程的反演及其数值方法".这是在工程上有广泛应用价值的重要课题,但由于它的非线性和不适定性,从理论到计算都有着特殊的困难.经过多年的努力,我们已发展出了一套有特色的反演计算方法和理论:通过单程波分解的技巧,以一种自然的方式将问题适定化,算法的计算量小,并且得到了一些较好的反演结果.单程波方程及其数值方法在地震偏移成像、吸收边界条件等实用技术方面有广泛应用.目前用这种方法构造的高效率大倾角偏移格式和深度偏移的因子分解算法等已经形成了实用软件并应用到石油勘探部门.

近 40 年来计算技术的飞速发展大大增强了数学工作者在自然科学、社会科学和工程技术等广阔领域的参与能力,反问题正是在这样的背景下应运而生的交叉性学科.它的生命力源于实际应用的迫切需求和反演工作者卓有成效的工作.反问题的出现为传统数理方程的研究开辟了新的疆域,也为数学家参与实际生活提供了新的切入点.应该看到,反问题的开展程度与工业和国防的现代化、科学技术在产品中的含量有着密切的关系.我们期待着这一新兴学科在国内能够健康地发展起来,为国家的经济建设作出它应有的贡献.

5 分岐、混沌和湍流的数值计算

5.1 世界是非线性的

分岐、混沌和湍流是非线性问题的本质属性.线性方程和非线性方程之间的本质差别是很明显的.一个线性方程的任何两个解可以加在一起构成一个新解,这就是叠加原理.叠加原理是用来解决线性方程的关键.例如,Fourier 变换方法,Laplace 变换方法都与解能够叠加有关.顺其自然,人们把问题分成许多小问题,然后把分散的解叠加起来,得到整个问题的解.

与此相反,一个非线性方程的两个解不能加在一起构成另外一个解,叠加失效了.因此,人们必须整个地考虑非线性问题;然而,即使最简单的非线性方程,人们也是不能企望有分析解的.

在物理上,线性与非线性行为之间的区别最好是从一些例子中概括出来.例如当水流以低速通过

一根管道时，它的运动是层流的并且有线性行为的特征，即运动是规则的、可预测的，可用简单的分析数学的术语来描述．但是当速度超过某一临界值时，运动就变成了湍流，其中具有局部涡，这些涡以具有非线性行为特征，以复杂、不规则且反复无常的方式运动着．另一个例子就是非线性单摆运动

$$\frac{d^2\theta}{dt^2} + \frac{g}{l}\sin\theta = 0,$$

其中 θ 是摆离开铅直线的角位移，l 是臂长，g 是重力加速度，上述方程显然是非线性的．如果 θ 很小，即运动只限于微小摆动，那么可以线性化为

$$\frac{d^2\theta}{dt^2} + \frac{g}{l}\theta = 0,$$

因而有解

$$\theta(t) = \frac{1}{\omega}\left(\frac{d\theta}{dt}\right)_0 \sin\omega t + \theta_0\cos\omega t.$$

式中 θ_0，$\left(\frac{d\theta}{dt}\right)_0$ 是初始时刻的角度和角速度，$\omega = \sqrt{g/l}$．显然，这个解是由两个解叠加而成的，它表明 ω 与振幅无关．这是线性的近似；当摆动是大的时候，可以得到非线性摆的一个积分

$$\frac{d\theta}{dt} = \sqrt{\frac{g}{l}}\cos\theta + c,$$

运动的周期

$$T = 4\int_0^{\theta_{max}} \left(\frac{2g}{l}(\cos\theta - \cos\theta_{max})\right)^{-1/2} d\theta,$$

在相平面 $\left(\theta, \frac{d\theta}{dt}\right)$ 上，$(0,0)$ 代表单摆最下方位能最小的点，它对应于稳定的中心点，而 $(0, \pm\sqrt{\pi})$ 代表上方位能最大的点，对应于不稳定的鞍点．

从木星的红斑通过湍流等离子体中的电磁辐射块区，到原子尺度的微观电荷密度波，空间局部化的、长寿命的波状激发，都充满着非线性系统。这些非线性波和结构反映了复杂行为之中存在着惊人的有序性。巨型红斑是17世纪后期首次观察到的，在木星大气的湍流喷火口中显然保持稳定，大约在4×10^{8}m的尺度上，即大致从地球到月亮之间的距离这样一个尺度上的一种特殊的有序结构，后来从"旅行者号"宇宙飞船上拍摄的照片中可以清楚地看到。

非线性现象包含这种有序结构的例子也可以从地面上某种运动看出，例如非线性海洋波形成的一种有序结构，它在几千海里尺度上传播没有本质的变化。从阿波罗-联盟号宇宙飞船上拍摄的照片上可以看到，在北苏门答腊附近安达曼海的公海范围上形成五条几乎成直线的表面波的波包，每条大约150km宽，每条相隔大约10km，以大约每秒2m的速度沿垂直于它们波峰的方向运动，这种波是由潮力和由于在海洋内部存在热梯度和盐梯度导致水面下的内波所激发的。

人们用一种称为隧道电子显微镜的成像过程来研究二硫化钽，可以发现六边形排列的拟序结构。这种电荷密度波相距大约正常晶格间距的3.5倍，为晶格中电子和原子核之间相互的非线性耦合的结果，这就是在10^{-9}m尺度上所显示的拟序结构。

自然界中的非线性现象，从10^{8}到10^{-9}m尺度范围内存在着一种拟序结构。纤维光学、Josephson传输线、等离子体腔子、流体中的旋涡、化学反应波非线性扩散阵面、金属中的位错、气泡和液滴等等。它们也都是由非线性微分方程所描述，在物理数学上，都表现为一种非线性动力学行为。

非线性动力系统，包括代数的、常微和偏微的或者它们耦合的非线性系统，即使是简单的，也具有极为复杂的动力学行为．非线性动力学的研究，成为物理学、力学和数学等学科中研究的主流之一．1923 年，G. I. Taylor 就从实验和理论上，分析了两个同心无限长的，以不同的角速度旋转的圆柱体之间粘性流体流动稳定性问题．当 Reynolds 数值低时，流动是定常的层流称为 Couette 流，当 Reynolds 数值增长到一定的数值时，Couette 流变为不稳定的，出现了有旋涡的流动，形成了绕旋转轴的环形涡流，再增加 Reynold 数，轴对称的涡流也变为不稳定，出现了一个地平经度的行进波，这种可以观察到的分歧现象，引起了很多数学家和物理学家的极大兴趣．1963 年，气象学家 Lorenz, E. N. 在数值实验中发现后来称为吸引子的混沌现象；1971 年，Ruelle, D. 和 F. Takens，对耗散动力系统引入了奇怪吸引子概念，建议用于描述湍流发生的新的机制，随后 May 的综述文章指出，生态学中一些非常简单的动力学数学模型，具有极为复杂的动力学行为；1975 年，Li 和 York 在一篇《周期 3 意味着混沌》的文章中，正式提出混沌这个概念；随后 Feigenbaum 和 Coullet 独立地发现了倍周期分歧现象中标度性和普适常数，揭示了混沌现象中也存在确定性规律；这些研究方向迅速融成一片，引起了众多物理学家和数学工作者的关注．

应该说，促进这一股研究热潮的基本动力之一，是寄希望于打开研究湍流的新门户．一百多年以来，湍流运动规律的研究一直没有取得根本性的突破，它至今仍是物理学领域内最为困难的一个基础理论问题．由于它具有广泛而重要的应用价值，又是自然界的所有非线性现象中一个典型代表，又由于自 60 年代以来，在非线性科学中相继发现的孤立子，拟序结构，确定性混沌，奇怪吸引子

以及分形结构等基本规律，极大地刺激了湍流的研究．科学家们逐渐认识到分歧、分形和混沌的研究最终将为了解湍流运动的本质和产生机制敞开了大门．1983 年，瑞典诺贝尔奖金学术委员会和美国科学、工程和公共政策委员会都对湍流和混沌现象之间的关系作过深入细致的探讨并广泛征求过意见，给出了比较乐观的估计．与低维动力学研究取得显著成就的同时，无限维动力学长时间行为研究也方兴未艾，尤其对 Navier－Stokes 方程、Kuramoto－Sivashinsky 方程、Cahn－Hiliard 方程、Ginzburg－Landan 方程、非线性 Schrodinger 方程、非线性反应扩散方程等吸引子及其分维数的估计，以及由此引出的惯性流形、近似惯性流形等新概念和建立于此基础上的新算法等等．

5.2 简单分歧例子

一个动力系统的解在某一范围内连续依赖于一参数，当参数变化超过临界值时，解出现突变现象，失去稳定性，破坏唯一性，出现几个分支，产生所谓分歧．当参数继续增大时，上述现象可能在新的水平上重复，产生次级分歧，次级分歧可能不断升级乃至无穷．分歧是非线性系统的本质属性，线性系统的解要么是零要么形成子空间，不会出现分歧．

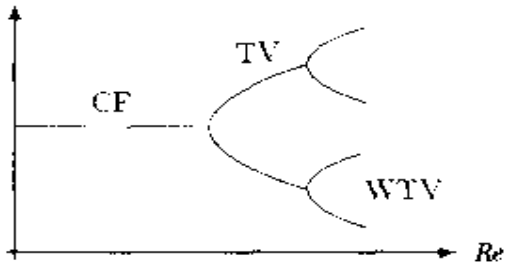
分歧点和里亚普诺夫意义下的稳定性息息相关．

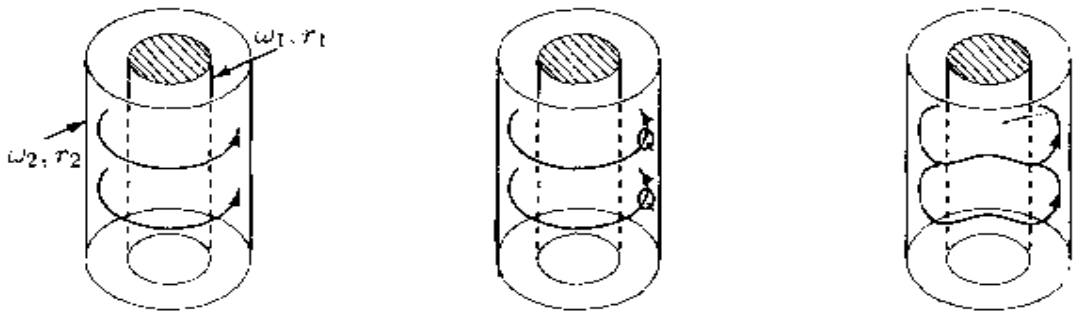
下面给出一个典型的无限维非线性系统的分歧的例子．

例 两个无限长同向旋转圆柱之间的 Taylor 问题

1923 年，G. I. Taylor 研究了两个无限长同轴以不同角速度旋转的圆柱之间粘性流的流动问题，发现了分歧现象．后来很多学者

研究了这个问题.实验、数值计算和理论分析表明,当 Reynolds 数 Re 很小时,两圆柱之间流体流动是层流,只有圆周方向的运动,粒子运动轨迹是一个同心圆,称之为 Couette 流(图 5－1(a)).当 Reynolds 数增加越过临界值 Rec(假设外圆柱不动,内圆柱以 ω_i 旋转,Reynolds 数增加就意味着 ω_i 增加),Couette 流出现不稳定,分岔出 Taylor 旋涡,即在通过轴的子午面内,沿 z 轴方向出现周期性旋涡,由一个一个旋涡组成,关于 $z=0$ 成镜面反射对称.Taylor 旋涡是环形涡,仍然是定常流动,它是稳定的(图 5－1(b)).





(a)Couette 流(CF)　(b)Taylor 旋涡(TV)　(c)Taylor 行进波(WTV)

图 5－1　两个无限长同向旋转圆柱之间的 Taylor 问题

当 Re 继续增大,越过第二临界值 $Rec2$ 时,Taylor 旋涡变为不稳定,形成 Taylor 行进波,是一种沿旋转轴均匀运动的波动.破坏了对时间和旋转轴的不变性,但仍是一种周期性运动,并且在一个适

当的旋转标架里，流动看来还是定常的，这样的周期运动，也称为旋转波(图 5－1(c))．

当 Re 继续增大时，第二次转变发生，流动变成拟周期，它的次频率作为调制旋转波，再经过若干阶段，进入湍流．

Navier－Stokes 方程在 Couette 流的第一临界值处，其导算子方程有零第一特征值，重数为 1，分岔出 TV 分支流；第二临界值处，在 TV 流上的导算子方程有一对共轭虚数，出现 Hopf 分歧．

如果外圆柱体旋转角速度 $\omega_0 \neq 0$，那么当 ω_i，ω_0 是同向旋转时，分歧图形和 $\omega_0 = 0$ 时相同，当 ω_0，ω_i 是异向时，情形要复杂得多．实验结果表明：存在一个临界值 ω_0^*，当 $|\omega_0| \leqslant \omega_0^*$ 时，分歧图像与同向相似，而当 $\omega_0 > \omega_0^*$ 时，则分歧途径为

$$\text{Coutte 流} \rightarrow \text{螺旋蜂涡} \rightarrow \text{波动螺旋蜂涡} \rightarrow \cdots.$$

Diproma 和 Grannich，Krueger，Cross 和 Diprima 的研究表明，反向旋转时，Couette 流失稳是由于有几个特征值穿过虚轴，这几个特征值是：二重零特征值和 2 个二重的一对虚共轭特征值．

5.3 分歧和稳定性

先看简单的平面自治系统

$$\begin{cases} \dot{x} = f(x, y); \\ \dot{y} = g(x, y), \end{cases} \tag{5.1}$$

设 (x^*, y^*) 为该系统的平衡解

$$f(x^*, y^*) = 0, \quad g(x^*, y^*) = 0,$$

考察在 (x^*, y^*) 处的线性化方程

$$\begin{bmatrix}\dot{x}\\ \dot{y}\end{bmatrix} = \frac{\partial(f,g)}{\partial(x,y)}(x^*,y^*)\begin{bmatrix}x\\ y\end{bmatrix}, \tag{5.2}$$

它有如下形式的解 $Ae^{\lambda t}$，其中 λ 为雅可比矩阵 $\frac{\partial(f,g)}{\partial(x,y)}(x^*,y^*)$ 的特征值，满足

$$\lambda^2 + p\lambda + q = 0 .$$

在参数平面 (p,q) 上，λ 的分布如图 5－2.

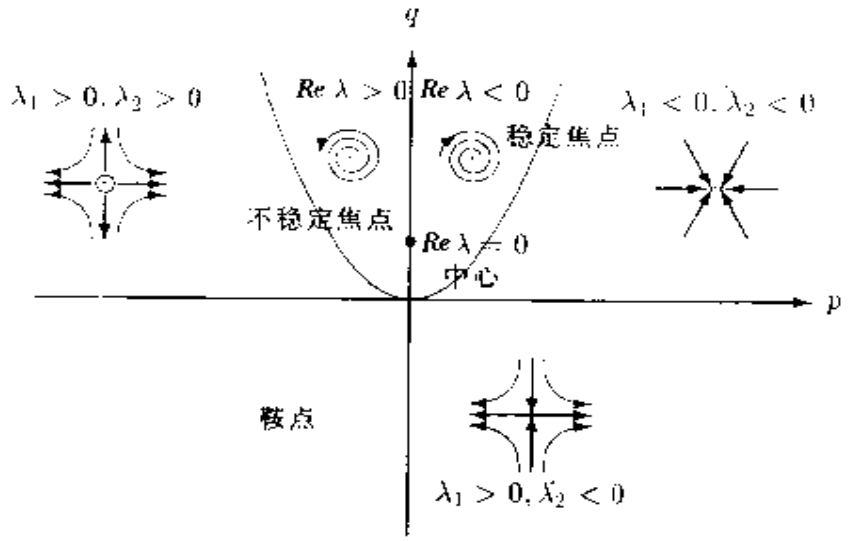


图 5　2　λ 在参数平面 (p,q) 上的分布图

(1) 双曲情况，此时 $Re\lambda \leqslant 0$

(A)$\lambda \neq 0$ 的实根. 线性化方程的解要么指数增长，包括不稳点结点，不稳定鞍点；要么指数衰减，包括稳点结点；

(B) 共轭复根. (5.2) 式的解为增长振荡，不稳定焦点；衰减振荡，稳定焦点.

这时非线性问题 (5.1) 式与线性化方程在平衡解 (u^*,y^*)

邻域和(5.2)式具有相同的拓扑结构.

(2) 中心点.$Re\lambda = 0$,线性方程方程的解围绕中心点振荡.

非线性方程(5.1)式的平衡解可能是中心点,也可能是焦点.

系统若是保守的,则只有中心点和鞍点;系统若是耗散的,则没有中心点,只有鞍点、结点和焦点.

(3) 结构稳定系统:若系统受到扰动,不改变相空间的拓扑结构,称为结构稳定性.

显然,中心是结构不稳定的;两个鞍点之间的连线也结构不稳定.

耗散系统中长时间状态的归宿,称为吸引子.

稳定的平衡解叫做定常吸引子.

除此以外,还有周期吸引子(如稳定极限环)、拟周期吸引子和混沌吸引子.

我们还需引入庞加莱截面:设 Σ 为系统轨道横截方向的一个平面,称为庞加莱截面,从 Σ 上任一点 x_1 出发,下一次在 Σ 上交于 x_2,我们说由 x_1 映到 x_2,同样由 x_2 映到 x_3,依此类推. 在 Σ 上,建立了一个映像关系:

$$x_{n+1} = f(x_n),$$

显然,当运动是周期时,则 $x_1 = x_2 = \cdots = p$ 就是一个点,它就是上面迭代式的不动点

$$x^* = f(x^*),$$

映照 f 的雅可比矩阵 $\frac{\partial f}{\partial x}(x^*)$ 的特征值 λ 称为 Floquet 乘子,简称乘子.

设 P 为双曲点,$W^u(P)$,$W^s(P)$ 分别是乘子 $|\lambda| > 1$ 和 $\lambda <$

1 的特征向量所构成的空间.

中心流形 $W^c(p)$ 等于雅可比特征值 $Re\lambda = 0$ 和乘子 $\lambda = 1$ 的特征向量所构成的子空间. 如果对同一个双曲点 P, $W^s(P) = W^u(P)$, 则当 $t \to \pm\infty$ 时, 所有轨道趋于一个点. 这个流形称为同宿轨道(homoclinic orbit); 相反, 设 P, Q 均为双曲点, 若 $W^s(P) = W^u(Q)$, 则 $t \to +\infty$ 和 $t \to -\infty$ 时, 轨道趋于不同的点. 这种流形称为异宿轨道(heteroclinic orbit). $W^s(P)$ 与 $W^u(P)$ 相交于一点时, 交点称同宿点; $W^s(P)$ 和 $W^u(Q)$ 的相交点称为异宿点.

庞加莱认为, 如果有一个同宿点或异宿点, 那么就有无穷多个同宿点或异宿点.

双曲点邻域内稳定流形和不稳定流形的演化反映了映射的伸长折叠性质. 斯梅尔已经证明有无穷多个同(异)宿点有马蹄(Horseshoe), 因而就有混沌.

5.4 分歧问题的数值方法

分歧图形可以用电子计算机通过数值模拟而显示出来. 分歧点的确定以及分歧点邻域内分支解的求解, 是分歧数值分析的主要任务. 现在, 分歧的全局分析和全局数值计算也有了飞快的发展.

在分歧点, 由于分歧方程导算子在这一点奇异, 一般牛顿算法或它的改进方法均已失效. 现代的连续弧长算法, 在正则点上, 可以有效地实现, 而碰到分歧点时, 也同样需要特殊处理. 数值求解分歧点及分支解的方法有两大类, 一类是直接方法, 另一类是间接方法. 直接方法是不需计算系统其他解而只计算分歧点, 间接方法

是通过计算奇异点附近的解来确定奇异点.

所谓扩充系统方法,是一种较为成功的直接方法.其主要思想首先由 H.B.Keller 提出的.其本质想法是将方程规模扩大,使其扩大后的新系统,在原奇异点上变为正则点,从而可以使用迄今为止一切行之有效的算法.那么,构造一个扩充系统便是数值求解分歧点首先必须解决的问题.

扩充系统方法的优点是将原问题奇异点转化为扩充系统的正则点.缺点是由于系统规模增大,计算机工作量也急剧增大.通常原问题若是 n 维的,扩充系统则是 $2n$,$3n$,…或 kn 阶.按照线性代数方程组 LU 分解,方程维数增加一倍,计算工作量增加约 7 倍;维数增加 2 倍,工作量增加约 26 倍.然而通常运用扩充系统求解时,不但可以求出分歧点,而且可得到分歧点诸多的信息,如特征向量等等.

5.4.1 非退化转向点

简单极限点,也称为转向点,非退化极限点称为简单转向点,退化的,称为非简单转向点.对于简单转向点,简单的扩充系统是

$$S(u)=\begin{pmatrix} F(x,\lambda) \\ D_xF(x,\lambda)\phi \\ \langle l,x\rangle-1 \end{pmatrix}=0,$$

其中 $u=\{x,\phi,\lambda\}\in X\times X\times R\triangleq Z$,$l\in Y$.那么可以证明,若 (x_0,λ_C) 为 $F=0$ 的简单转向点,则 $u_0=\{x_0,\phi_0,\lambda_0\}$ 为正则点,其中 ϕ_0 为 DF_0 对应于零特征值的特征向量.反之亦然.

实际上,如果 s 作为通过转向点分支解的弧长,那么系统

$$\begin{cases} F(x(s),\lambda(s)) = 0, \\ \dot{x}(s)^2 + \dot{\lambda}(s) = 0 \end{cases}$$

在转向点处，即为正则点.

5.4.2 简单分歧点

设新变量 $u = \{x, \lambda, y_1, y_2, y_3\}$，扩充系统

$$S(u) = \begin{bmatrix} F(x,\lambda) + \langle y_2, D_xFy_1\rangle y_2 \\ \langle y_2, D_xFy_3 + D_\lambda F\rangle \\ D_xFy_1 + [\langle y_2, y_1\rangle - 1]y_2/2 \\ D_xF^* y_2 + [\langle y_2, y_2\rangle - 1]y_1/2 \\ D_xFy_3 + D_\lambda F + \langle y_2, y_3\rangle y_2 \end{bmatrix},$$

当 $F \in C^3$ 时，$S(u)$ 有意义且 C^2 连续.

可以证明，当 x_0, λ_0 是简单分歧点时，则

$$u_0 = \{x_0, \lambda_0, \varphi, \psi, v_0\}$$

是 $S(u)$ 的正则点，其中

$$\mathrm{Ker}(D_xF_0) = \mathrm{Span}\{\phi\},\ \mathrm{Ker}(D_xF^*) = \mathrm{Span}\{\psi\},$$

$$v_0 \in \mathrm{Range}(D_xF_0),\ D_\lambda F_0 + D_xF_0 v_0 = 0.$$

扩充系统的优点，是它在 u_0 的导数算子具有形式

$$DS(u_0) = \begin{bmatrix} A(u_0) & 0 \\ D(D_xF_0\phi)\cdot & D_xF_0 + \langle\phi,\cdot\rangle\psi \\ D(D_xF_0^*\psi)\cdot & 0 \qquad D_xF_0^*\phi\langle\psi,\cdot\rangle \\ D^2F_0\phi(v_0,1)\cdot & \psi\langle v_0,\cdot\rangle \qquad 0 \quad D_xF_0\psi\langle\phi,\cdot\rangle \end{bmatrix}$$

其中 A 和 $D_xF_0 + \psi\langle\phi,\cdot\rangle$ 在 $X \times R, X$ 上分别可逆，并且 $A(u_0)$ 的逆可以求出来. 运用牛顿方法求解，可以大大减少求解

规模，许多作者都设计各种类似的扩充系统，其求解规模都比原系统大得多.

5.4.3 多重奇异点和 Hopf 分歧点

对于余维数为 2 的分歧点、极限点、对称破缺的分歧点，均可以构造可分裂迭代的扩充系统.

为了求 Hopf 分歧点，也可以构造很多类型的扩充系统，最简单的一种是

$$S(y)=\begin{vmatrix} F(u,\lambda) \\ [D_uF(u,\lambda)^2+\omega^2 I]p \\ (p,p)-1 \\ (q,p) \end{vmatrix}=0,$$

其中求解参数为 $y=(u,p,\lambda,\omega)$，D_uF 的特征值

$$\mu(\lambda):\mu(\lambda)=\sigma(\lambda)+i\omega(\lambda),\mu(0)=i\omega_0,$$

对应的特征向量为 $\phi=\phi_1+i\phi_2$，q 是选定向量，但 q 在 Span$\{\phi_1,\phi_2\}$ 上的投影不为零，当

$$\frac{\partial u}{\partial t}=F(u,\lambda)$$

满足 Hopf 分歧点(u_0,λ_0)的存在条件时，则必存在 $p_0\in$ Span$\{\phi_1,\phi_2\}$，使得 $y_0=(u_0,p_0,\lambda_0,\omega_0)$ 是扩充系统的正则点，而且可以证明，它是 $S(y)=0$ 的正则点.

5.4.4 连续算法

连续算法，也称为预估校正方法、连续弧长算法和路径跟踪算法等，它的任务是数值求解通过奇异点邻域各个分支解，因此

需要解决奇异点判别、解分支方向确定和跟踪、步长选取等问题. 目前有许多软件可供使用，如 Auto(Doeddel)，Pitcon(Rheinboldt) 及 E. L. Allgower 和 Kart Georg 的书，书末附有各种程序.

为简单起见，我们仅仅阐明有限维问题连续算法的基本思想.

设 $F(u,\lambda): R^n \times R \to R^n$ 的光滑映照，$x_0 = (u_0, \lambda_0)$ 为它的零点，满足

$$F(x_0) \equiv F(u_0, \lambda_0) = 0, \ \mathrm{rank}DF(u_0, \lambda_0) = n,$$

由隐函数存在定理，存在一个开区间 J 和唯一的一条光滑曲线性 $x(s) \in R^{n+1}$ 使得

$$F(x(s)) = 0, \mathrm{rank}DF(x(s)) = n, \ \dot{x}(s) \neq 0,$$

其中 $\dot{x}$ 表示关于 s 的导数，这时有很多方法可以跟踪这条曲线，如设 $D_u F_0$ 非奇异，则

$$D_u F(x(\lambda))u'(\lambda) + D_\lambda F(x(\lambda)) = 0,$$

由此可解出 $u'(\lambda_0)$，取初值

$$u_0 = u(\lambda_0) + \delta u(\lambda_0),$$

应用牛顿迭代

$$u^{k+1} = u^k - D_u F(u^k, \lambda_k)^{-1} F(u^k, \lambda_k),$$

其中 λ_k, u^k 为已知，$\lambda_{k+1} = \lambda_k + \delta$，这个过程可以一直循环下去，得出一条解曲线；如果 $D_u F_0$ 奇异，但是 $D_\lambda F_0 \notin \mathrm{Range}(D_u F_0)$，那么可以令

$$S(x,s) = \begin{bmatrix} F(x) \\ N(x,s) \end{bmatrix} = 0,$$

$$x = (u, \lambda),$$

$$N(x,s) = \|\dot{x}\|^2 + |\dot{\lambda}|^2,$$

于是可以证明 $D_x S_0$ 非奇异，对 $S(x,s) = 0$，可以用上述方法求出

曲线来，如已知 $x_i = (u_i, \lambda_i)$，则求切线方向：

$$D_\lambda F_i \dot{\lambda}(s_i) + D_u F_i \dot{u}(s_i) = 0 \quad (F_i \text{ 表示 } F \text{ 在 } s_i \text{ 取值}),$$

预估：取 $S_i = s - s_i$，

$$x^0(s) = x(s_i) + S_i \dot{x}(s_i),$$

校正

$$D_x S(x^k(s), s)(x^{k+1}(s) - x^k(s)) = -S(x^k(s), s)$$

$k = 0, 1, 2, \cdots$ 迭代，可得 $x(s_{i+1}) = x(s_i + \delta_i)$ 的近似值

对于分歧点情形，我们首先得辨别奇点类型．设分歧点是简单的，即 $\dim N(D_u F_0) = 1$，那么令

$$\sigma(s) = \det \begin{pmatrix} D_u F(s) & D_\lambda F(s) \\ \dot{u}^T(s) & \dot{\lambda}(s) \end{pmatrix},$$

则当通过 (λ_0, u_0)（即 $s = s_0$）时，如果 $\sigma(s)\det(D_u F(s)) > 0$，则 (u_0, λ_0) 为简单的非退化分歧点；如果 $\sigma(s)\det(D_u F(s)) < 0$，则 (u_0, λ_0) 为简单极限点．

得到奇异点之后，必须也求通过奇异点各个解分支的方向，求解代数方程就可以得到解分支方向．如，设 y_i 为代数分歧方程的解，那么

$$\dot{x}(s_0) = \sum_{i=1}^{m+1} y_i \phi_i,$$

有了切线方向，就可以跟踪相应的解分支．

5.5 混沌和湍流

非线性动力系统按其特性可分为两大类．其一是耗散系统，它的解在相空间内的相体积在时间演化过程中收缩到零，收缩到低

维空间.这种系统存在吸引子,相空间中所有运动轨道最终都要被吸引到这种吸引子上.相反,另一类是保守系统,它的解在相空间中始终保持体积不变,这种系统不存在吸引子,任何轨道都不被吸引,因此只有椭圆和双曲两类奇点.

前面我们讨论是随着系统控制参数的变化,系统物理量发生质的变化的规律,即分歧和突变理论,它与流体流动的转换问题有关,而动力系统的物理量的长时间的演化规律,即渐近行为,涉及到吸引子、分形和混沌.

湍流的流场结构一直是人们特别关注的一个问题.事实上,在混沌理论中占有重要地位的奇怪吸引子和分形等新概念就是在研究湍流特性中提出的.系统中运动轨道处于一种不规则的非周期的混沌运动姿态就是由于某种具有奇特结构的奇怪吸引子作用的结果.简单地说,奇怪吸引子就是在相空间的一个有界区域内,由无穷多个不稳定的点集组成的一个集合体.它实际上是由系统中存在的无穷多个双曲不动点(包括周期轨道)的所有不稳定流形的闭包组成,即由所有不稳定流形总和及它们的邻域组成.这里不稳定流形构成一个闭环.这个点集吸引所有在它附近而不属于它的运动轨道并且作为一个整体,由其中任何一点出发的运动轨道,只要时间充分大,总能无限接近这个奇怪吸引子所有其他点.

奇怪吸引子还具有如下特点:

对初始条件有非常敏感的依赖性,在初始时刻从这个奇怪吸引子上任何两个非常接近的点出发的两条运动轨道,最终必然会以指数的形式相互分离.如果定义里亚普诺夫特征指数

$$\lambda_i = \lim_{t\to\infty} \log \frac{l_i(t)}{l_i(0)},\quad i = 1,2,\cdots,n,$$

其中 $l_i(0)$, $l_i(t)$ 分别对应于初始时刻和 t 时刻经过初始相邻近两点的运动轨道沿第 i 个特征方向之间的距离，n 是相空间的维数.那么奇怪吸引子的里亚普诺夫特征指数至少有一个要大于零，也必然有小于零的里亚普诺夫指数；

系统被激发的特征频率有无穷多个；

系统在运动过程中存在拉伸和折叠现象，即所谓马蹄现象.这就意味着存在双曲不动点，存在着不稳定流形；

奇怪吸引子具有非常奇特的拓扑结构和几何形式.它是具有无穷多层次自相似结构的维数是非整数的一个集合体.它在某些维数方向上是连续的，而在另外一些维数方向上则具有类似康托集那样的结构，沿着不稳定流形的方向是连续且具有指数发散性；而且在垂直不稳定流形方向上存在类似康托集那样的结构.

为了描述奇怪吸引子的这种奇特的结构.Mandelbrot 引进了分形(维数为非整数的对象)的概念.分形物体没有特征长度，但却有自相似性.分形物体维数，可以用分维数 d_f 来计算.如 n 维数空间中一个奇怪吸引子，$\{x_i\}_i^N{}_n$ 是它的一组点集，用边长为 a 的 n 维小单元体全部复盖住这个点集，所需最小数为 N，则

$$d_f = \lim_{a\to 0} \frac{\ln N}{\ln(1/a)},$$

分形物体除了它的维数不是整数外，还有一个极其重要的特征是自相似自嵌套结构，即某些局部区域的精细结构和它的总体结构具有相似的形态.因为分形物体不是完全不规则的.湍流运动中观察到的大大小小不同尺度的涡旋结构，与分形物体有很多相似之处.因而，研究湍流的几何和拓扑结构，是应该特别强调的.

实际上，湍流的耗散区域，即湍流中涡量高度集中的区域是一

种间歇状的分形结构，其主要特征是局部自相似性。湍流中小尺度结构是一种涡转或细涡管的集合体。这是因为流体在高雷诺数运动时，流体受到一种拉伸的作用，因而涡受到拉伸作用而变形。在三维空间的湍流流动，即某一初始时刻集中在一起的一团高涡量区很快被拉伸成涡管（两个方向上压缩，一个方向上拉伸）或涡面（两个方向上拉伸，一个方向上压缩）。最新的计算表明，高涡量区主要演变成管状或绳状的形式，低涡管或涡面在不断拉长的过程中，由于周围流体的作用而被折叠变形。根据 Helmholtz 定理，在粘性作用很小的尺度范围上，涡管不能被破坏，因此涡管必定是自回避地反复折叠着，它们在空间分布必定不均匀，从而形成间歇状的结构。但是涡管不能被无限制地拉伸。当它细到一定程度时，粘性必然要起作用，这时涡管就会发生瞬时破裂和重新联结的一种转换过程。这种瞬时转换过程对应的就是所谓局部猝发现象，是一种分歧现象。

6 计算流体力学

6.1 流体力学问题

计算流体力学是一门研究如何在高速电子计算机上求解流体力学方程组的科学.一说到流体,人们往往认为就是液体状态的水,其实气体也是流体,在高温高压下甚至金属也可看成是流体.事实上流体在这里指的是一切可看成是连续介质的物质,也就是说该物质是由无数微团连续组成的,物质的密度、压力、速度等宏观量就是在每个微团内进行统计平均而得出的,并代表该微团所具有的值.

世界上许多物质运动的现象,这些现象中一些特征量的变化都可以用流体力学方程组来描述.不尽长江滚滚来,长江大桥桥墩周围水流的变化,附近旋涡的形成,就可以用流体力学方程来描述.在设计大桥桥墩的时候,它的外形、强度等等就都要根据流体力学运动的参数来确定.天气的变化,刮风下雨这

些人们常见的自然现象，实际上是空气流动的结果，而空气流动也可以用流体力学方程来描述．气象工作者根据今天世界各地的气象实况，通过求解流体力学方程组得到明天、后天或半个月后空气流动的变化来做 24 小时、48 小时或更长时间的天气预报．飞机和各种飞行器的设计很关键的是它们在飞行过程中周围气流的变化参数，现在人们已经可以用解流体力学方程组的办法来得到这些参数．战争中在面对敌方坦克进攻的时候，往往用穿甲弹去打击敌人的坦克，穿甲弹是用炸药爆轰挤压一个锥形的药型罩，形成一个能量集中高速运动的射流去击穿坦克的厚可达 1m 左右的防护板，药型罩是用重金属材料制成的．在由炸药爆轰形成的高温高压条件下，金属也是一种连续介质，它的运动也可以用流体力学方程组来描述．因此解流体力学方程就可以再现药型罩被挤垮、形成射流并穿透防护板的全过程．在国民经济建设和国防建设的许多领域，还有很多问题都可以用流体力学方程组来描述．

流体力学方程组是一组非线性的偏微分方程．除了极个别的特殊情况外，这组方程的解是无法用初等函数来表达的，因此在现代高速电子计算机出现以前，科学家们是根据具体问题，用深刻的物理概念和高超的数学技巧，将方程简化到可以用最简单的计算工具，例如算尺或手摇计算机进行计算的程度，然后用数值方法求解．有时，这些简化方程的解可以用初等函数表达，但是这种简化方程的解只能定性地或局部地反映真实的流体运动的图像．要想直接求解流体力学方程组，只有求助于更先进的计算工具．现代电子计算机可以说是应数值求解流体力学方程组的需要而诞生的．第二次世界大战期间，科学家们在进行原子弹的理论设计时需要求解流体力学方程组，而当时一般的计算工具完成不了这个任务．

von Neumann 便提出设计了电子计算机.

流体力学方程组是根据质量守恒、动量守恒和能量守恒的规律建立起来的.力学量密度、速度、压力、能量变化所满足的方程是由它们对时间和空间的偏导数之间的关系来描述的.导数是一个极限运算过程,极限过程是一个无限的过程.而人们生活在一个有限的时间和空间的范围内,因而人们对于极限过程只能从思维上把握它,而在实际操作上只能用有限去逼近无限.空间的任何一个区域都包含有无限多个点,用数值的方法绝无可能根据方程求出区域内每一个点上的力学量的变化情况.这就有一个用有限个点上力学量之间的关系去代替原来的偏微分方程,用有限的代数运算去代替无限的微分运算的问题.最简单的譬如说用一个力学量在两点上的差被这两点的距离去除来代替空间的微分,这个做法称为将微分方程离散化.这就是将一组在一个区域上的偏微分方程离散化为一组在这个区域中若干个点上的代数方程组.在电子计算机上求解的流体力学方程组实际上是这一个离散化了的代数方程组,因此计算流体力学也就是研究如何将流体力学方程组离散化的科学.

离散点的选取往往是流体力学计算的第一个步骤,也就是将求解的区域划分为网格.表征网格大小的是网格的步长,即同一方向上相邻两个网格点之间的距离.网格的大小要看研究的对象而定.研究地球气候的变化,要将整个地球划分为网格,它的步长往往是几 km.如果要研究前面说的破甲弹,由于它的药型罩的厚度只有几个 mm,因而它的步长最多也只能是 mm 量级的了.人类是生活在三维空间中的,在取定了坐标系以后,空间的任一点可以用三个数(坐标)来确定.一个三维区域划分为网格,其最后所得的网

格数往往是很可观的.例如将一个 1L 的正立方体在长宽高三个方向上均匀地将网格步长取成 1cm,则总的网格数就是 1 千.如果将网格步长缩小一个量级,取成 0.1cm,则总网格数就提高三个量级,增加为 100 万.网格数越多,计算量越大.对计算机速度和内存的要求就越高.因此在巨型机出现以前,人们很难进行三维问题的计算.事实上有不少实际问题并不是三维的.例如计算区域是一个球体,而在每一个球面上的物理量的状态都是一样的,只要求出任意一条半径上有限个点处的物理量的状态就够了.这样的问题称为是一维球对称的.网格的划分只要在一个一维直线段上进行就可以了.这样网格数就大为减少.也有二维柱对称的问题,那就是在以一根轴线上任一点为圆心的一个圆周上的状态都是相同的.这时只要在一个平面区域上划分网格就可以了,而实际上的求解区域则是这个平面区域绕对称轴线旋转而成的旋转体.如何划分网格,这也是计算流体力学研究的课题之一.一般来说,在力学量变化缓慢的区域上,网格可划分得粗一些;而在变化剧烈的区域上,网格应划分得细一点.在二维问题的计算中,有的求解区域几何形状很复杂,例如某些飞行器的外形,这时就需要将物理平面用数学变换的方法变换到一个规整的矩形区域上,再划分网格进行求解.

流体力学方程由于其为非线性,所以解的状态往往带有一些奇性,其中最困难的问题之一是它的解在很多情况下会出现间断,即出现力学上称之为冲激波、接触间断、切向间断等等的现象.如何用数值方法去逼近这种间断解是计算流体力学一开始就面临的一个难题.作为计算流体力学创建者的 von Neumann 的杰出贡献就在于他在流体力学方程中引入了人为粘性,于是将间断解(冲激

波)转化为在一个很窄的区域上剧烈变化的连续解,然后用数值方法去逼近这个连续解作为一种近似.半个世纪以来,从事计算流体力学的人们,构造了多种多样的计算方法,除在个别领域外(例如地球流体力学的计算),所有这些方法都不能回避如何计算间断解的问题,都是从不同角度出发,按照不同的观点,构造不同的人为粘性.

计算流体力学的内容十分丰富,下面结合几个具体的实例来介绍计算流体力学的内容和它的应用.

6.2 计算地球流体力学问题

地球流体力学研究地球上的流体宏观运动的共同规律及其有关的实用问题.所谓地球上的流体主要是指大气、海洋、河流、湖泊、空间以及包括生态过程在内的环境.地球流体力学区别于一般流体力学的一个主要特征是必须考虑重力和地球旋转力.由于自然界流体运动受很多因素和复杂的相互作用过程所控制,不进行大规模的计算难于得到正确的研究结果,特别是无法解决一些主要的实用问题,例如预报天气剧烈变化过程,预测数月至数年的旱涝等气候灾害,预测年景丰歉,预测气候和环境变迁,预测海洋状况变化,预测鱼群回游路线,防治环境污染和进行环境生态规划等等.因此大规模科学计算在这些领域中得到了广泛的重视和大规模的应用.研究工作经过从个别到一般、具体到抽象的不断上升,就形成今天的计算地球流体力学.

下面,就气候的数值模拟对计算地球流体力学作具体的阐述.

6.2.1 当代气候问题

气候变化及其对经济和社会发展影响的问题已成为当前世界各国政府和科学家们所关注的重大问题.近几年来,世界范围内的气候异常给许多国家的粮食生产、水资源和能源带来严重的影响,例如,1988 年,美国中西部严重干旱造成粮食减产 37%,而孟加拉国却遭受前所未有的水灾;非洲持续多年的干旱,使许多国家遭受本世纪最严重的粮食危机.因此,能否预报气候的异常变化已成为一个迫切需要解决的重大科学问题.近四十年来,我国的气候灾害也频繁发生,每年都有数省或大江大河流域面积大小不等的地区遭受干旱和水灾,而且往往是同一年份同一地区这季干旱而另一季洪涝,发生大范围的持续性的大旱大涝年份也为数不少.平均而言,每年因受旱涝灾害少收粮食约 2000 万吨.因此,提高对气候灾害的预报和防治能力已是我国各级政府和有关部门的迫切要求.同时,未来全球规模的气候变化趋势,我国国民经济的发展和各地区对各类资源的开发所带来的对气候和环境的影响,也是十分值得关注的研究课题.

在 50 年代以前,人们把气候当作静态来研究,当代则把气候看作是不断变化的.它有着年代际尺度、十年际尺度、百年际尺度和千年际尺度不同的变化.

除太阳辐射这个主要能源之外,气候的形成和变化受诸多因素的影响,它不仅是大气内部的状态和行为的反映,而且是与大气有明显相互作用的海洋、冰雪圈、陆地表面及生物圈等所组成的复杂系统的总体行为(图 6-1).各子系统内部及各子系统之间的各种物理、化学乃至生物过程的相互作用决定气候的长期平均状态

以及各种时间尺度的变化，从这个意义上说，气候学是大气科学、海洋学、地球物理学、生物学等众多学科相互渗透共同研究的交叉学科.

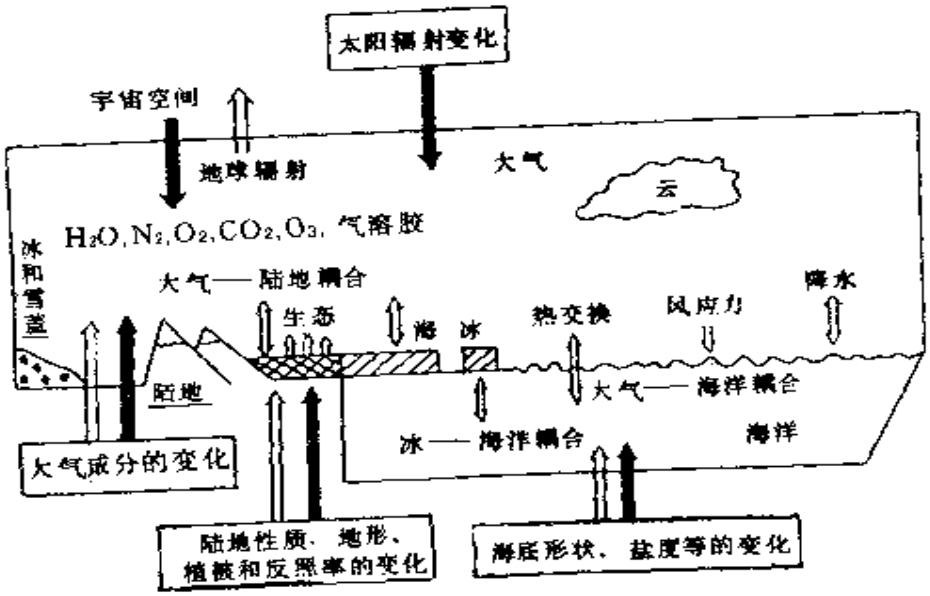


图 6-1　大气—海洋—冰—陆地—生物圈组成的气候系统示意图

经典气候学多是统计、定性的描述，近代气候学需要推理和定量的研究，而当代气候学则要求对气候系统进行定量观测和综合分析，并对气候形成和变化的动态过程进行理论研究和数值模拟. 目前用数值模式模拟气候虽已取得很大的成功，但由于气候系统的复杂性，其模拟结果仍存在着很多问题.

6.2.2　气候数值模拟问题

研究气候变化的一个主要方法是"物理—动力方法"，这就是根据基本的物理定律(如牛顿运动定律、能量守恒定律和质量守恒

定律等)建立起来的流体力学方程组来确定"气候系统"的各个部分(大气、海洋、冰雪、植物及陆地表面等)的性状,由此构成了气候的数学模式,也就是人们所说的气候模式.气候模式不仅可用于模拟当代气候,而且可用于模拟某些"外部"条件(如地球大气所接受的太阳辐射等)和内部参数(如大气中的 CO_2 浓度等)的变化所引起的气候变化.因此,如果说人类能够用试验方法来研究气候及其变化的话,那么最重要的试验手段就是利用气候模式进行数值模拟.当然任何模式都只是实际气候系统的某种程度的近似,在利用它们模拟气候的同时,还必须不断地检验它们本身的可靠程度并加以改进.

最基本的流体动力学模式就是大气环流模式(AGCM),也常简写为 GCM,比较完善的气候模式还应包括大洋环流模式(OGCM),陆面过程模式(LPM)等.

大气环流模式的控制方程组基本上是流体力学方程组.为了求得数值解,通常先将大气沿垂直方向(即从地面到天空)划分为少量的若干层(例如 7 层,9 层),将要计算的变量(包括预报量和诊断量)安排在各层中或者层与层之间的界面上,变量在每一层上的水平变化可以由一张覆盖着整个地球的格网上的值来表示(图 6-2),也可以由有限个基函数的线性组合给出,前者称"格点"模式或者有限差分模式,后者则称"谱"模式.

模式变量的时间变化也是离散的,给定预报量在某一时刻的值(称为"初条件"),利用模式方程组按一定时间步长外推(称为"时间积分"),就能求得它们在任意指定时刻的数值,该时刻的诊断量的数值则由已求得的预报量按诊断方程式计算.为了避免计算不稳定的发生,时间积分的步长通常不能超过某个临界值,它是

由模式所包含的物理过程、网格步长的下界(或者谱模式中具有最小空间尺度的基函数)以及时间外推方法所决定的.

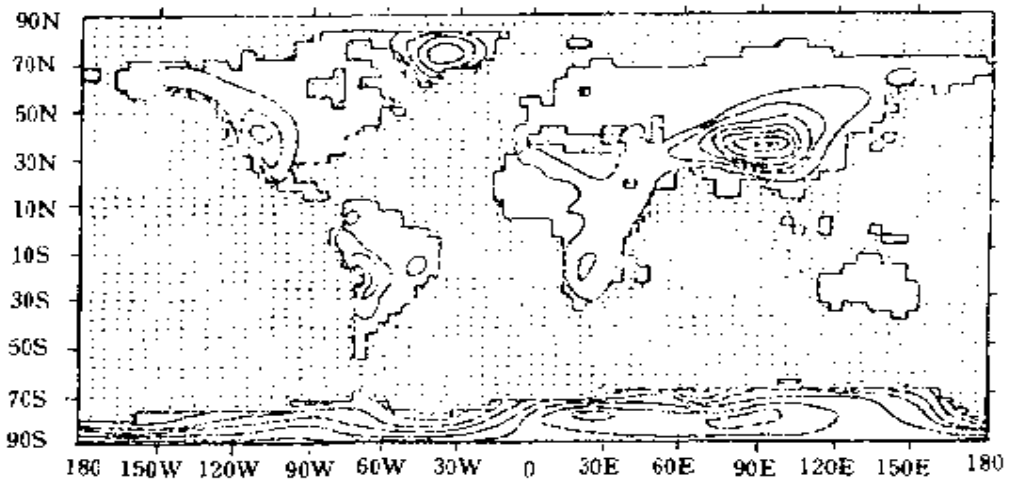


图 6－2　大气环流模式的陆面高度及海陆分布示意图

6.2.3　时间积分问题

为了进行气候模拟,往往需要将大气环流模式积分几个月、几年乃至几十年、上百年.由于计算机的运算速度和存储量的限制,模式的空间分辨率(垂直层数和水平网格点数或基函数的个数)就不能取得太高,即使如此,气候模拟的计算量还是非常大的.国外气候模拟早已使用像 CRAY－XMP 和 CYBER－205 这样的巨型机.如表 6－1 给出的就是美国普林斯顿的地球流体力学实验室((GFDL)在最近 30 年中使用过的计算机.可以说气候模式的发展总是同计算机的发展相伴随的.而且前者对后者的要求甚至是超前的.就以中国科学院大气物理研究所现有的 9 层大气环流模式来说,在该所的 Convex 3210 机上数值积分一年需 CPU 时间 4

天，积分十年需 CPU 时间一个月，而对于气候模拟问题，往往需要连续数值积分几十年至一百年，所需的 CPU 时间就是数月至一年的时间，如果再把垂直分层和水平分辨率增加一倍，其计算量将增加一个量级以上。因此，在我国计算机条件明显落后于国际先进国家的条件下，要在气候模拟这类问题上赶超世界先进水平，除了努力发展运算速度更快的大型电子计算机外，积极发展各种快速算法具有更现实的意义。

表 6－1　　GFDL 电子计算机更新情况

年　代	计算机型号	速度(万次/秒)	内存(万字)
1963	IBM 704	1.0	3.6
	7090	23.0	3.2
1965	IBM 7030	65.0	1.6～26.2
1967	CDC 6600	300.0	3.2～13.1
	UNIVAC 1108	130.0	3.2～13.1
1972	IBM 360/91	500.0～600.0	12.1～102.4
1976	CDC 7600	8000.0	52.1
1982	CYBER 205	32000.0	400.0
1984	两台 CYBER 205		

6.2.4　快速算法

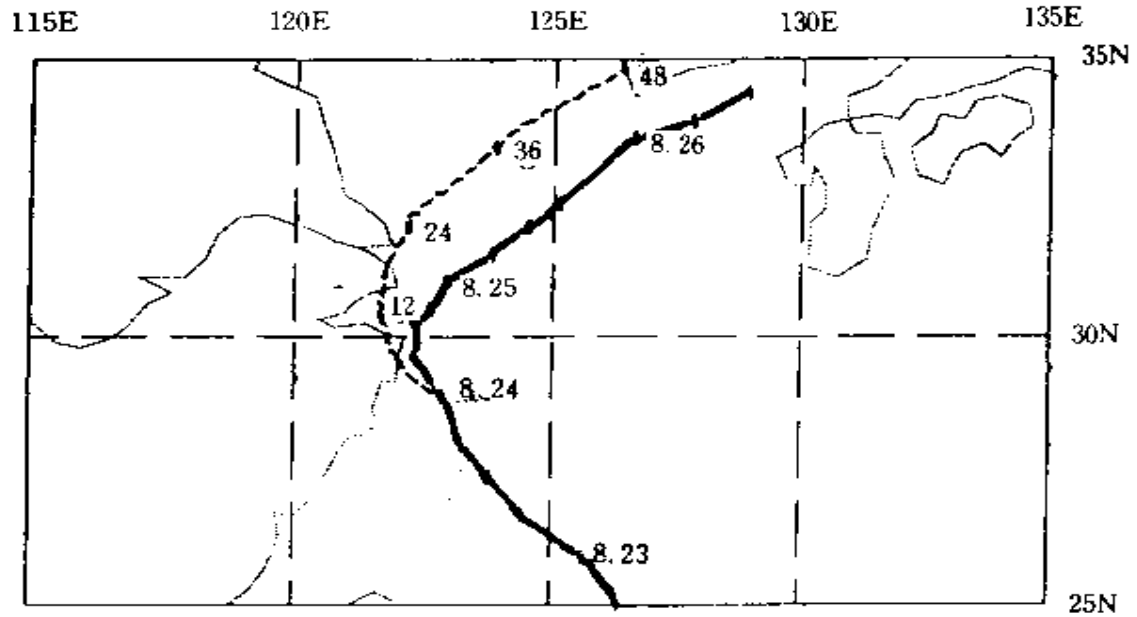
就以有限差分法来说，传统的模式主要有两类：一类是采用显式中心差分格式(leap－frog)进行时间积分，这种方法的时间步长不能取得太长，通常为几分钟至十几分钟，又由于时间中央差格式是三时间层的，由初值和积分过程中产生的误差会分离出一个虚假的

计算解，往往易于产生非线性计算不稳定，使数值积分难于长时间进行下去；另一类模式，采用隐式时间积分，时间步长可以取得较长，但往往需要进行非线性迭代，同样耗费机时过多。人们自然会问，能否设计出既能显式求解又能长时间计算稳定的计算格式呢？回答是肯定的。从"七五"后期至"八五"期间，中国科学院大气物理研究所发展了显式完全平方守恒的差分格式，就是能初步满足上述要求的一类新格式。与之相配合，还在原有算法的基础上发展了多种经济有效的新算法（如特征方向法、区域分离法和改进的分解算法等）均取得较好的效果。图6-3给出台风路径数值模拟的计算结果，图6-3(a)是原方案计算的，而图6-3(b)是新方案计算的，从图上不难看到新方案计算效果是有明显改进的，更重要的是新方案节省计算时间较多，只需原方案计算时间的1/9即可。从表6-2可以

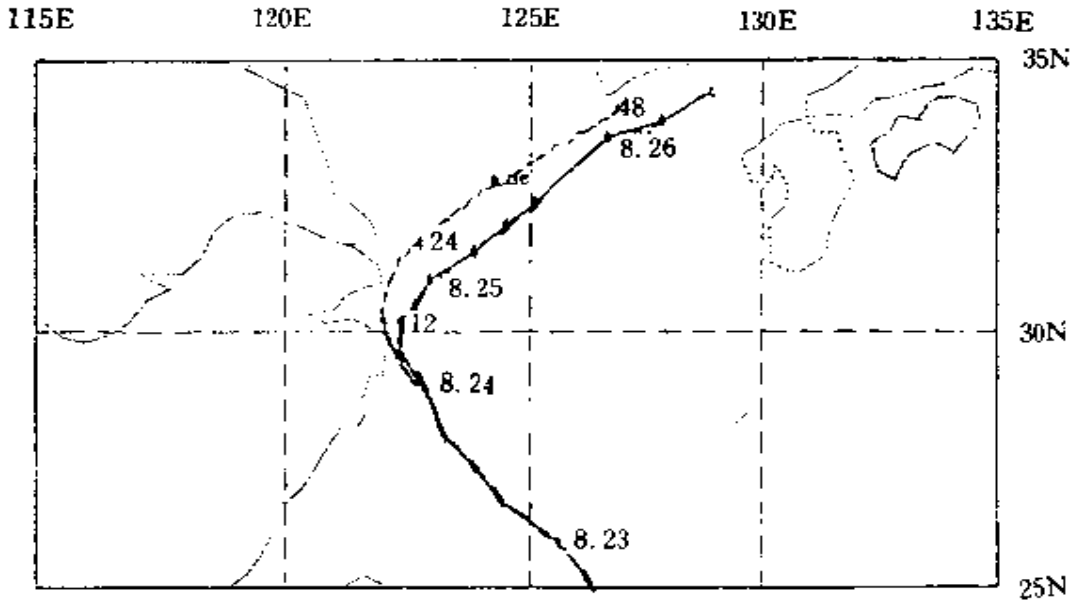
表6-2　　四种方案的结果对比

	原隐式方案	显式一阶计算方案	显式一阶简化计算方案	显式二阶计算方案
积分时间(d)	30	30	30	30
CPU时间(min)	9.80	2.60	2.12	2.92
$\overline{V}_{max}$(cm/s)	25.75	25.62	25.64	25.88
位置(I,J)	(26,23)	(26,23)	(26,23)	(26,23)
H_{max}(cm)	18.06	17.98	17.99	18.04
位置(I,J)	(5,47)	(5,47)	(5,47)	(5,47)
H_{min}(cm)	41.09	40.92	-40.94	-41.05
位置(I,J)	(47,4)	(47,4)	(47,4)	(47,4)

注：$\overline{V}$为海流速度，H为海面起伏。



(a) 显式瞬时平方守恒方案



(b) 显式完全平方守恒方案

图 6 - 3　7910 号台风 72 小时路径预报(初始时刻 8 月 24 日 8 时，虚线为预报结果，实线为实况)

看到四种方案计算时间也有很大不同，显式方案比隐式方案省时80%.

6.2.5 能量守恒的半拉格朗日算法

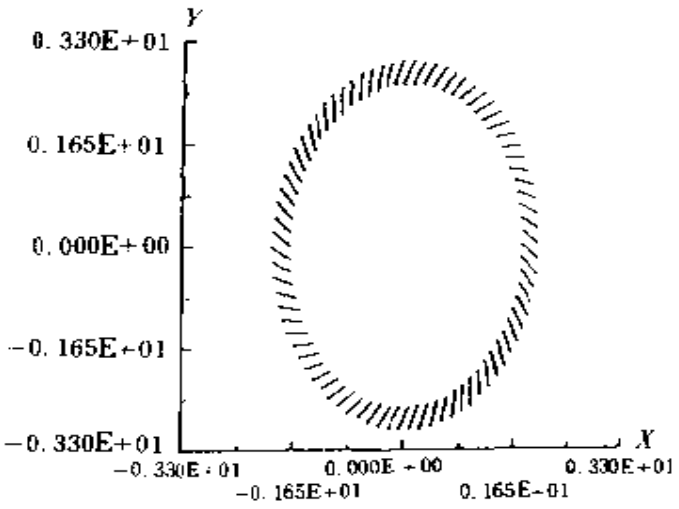
为了节省计算时间，改进计算效果，近十多年来在国外积极发展了一种半隐式半拉格朗日格式.所谓半拉格朗日格式，是以随质点运动的半拉格朗日观点处理平流部分，它考虑在时间截断误差和空间截断误差相当情况下，最大限度地放大时间步长，而不损失计算精度，而且能保持计算稳定.这种格式没有欧拉格式的非线性计算不稳定性，实际应用时，将它与半隐式格式结合起来可使时间步长进一步放大，约为欧拉半隐式格式的4～6倍.目前，这种半隐式半拉格朗日格式逐渐为世界各国天气预报所采用，例如欧洲中心(ECMWF)、加拿大的CMC、美国的NMC和NCAR等，它们不仅应用在中期天气预报上，而且也应用于气候预测中.纵观半隐式半拉格朗日格式从80年代到90年代的发展和应用，有两个问题需要解决，一是此格式需要四周格点内插出流体质点的上游和中途位置，这就带来预报场的人为光滑；二是它不能像欧拉格式那样保持主要物理量的守恒.这两个问题给长时间数值积分带来问题，现在世界上不少人致力于此问题的研究.近年来，中国科学院大气物理研究所已提出完全不需要任何内插的半拉格朗日格式方案，通过数值试验表明效果良好.在"八五"期间，还进一步发现，前述的特征方向法就是一种能保持完全总能量守恒的半拉格朗日算法，这实际已为解决上述第二个问题向前推进了一大步.

6.2.6 平方守恒格式的应用

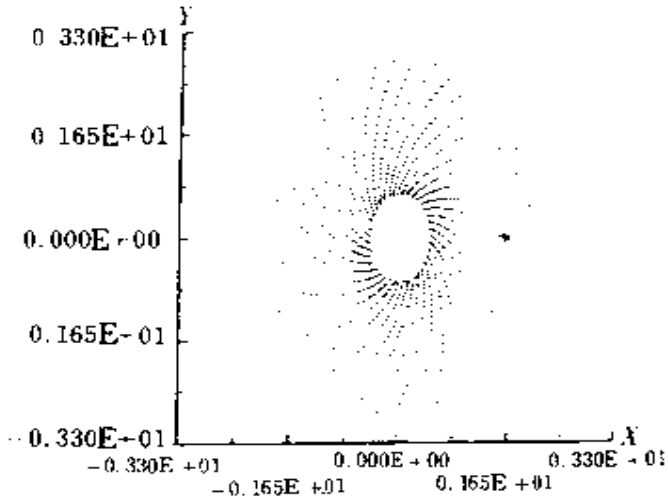
目前，平方守恒格式除了在台风路径数值模拟和近岸海流数

值模拟工作中取得良好效果外，不久前，针对河流泥沙冲积和三角洲发育问题，中国科学院大气物理研究所又发展了一个二维数值模式．数值试验表明，该模式有较好的数值模拟能力，能模拟出随着泥沙在邻近河流入口处地区的沉积，使那里水流变浅，流速变大，到一定程度之后，湍流过程增强．原河床沉积的泥沙被扬起，河床被冲刷，泥沙加速向下游和漫滩区输送并沉积，最终发育成完好的水下三角洲，逐渐向下游推进，并形成河口江心洲型的水下岛屿，三角洲边缘十分陡峭．这样的研究对河道航运、水利灌溉，水利工程和港口建设等有一定的启发指导意义．

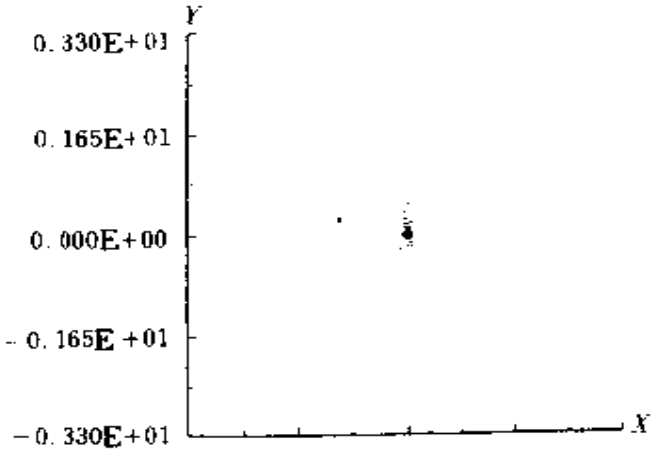
Runge-Kutta 方法是常微分方程数值求解中常用的数值方法，但当该方法作较长时间的积分后数值解会表现出较大的耗散性，如图 6－4 所示的一个常微分方程的解，当积分到 10^6 步时，已将一个椭圆耗散为一个点，这是不符合物理规律的．为了使这类方法能适应气候模拟这类需作很长时间数值积分的计算问题．



(a) 10^6 步积分结果(每隔 10^3 步输出一点)



(b) 10^7 步积分结果(每隔 10^4 步输出一点)



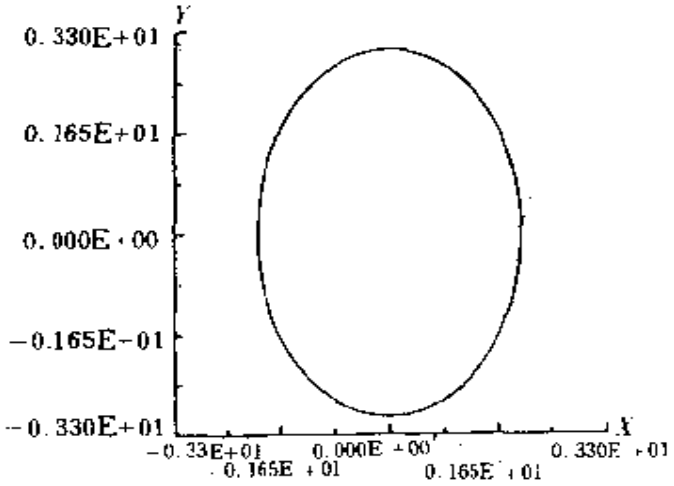
(c) 10^8 步积分结果(每隔 10^5 步输出一点)

图 6-4　用 Runge-Kutta 法所得的近似解

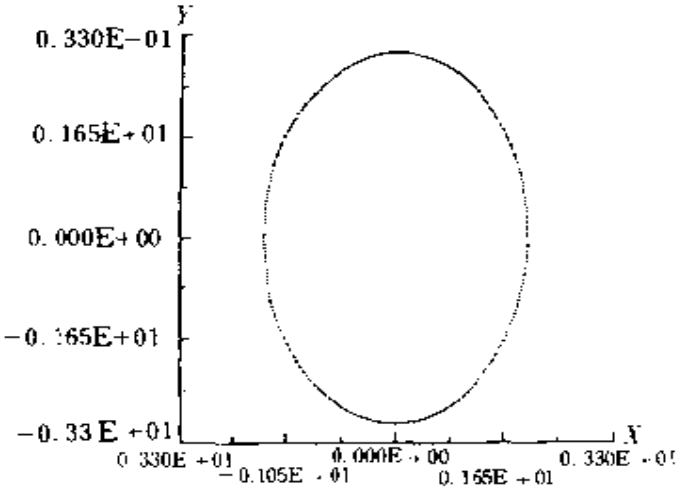
人们将平方守恒的概念引入到 Runge-Kutta 方法中,使该方法获得新生,图 6　5 给出用改进后的 Runge-Kutta 方法计算上述同一

问题的结果，无论积分 10^6 步、10^7 步还是 10^8 步，方程的解始终保持在同一个椭圆上.

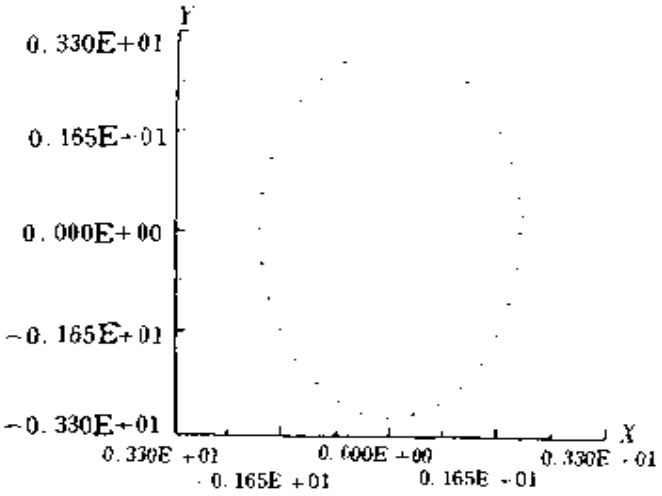
根据大气、海洋问题的物理特点，还将计算数学中的一些新方法(如辛格式等)加以发展推广，并与平方守恒概念结合起来，发展了辛算子法和广义哈密顿算法等，它们在数值试验和应用中也取得很好的效果.



(a) 10^6 步积分结果(每隔 10^3 步输出一点)



(b) 10^7 步积分结果(每隔 10^4 步输出一点)



(c) 10^5 步积分结果(每隔 10^5 步输出一点)

图 6-5 改进的 Runge-Kutta 方法的计算结果

6.3 飞行器和计算流体力学

航空与航天是 20 世纪人类认识和改造自然进程中最活跃、最有影响的科学技术领域,也是人类文明高度发展的标志.航空是指飞行器在地球大气层内的航行活动,航天是指飞行器在大气层外宇宙空间的航行活动.飞行器的外形不断改进,性能不断提高,使人类从大气层内的航行扩大到了大气层外宇宙空间的航行,这些都是与人们对飞行器的空气动力特性认识不断深化分不开的.计算流体力学在这一认识过程中起到了关键性的作用.

为了认识飞行器的性能必须研究空气动力特性,也就是研究飞行器和空气相对运动时所产生的力、热和其他物理现象.计算流体力学推动了空气动力特性的研究工作.自从 1687 年牛顿定律公

布以来，直到本世纪50年代初，研究气体流动特性的主要方法有实验研究(以地面实验为研究手段)和理论分析方法(利用简化的流动模型假设，给出所研究问题的解析解)两种。在后一方面，理论工作者在研究气体运动基本规律的基础上，建立了各种简化流动模型，给出了一系列解析解，推动了空气动力学的发展。然而，仅采用这些方法研究复杂的非线性流动现象是不够的，已不能满足高速发展起来的航空航天事业的要求。50年代以来计算流体力学的兴起促进了实验研究和理论分析方法的发展，将实验研究与理论分析方法联系起来。然而，更重要的是计算流体力学采用它独有的新的研究方法，即数值模拟方法，研究空气动力特性。

计算流体力学在飞行器空气动力特性的研究中是用数值方法借助于电子计算机求解满足初始条件和边界条件的空气动力学基本方程组，即描述空气运动的动力学特性应满足的基本物理规律：质量守恒律、动量守恒律和能量守恒律的数学方程组，对飞行器周围空气流动的流场进行数值模拟，研究飞行器与空气相对运动时所产生的力、热和其他物理现象以及对空气动力特性的影响。在这一领域，计算流体力学自60年代后期以来得到了迅速的发展。主要原因是利用计算流体力学的方法，不但可以对外形给定的飞行器进行空气动力特性的分析，还可以按预定的空气动力特性要求设计飞行器的外形，得出空气运动流场中物理量的细节分布，不存在风洞实验中洞壁和支架干扰等一系列理论分析方法无法解决的问题。这就大大促进了先进飞行器的发展。计算流体力学发展很快的另一个原因是计算机的运算能力迅速提高，计算花费的时间不断下降。计算流体力学在飞行器的空气动力分析和设计中发挥着越来越大的作用，使飞行器的设计过程发生了根本性的变革。

航空航天飞行器周围空气运动的流场非常复杂，且随飞行器的飞行高度和飞行速度有很大变化。飞行器的发展过程也是从简单逐渐走向复杂，从易到难的过程。计算流体力学的发展正是适应这个过程的转换。下面以飞行器外形及飞行速度的发展为例来说明这个问题。

30年代，航空工业发展的初期，飞机的飞行速度较低，飞行马赫数，即飞行速度 v 与音速 a 之比 $M=v/a<0.3$，飞机的外形为简单的细长机身及薄翼的组合体，且飞行仰角 α 很小，故飞行器对其周围空气运动流场产生的扰动为小扰动。这种情况下，可忽略空气的压缩性和粘性，空气运动的基本规律可用线性的位势流方程——拉普拉斯方程来描述。求解的方法是基本解的叠加，如数值解方法——面元法。以后，为了考虑粘性效应，有了边界层方程的数值计算方法。随着计算机和计算技术的发展，为了研究飞行器壁面附近粘性流与外层无粘流的干扰对空气动力特性的影响，位势流方程发展为外流无粘流方程与内流边界层方程相结合，通过迭代，数值模拟空气粘性干扰场的气动特性。随着飞行器飞行速度继续增高，飞行器周围空气运动流场更加复杂。当飞行马赫数 $M>0.3$ 时，就必须考虑空气的压缩性。当飞行速度接近音速时，在飞行器周围空气流场（也称为飞行器的绕流流场）中会出现局部的超音速区，即局部马赫数 $M>1$ 的区域，在其后形成弱激波。若飞行器的外形仍为细长体，飞行器对空气的扰动仍假设为小扰动，可假设速度的位势存在，空气的粘性可忽略，这时空气运动的基本规律可用非线性位势流方程来描述，从而在计算流体力学中该类方程的数值求解方法得到了发展。

当飞行速度超过音速几倍时，飞行器绕流流场变得十分复杂，

首先将出现强激波，高速气流过激波后温度升高，使得原有的尖头飞行器已不适用，否则将被烧掉，出现了钝头体飞行器．在图 6－6 中给出了外型为球—锥体的飞行器(如超音速导弹弹头)绕流流场特征示意图．在飞行器前缘出现了弓形激波，激波后在飞行器对称轴附近流场为亚音速流场($M<1$)，这时原来对于简单细长体绕流合适的小扰动假设不成立．若考虑飞行雷诺数足够大，则粘性较大的区域只在飞行器壁面附近的小区域内(参见图 6－6)，故仍可假设空气无粘性，这时可用非线性欧拉方程来描述飞

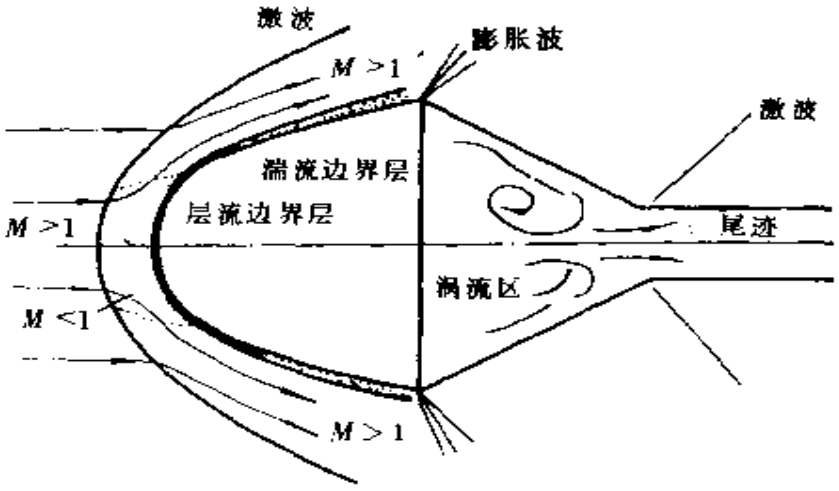


图 6－6　超音速钝头体绕流

行器的空气动力特征．从 40 年代开始，随着喷气式飞机、超音速导弹等的出现，在气体动力学中提出了很多需要解决的实际问题，以位势流方程、线性位势流方程为基础方程的计算方法已不足以解决这些实际问题．这进一步促进了计算流体力学的发展．中心问题是非线性激波的数值处理，从 50 年代中期到 60 年代初，针对这些问题发展了一系列计算方法，用以求解非线性欧拉方程的定常解．

前面提到的钝头体绕流的数值模拟是一典型例子.在60年代初,该问题成为飞行器绕流的一个难题,当时著名的流体力学家Lighthill曾认为"这是近期难以解决的问题".但原苏联的计算流体力学专家多罗德尼臣(Дородницын)提出的积分关系式方法和范达克(van Dyke)提出的反方法以及以后捷列宁(Теленин)提出的直线法用以求解欧拉方程,成功地解决了该问题.这是计算流体力学对航空航天飞行器发展的一个重要贡献.

60年代中后期,随着电子计算机和计算技术的飞速发展,计算流体力学在航空航天飞行器的发展中起到了越来越大的作用,很多难以解决的问题通过计算流体力学方法逐步得到了解决.这里以航天飞机为例.当航天飞机以超音速飞行时(飞行速度8倍于音速),飞行器周围空气流场除了前面提到的钝头绕流区以外,由于航天飞机的座舱(或其他凸起物)将导致气流的分离.这是壁面附近的粘性气流受到物面形状的较大变化而引起的.分离区的存在又诱导出分离激波.后体由于机翼的存在也将产生激波和气流的分离.这些流场中产生的激波与航天飞机前缘形成的弓形激波相交将产生激波—激波干扰;与壁面粘性层之间将产生激波—附面层干扰,流场十分复杂.研究该类流场的空气动力特性,理论分析方法较困难,单纯的实验研究也难以解决,必须依靠计算流体力学方法与实验研究相配合.描述该类流场内空气运动规律的数学方程,非线性欧拉方程已不适用,必须求解三维可压缩Navier-Stokes方程(以下简称N-S方程),数值模拟该类流场,研究其空气动力特性.60年代中期以来,基于双曲型方程数学理论的时间相关方法开始应用于求解航天飞行器绕流的定常问题.其基本思想是从非定常欧拉方程或非定常N-S方程(或简化N-S方程)

出发，利用双曲型方程或双曲—抛物型方程的特征，沿时间方向推进求解，所得的对于时间的渐近解为所求的定常解。

另外，航天飞机再入大气层后，飞行马赫数最大达 26～28，即飞行速度为音速的 26 倍左右，飞行轨道的大部分区域，由于存在弓形激波，激波后气流温度可达 12000K 以上，故空气分子内部自由度激发，产生离解和电离，处于化学非平衡状态。这对空气动力特性，特别是气动热产生较大影响，称为非平衡化学反应效应或真实气体效应，这是地面实验难于模拟的，理论分析方法更是无能为力，只有依靠计算流体力学方法进行研究，且已获得了很好的成果。对于该类问题描述空气运动基本规律的数学方程是 N－S 方程或简化 N－S 方程和组元守衡方程，其中包括有组元的质量扩散项和扩散引起的能量输运项等。

70 年代以来，计算流体力学取得巨大成功的领域之一就是采用时间相关方法求解可压缩 N－S 方程，数值模拟飞行器超音速（$1.5< M <5$）、高超音速（$M >5$）粘性绕流复杂流场。现在已可用于模拟包含有各种宏观尺度结构的非光滑流场，如包含有激波、粘性干扰、分离流、真实气体效应等物理特性的流场。今天已可利用巨型计算机，采用合适的网格生成技术和有效的计算方法，求解非定常可压缩 N－S 方程，数值模拟航天飞机整机的超音速、高超音速粘性绕流流场。

现在国外已有在飞机的设计过程中用数值模拟代替风洞试验的做法。例如 Lockheed 公司在设计 F－22 的过程中就用求解 Navier-Stokes 方程的大型软件 TEAM 计算了整机在大约 370 种状态下的气动参数，包括马赫数从 0.6 到可能最大的值，仰角从 －4°到 24°，以及机翼前后缘襟翼，水平尾翼和方向舵等的各种偏

转情况，计算区域是三维的，共划分了125万个网格，在CRAY机上共计算了1600小时，Lockheed公司宣称，由于采用了数值模拟的手段替代风洞试验取得了飞机设计所需的气动参数，因而大大降低了成本并且缩短了研制周期。人们估计，为使飞行器的数值模拟结果能达到工程需要的精确度，计算一个问题大约需要10^{15}～10^{15}次运算，普遍认为到本世纪末计算机可以达到每秒进行10^{12}次运算的水平，因此从计算机的发展来看是能够满足飞行器设计的需要的。

随着我国航空航天事业的发展，在飞行器的研制中，我国在60年代就有了自己的计算流体力学研究队伍，最早的工作是研究钝头超音速无粘绕流的计算方法，针对第一代飞行器的要求，开展了欧拉方程定常解的数值方法研究，给出了钝体超音速和高超音速无粘绕流的计算结果，为战略导弹的第一代弹头的气动设计做出了贡献。1980年5月18日，我国洲际导弹从本土向太平洋发射成功，宣告了战略导弹第一代弹头的研制工作告一段落。随着我国宇航飞行器的进一步发展，提出了更为复杂的空气动力特性问题，如粘性干扰、分离流、真实气体效应等，为此，在70年代中期开展了采用时间相关方法求解欧拉方程、N－S方程和简化N－S方程的计算方法研究，在差分格式的构造、计算方法精度的改进、提高求解效率、提高激波分辨能力等方面作了深入的研究，建立了我们自己的计算方法，并利用这些方法求解欧拉方程和可压缩N－S方程，给出了各种飞行器绕流流场的计算结果。目前我们有能力采用自己的方法和适当的网格生成技术，数值模拟航天飞机整机的超音速、高超音速绕流的复杂流场。在图6－7～图6－10中给出了采用我们自己提出的耗散比拟方法，求解三维N－S方程，数值

模拟航天飞机哥伦比亚号绕流流场、飞行马赫数 $M=8$、飞行仰角 $\alpha=5°$的计算结果.

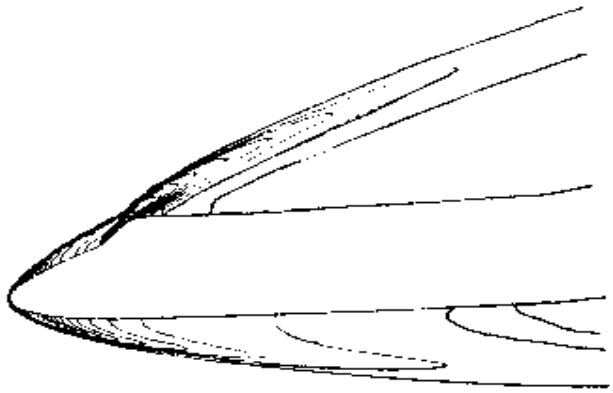


图 6－7　航天飞机哥伦比亚号纵对称面的压强等值线分布，$M_\infty=8$，$\alpha=5°$，$Re_L=1.15\times10^6$. 最大最小值之间 71 个等距间隔.

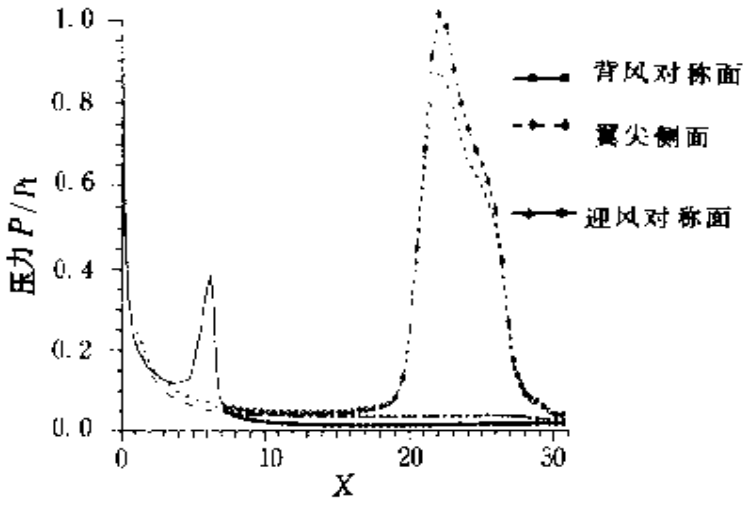


图 6－8　航天飞机哥伦比亚号沿轴向压强分布. 无符号线：网格数 $(39\times41\times46)+(51\times53\times55)$. 带符号线：后体网格数 $51\times41\times46$

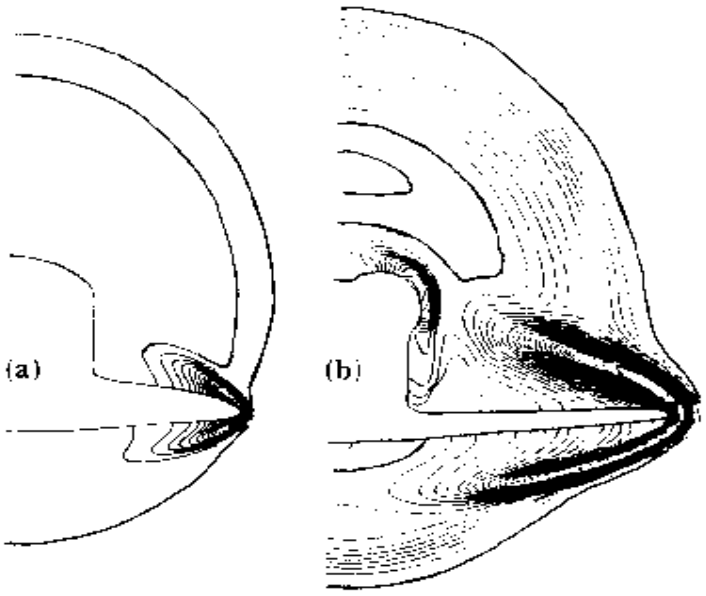


图 6 9 航天飞机哥伦比亚号横截面内的压强等值线分布，(a)X = 22 (b)X = 28

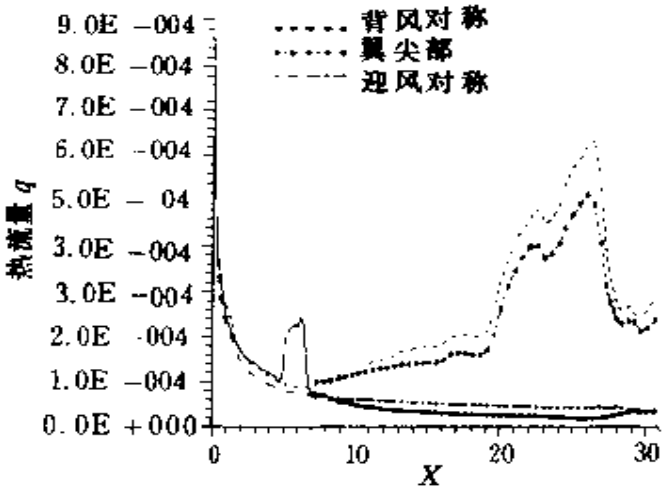


图 6 − 10 航天飞机哥伦比亚号沿轴向的热流强度分布，无符号线：(39×41×46) + (51×53×55)，带符号线：后体网格 51×41×46

从以上简单的叙述可以看出，计算流体的研究工作为20世纪人类走向大气层外的宇宙空间起到了重要的作用.

6.4 高速碰撞和计算流体力学

计算流体力学在研究爆炸和高速碰撞现象中发挥着重要作用，为深入揭示各种爆炸过程的内在机理提供了一种强有力的手段，改进了这一领域过去主要靠实验的研究方法的状况，推进了研究工作的深度，提高了人们控制和应用爆炸现象的水平和能力.

提及爆炸和高速碰撞现象，人们首先会想到军事上的应用，各种弹药如炮弹、航空炸弹、破甲弹、鱼雷、水雷等都是利用爆炸作用的典型例子.其实，在国民经济建设中爆炸现象也广泛存在，有的是需要加以防避的，如工厂的粉尘爆炸、煤矿的瓦斯爆炸等，有的则是进行利用的，例如开山炸石、筑坝填海、爆炸拆除各种大型工程等.民用工业中的爆炸切割、爆炸焊接、爆炸成型、油井射孔等，也是利用爆炸作用来实现预期的目的.

为了利用爆炸现象，人们必须充分掌握爆炸的规律，对各个具体应用项目做出可靠有效的设计，为此，研究设计人员就需要对一系列与爆炸及其作用有关的课题进行深入研究，例如炸药的爆轰过程、爆轰波与物质的相互作用、高温高压状态下物质的变形和运动，物质间的高速碰撞，物质的相变、熔化和断裂等等.过去在没有数值计算这一手段时，为做爆炸应用项目的设计，研究设计人员只能依靠一些简化的理论公式和经验公式，再凭借已有的实践经验，作出初步的设计方案，进行实弹试验，直到获得一个达到要求的方案为止.这样的设计，工作量相当大，要消耗大量的人力物力财力、

要花很长的时间周期.例如为设计一个破甲弹,有时就要做上千发实验.不仅如此,这样做出的设计方案,还不一定就是最好最满意的.因为,经验公式和爆炸实验所能给出的结果是有限的,给不出爆炸作用全过程及其各种细节的结果.这样人们就难以透彻了解爆炸装置的内在作用机制,自然也就不易做出理想的设计方案.

事实上在爆炸作用下,物质处在高温高压高速运动的状态中,这时物质可看成是连续介质,它的运动参数(速度)和状态参数(密度、压力、内能等)仍然满足流体力学方程组.更精细一点的理论,还要考虑到某些物质的弹塑性,因而在一定的条件下物质参数满足的是弹塑性流体力学方程组.所以可以用求解流体力学方程组或弹塑性流体力学方程组部分地或大部分地代替实验取得设计所需要的参数.并且人们可以通过数值计算结果的输出细致地了解物质在爆炸作用下变化的全过程,可以通过不同状态参数的计算结果的对比,分析不同因素的影响,从而掌握爆炸装置的作用机制,就可以作出更为满意的设计方案.

应该指出,流体力学方程组反映了连续介质运动的共性.但是,任何一种运动又是具体的、性质各异的,人们实践中碰到的和要研究的都是具体的运动.就以爆炸现象为例,就有核爆炸、化学炸药爆炸、煤气爆炸、粉尘爆炸等等,它们的性质和作用是大不相同的.爆炸又可能发生在空中、地下、水中、建筑物里、车船上等不同的地方,于是,它的作用对象就可能是空气、泥土、水、混凝土、金属等等不同物质.这就是说,研究爆炸现象跟研究自然界任何现象一样,不仅要描述运动的共性,而且还要描述其个性.所以,在研究具体问题时,除要用到上述的守恒方程外,还必须增加表达具体运动特性的关系式和表达具体物质性质的本构关系、状态方程和物

性参数(如杨氏模量、剪切模量、屈服强度、声速、粘性系数等等).举例来说,若研究炮弹爆炸问题,则就需要增加表达炸药化学反应的能量释放的方程,增加表征弹壳金属的物性参数和状态方程等.

这样以守恒方程为核心再加上本构方程、状态方程和各种相应的关系式组成的方程组,就构成对爆炸和高速碰撞现象进行理论描述的数学模型.当守恒方程中考虑应力应变,再加上表达应力应变关系的本构方程时,这就是通称的弹塑性流体力学方程组.

6.4.1 数值计算方法

众所周知,研究流体力学问题可以从两种观点出发.一种观点是在固定的空间坐标中讨论流体力学的运动,也就是空间坐标位置固定不变,研究流体在各坐标点上的变化情况,这称为欧拉观点;另一种观点是跟踪流体各质团研究其运动,坐标系就建立在流体上,跟随流体一起运动,这种观点称为拉格朗日观点.与此相对应,数值计算方法也分为欧拉方法和拉格朗日方法两大类.

欧拉方法是将物理空间中的求解区域划分为网格后,网格固定不变,求解每个网格上的力学量的变化.这种变化一方面是由于压力(包括外力)分布不均匀造成的,另一方面是流体从邻近的网格流入本网格,或本网格的流体流出进入到邻近网格而引起的,这就要求计算流体通过网格边界的流量(或称输运量).输运量的计算是一个需要专门研究的问题,不恰当的算法,会导致整个计算的不稳定性即计算中难以避免的误差(例如舍入误差等等)成无法控制的增长发展.欧拉方法面临的一个难题是如果要求计算的问题包含有多种介质.这时即使初始时刻不同物质之间的界面是清晰的并有确定的位置,但流体运动后,这种物质界面也会变得模糊.

一定会出现在一个网格中存在两种不同物质的情况，这种网格称之为混合网格．如何判断物质界面的位置，如何计算混合网格上的热力学量，如何计算混合网格向邻近网格的输运量都是需要特别加以处理的问题．有一种做法是在欧拉网格上增加一些质点或标志，不同的质点或标志代表不同的介质．然后在整个计算过程中跟踪这些质点或标志，就可以显示不同介质运动变化的情况．根据这些质点或标志的空间分布就可以大体上勾划出物质界面的位置．再按照混合网格中不同物质所占的体积份额来计算混合网格上的热力学量，以及混合网格向邻近网格的输运量．

拉格朗日方法是将物理空间中的求解区域划分为网格后，一个网格就是一个固定的质团，网格跟踪质团运动，因而这种方法适宜于计算多种介质的问题，可以始终保持界面的清晰．由于质团之间没有物质交换，因此无须在网格边界上计算输运量．网格的形状就是质团的形状．因为在高维（二维、三维）流体运动中，质团会变形，于是网格也随之而变形．有时质团变形很严重，呈现出扭曲等畸形状态，网格也就会扭曲，甚至发展到网格翻转、对边相交等不合理的状态，以至计算无法继续下去．这种大变形的计算是拉格朗日方法面临的最大的难点．解决办法通常是在计算过程中适当地重新划分网格，也即中止跟踪原来的质团而跟踪新的质团，当然这种做法就不是原来意义上的拉格朗日方法了．容易理解，如果每计算一步都将网格重新划分一次，并取新网格为初始网格，那么这就成了欧拉方法了．

现在有一种方法称之为任意拉格朗日－欧拉方法，是将欧拉方法、拉格朗日方法和在拉格朗日方法中加上重分网格的方法统一结合起来，将守恒方程建立在运动区域（网格）上．这样在物理空

间中建立了网格后，如果网格固定不动，就是欧拉方法．如果网格按照流体的速度移动，则是拉格朗日方法．如果网格移动，但移动速度并不取作流体运动的速度而是任意人为规定的，这就相当于在拉格朗日方法中重新划分网格．由于这种方法具有一定的灵活性，所以得到广泛的应用．

由于欧拉方法和拉格朗日方法各有其优缺点，而且它们的成功与不足之处又往往是互补的．因此人们在实际计算中有时将两种方法用不同的方式结合起来使用．例如前面提到的用欧拉方法计算多种介质问题时，引入质点或标志，跟踪质点，这就是拉格朗日方法了．有时可将两种方法的程序结合起来使用，例如当运动中的物质变形不严重时，则用拉格朗日方法计算，而当物质发生大变形时，就改用欧拉方法计算．现在国外已做成了由多个各种功能的程序组成的程序包，用它来进行计算，可以根据运动发展情况适当使用合适的程序，这就使对复杂问题的计算能有效进行下去并获得好的结果．

还有一种做法是把解析方法和数值方法结合使用．有的问题在整体上很复杂，但其运动的发展在某一阶段上或在局部地方是可以用理论的解析解描述的，则在这些地方就使用解析解而省掉数值计算，而在其他地点和时间仍照做数值计算．这种解析与数值相结合的计算方法，能节省计算工作量，有时还能绕过数值计算的难点使计算得以顺利进行，取得很好的结果．

6.4.2 数值计算与理论设计

利用解流体力学方程的办法，也就是用数值计算的方法进行产品的理论设计，其工作过程可用以下框图表示(图 6－11)．

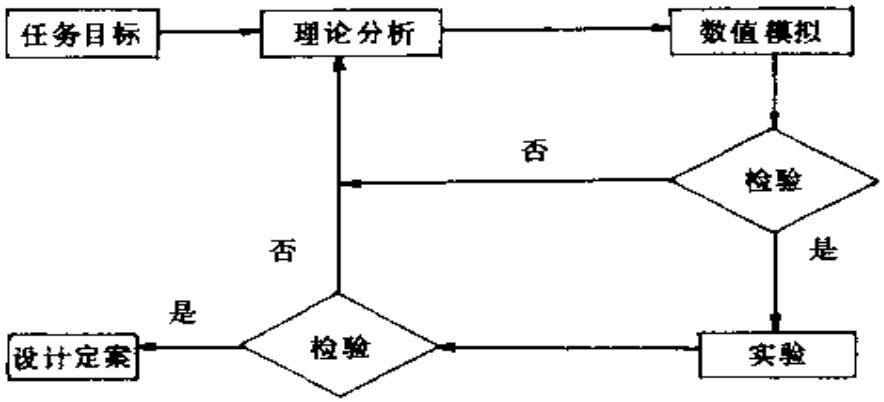


图 6－11　产品设计工作过程框图

这过程简单说来就是：先按设计任务的要求进行理论分析，提出设想和设计方案；然后，把方案做成计算模型进行数值计算；接下去是对计算结果进行分析评价，找出设计方案的优缺点，定量审核方案是否达到原定目标的要求．这中间为了掌握各种材料配置的作用，以求全面了解产品动作的机制和规律，还可以特别设计一些方案，改变材料的品种、质量，以及几何形状等等，作一些对比的数值模拟．工作到此可算为一个阶段，若分析认为方案未达标，则需修改方案，这时就返回去又从提新的设计方案开始，重复前面的工作过程．这样，一次一次地进行“分析—计算—分析”的工作循环，直到得出满意的设计方案为止．然后，就转入下一阶段工作，对设计方案进行试验．假若试验结果与已有的计算结果不相符，则工作就返回去又从最开头做起，进行“分析—计算—分析”的大循环，直到试验与计算相符，设计方案得到试验证实为止．这种设计方法的内容可简要地归纳为几句话：以理论分析为指导，用数值计算为主要手段，揭示机制，优选方案，最后用试验检验定案．作到了理论、计算、试验三者紧密结合，各显其能，互为依存和补充．

多年的实践表明，用数值模拟的方法进行常规弹药和其他复杂爆炸装置的设计，是非常有效的．每做一项研究设计任务，只要开始认真作好理论分析，提出合理的设计思想，然后，一般经过几轮数值计算就可做出较好的理论设计方案，并且，对该方案的实弹试验，多数情况能做到一次成功，试验结果与理论基本相符．能取得这样的效果，主要是数值计算起了重要作用．对设计方案进行的每一次数值计算，都可获得大量重要数据，清晰地显示该方案的作用过程及其性能．所以设计一个模型，做一次数值计算，就相当于在计算机上做一次"实弹试验"．充分用计算机的"试验"对设计方案进行有效的修改和优化，大大节省实际的实弹试验．

6.4.3 应用举例

作为实际应用的例子，下面介绍用数值模拟的方法设计石油射孔弹的情况．

石油射孔弹是采油必须的一种装备．每口油井打成后，其井壁是封住的，要开始采油时，须把许多射孔弹放入井内将井壁射穿，并在井洞四周的油层中射出许多细而深的小孔，深度一般需达到几百毫米．小孔越深，出油率越高，所以，射孔弹性能的好坏，直接关系到出油率．

石油射孔弹是一种爆炸装置，它利用炸药的爆炸作用实现射孔．由于受油井狭小空间的限制和井下安全的需要，石油射孔弹只能作得很小和装很少的炸药．要用很少的炸药射出很深的孔，为此，石油射孔弹采用了一种特殊的装药结构，即所谓的聚能装药结构．这种结构是柱形对称的．图 6－12 是它沿对称轴切开的剖面示意图．最外层是弹壳，中间装填高性能炸药，内层是一个非常薄的

锥形金属壳，通常称为聚能罩或药型罩，罩以内是一个空腔．石油射孔弹的作用原理是这样的：雷管点火引爆炸药，炸药爆炸产生的高温高压推动药形罩迅速向中心轴上聚拢，当它聚拢到中心轴上时，罩物质发生强烈的对碰，产生很高的温度和压力，金属罩被熔化，其中少部分物质以很高的速度沿对称轴向前方射出去，形成一股细而长的高速金属射流．石油射孔弹就是利用这股射流进行射孔．用数值计算进行射孔弹设计时，就要求数值模拟在定性和定量上都能准确地模拟射孔弹动作的全过程．很显然这是一个多介质大变形的问题．因此单纯的欧拉方法或拉格朗日方法都无法模拟这个过程，下面给出的计算结果是采用欧拉方法加上特殊的跟踪物质界面的措施得到的．

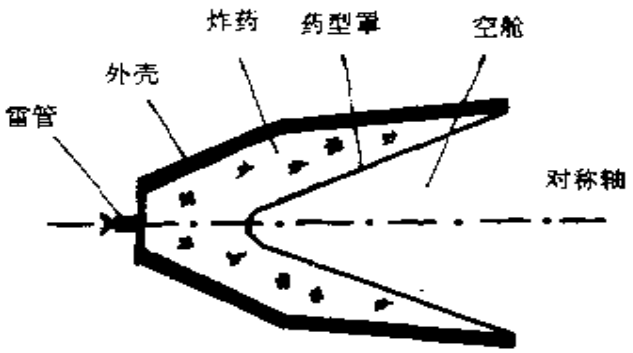


图 6－12　石油射孔弹剖面示意图

石油射孔弹的设计，就是在给定弹体体积和炸药装置的条件下，选取一种结构使射流具有尽可能大的侵彻能力．根据射流侵彻的近似理论，射流侵彻深度与射流的速度、密度、长度、断裂时间等诸多因素有关，可以简单地说，射流的速度和质量越大，其侵彻能力就越强．而又根据射流形成的近似理论知道，射流的速度和质量与炸药性能、药形罩的形状、质量、材料等因素有关．仅就药形罩形状来说，定性知道，罩的顶角越小，则射流速度越大，但弹质量却越小，所以，这两者又互相矛盾．此外，药形罩还可以做成双锥的、多锥的、曲面的等等．可见，射孔弹的设计，涉及的因素很多，只有从

定量上弄清各个因素的作用及它们相互的影响，才能作出一个好的设计.

有了数值计算手段，每做出一个射孔弹的设计方案，都能定量地算清楚它作用的全过程.具体地说，可以算出：雷管点火后炸药的爆炸过程，包括爆轰波的形状、传播速度、波后的压力、速度、密度等；爆轰波对药形罩的压垮过程，包括罩的变形折转、压垮速度、压垮角等；药形罩向对称中心聚合过程，包括聚合的速度和加速度、聚合对碰后的压力、密度、速度等；射流的形成过程，包括射流射出的速度、密度、射流半径、质量等；射流的运动过程，包括整个射流各点的速度、速度变化、拉伸、断裂等；射流碰靶和侵彻过程，包括开坑情况，侵彻速度、射孔半径、射孔深度等.而在实弹试验中，利用现有的测试手段是不可能在一次试验中取得如此丰富的数据的.有了这些结果，实际上就知道了射孔弹的性能，也就知道了设计方案达标与否.下面给出一个具体射孔弹的数值计算结果.

图 6－13 是射流形成和运动过程的计算结果的图形显示 .图中 t 代表时间，单位是 μs，雷管点火为时间零点 $t=0$；空间坐标取为柱面坐标(z,R)，z 是轴向坐标，其原点 $z=0$ 取在射孔弹的底端面上，R 是径向坐标，其原点 $R=0$ 在弹的对称轴上，坐标长度的单位为 mm .图 6－13(a)是雷管点火后不久 $t=2.39\mu s$ 的图像，爆轰波形状呈球形，已传播到了药型罩上；图 6－13(b)，炸药已全部起爆完毕，药型罩顶部已被压垮到对称轴上，射流已开始出现；图 6－13(c)，罩大部分已被压垮；图 6－13(d)，$t=20.25\mu s$，药型罩已全部被压到轴上，射流已全部形成，其头部已运动到 50mm 处.图 6－14、图 6－15、图 6－16 中给出了这时刻射流的几个重要物理量的数值结果.

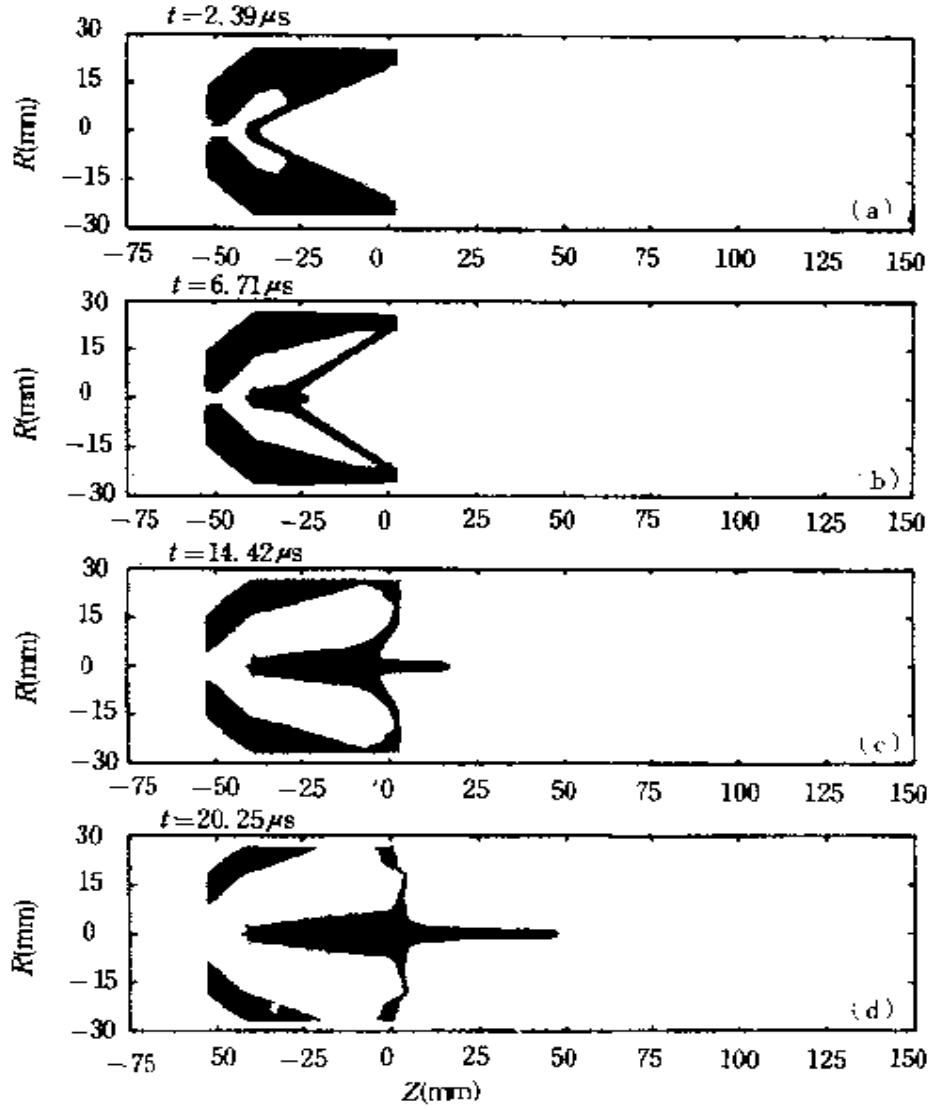


图 6－13　射流形成过程计算结果的图形显示

图 6－14 是射流各点的速度值，头部的最大，尾部的最小，这速度分布决定着射流的拉伸情况和穿靶能力．顺便说明一下，图中共列出了四个设计方案的结果，每一条曲线对应一种方案，将它们画在一张图上，便于对各方案进行比较．其中曲线 d 对应图 6－13 的方案．

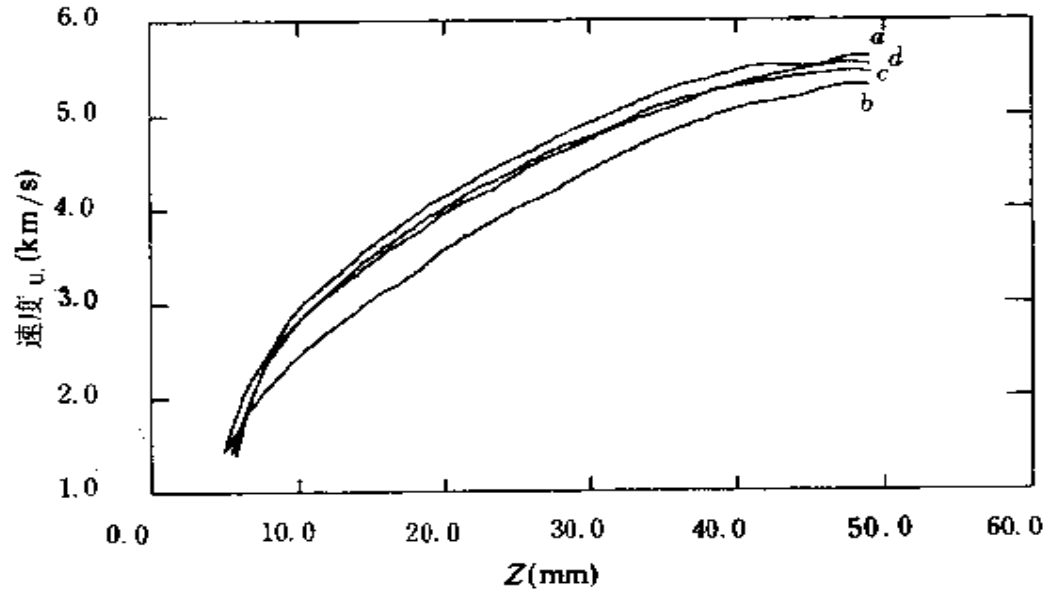


图 6－14　射流速度分布

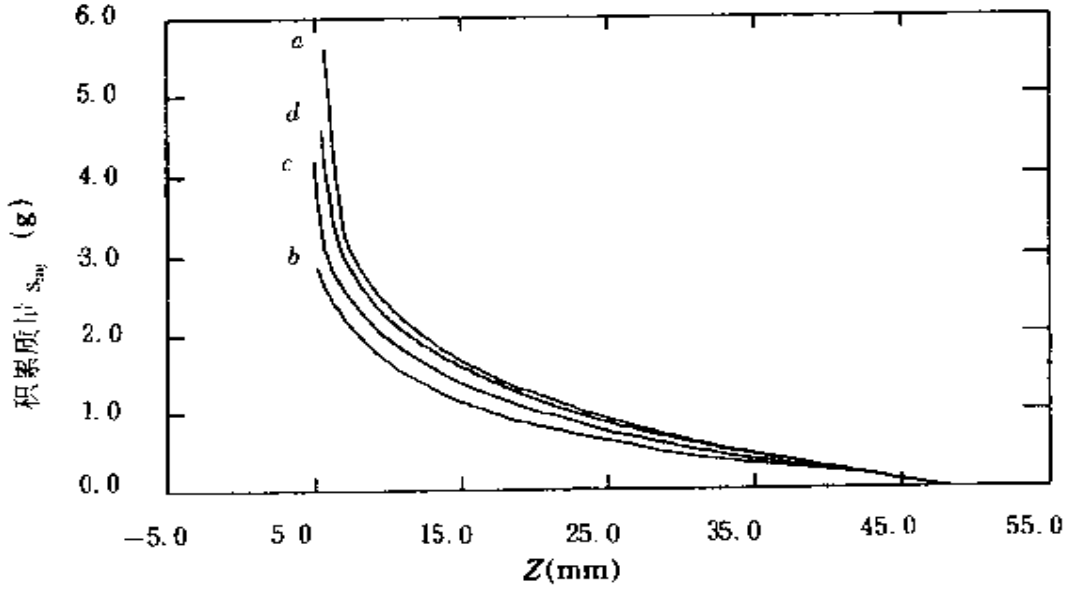


图 6－15　射流积累质量分布

图 6－15 是射流积累质量的分布曲线.所谓积累质量就是从射流头开始将各段射流的质量逐段相加所得的质量的和.例如,图上 $z=25$mm 处曲线的值就代表从射流头($z=50$mm 处)到 $z=25$mm 处这一段射流的质量.所以,射流尾端处曲线的值就是整个射流的总质量.

图 6－16 是射流积累动能的分布曲线.积累的含义同前图.射流的总动能越大,其侵彻能力就越强.

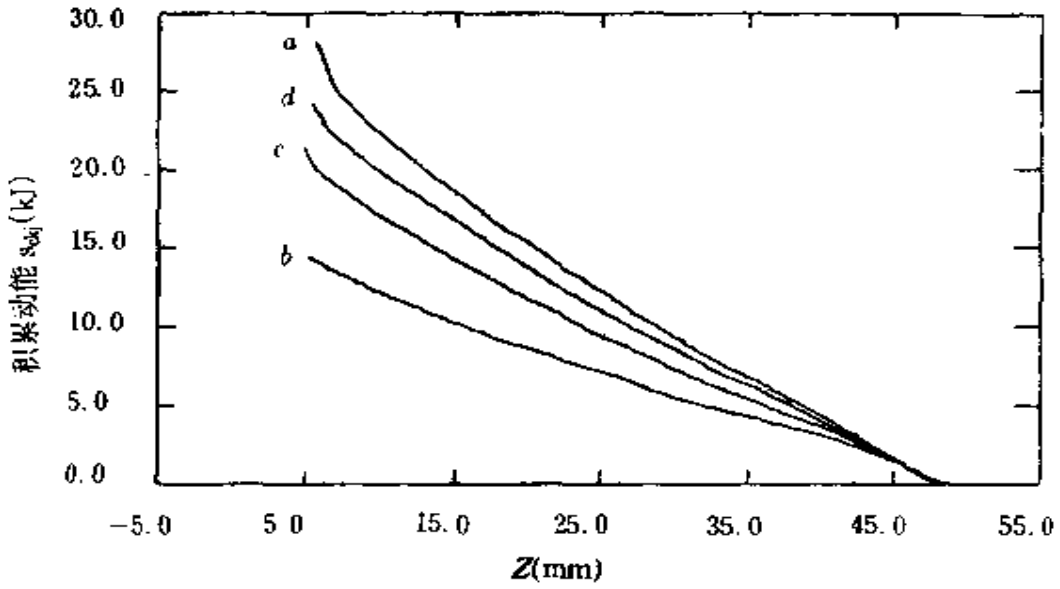


图 6－16　射流积累动能分布

从以上的图形和曲线.就可以直观地看到射孔弹作用的全过程的物理图像,并获得对射流性能的定量认识.

图 6－17 是射流碰靶后侵彻过程的计算结果的图形显示.图中的时间和空间坐标同图 6－13.B 代表靶,靶物质为低碳钢.图 6－17(a),射流A刚碰到靶.图6－17(d),侵彻过程结束.图中显

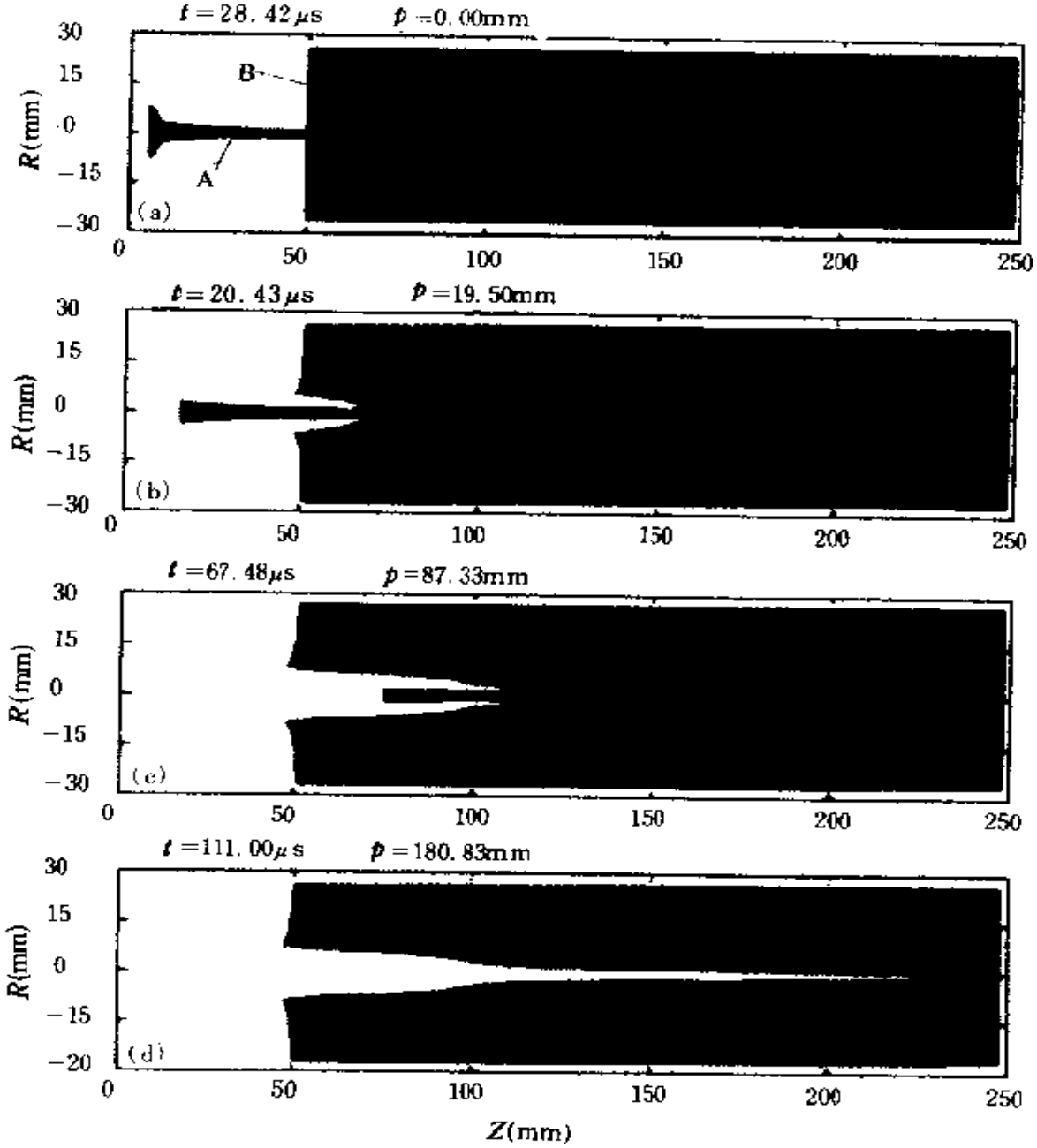


图 6－17　射流侵彻过程计算结果的图形显示

示了射流穿孔的深度、孔径和孔形.图 6－18 就是与此对应的侵彻曲线,即射流的穿靶深度随时间的变化,曲线终点的值就是该射孔弹能够穿孔的最大深度.这曲线也就回答了这个设计方案是否达到了指标要求.

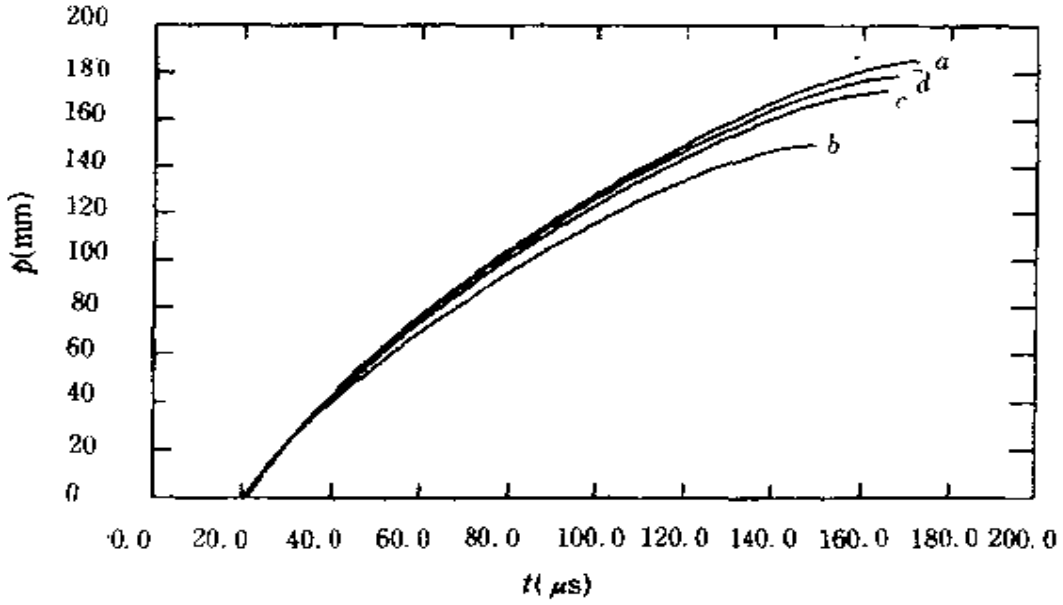


图 6－18　侵彻深度随时间的变化

7 有限元在航空工业中的应用

7.1 从波音 777 型飞机“无纸设计”谈起

美国波音公司研制生产的民航飞机 700 系列，是在世界各大洲上空和洲际穿梭飞行的主要机型．它们高亚音速的飞行速度和低成本的高有效载荷能力，使世界变小了，人们十分惬意地享受着现代航空高科技的实惠．舒适安全的空中交通，加快了国际商贸活动，便利了科技文化的交流与合作，使国际休闲旅游漂洋过海成为可能．

波音 700 系列主要机型是 737，747，757，767 和最近诞生的 777．波音 777 型不是载客量最大的飞机，却是航空高科技最高水平的一个象征，是计算机应用技术的完美体现．波音 777 型飞机是世界上首次实现“无纸设计(paperless design)技术”取得成功的航空高科技产品．这一技术标志着计算机辅助飞机设计 CAD 技术、计算机辅助工程分析 CAE 技术、

计算机辅助验证测试CAT技术、计算机辅助制造与系统工程的组织管理CAMM的技术已经提高到一体化的全系统性的综合优化水平，对显著缩短设计周期、降低设计制造成本有重大意义。

飞机无纸设计技术这一辉煌成就，像原子弹爆炸一样，产生了巨大威力，但这一技术发展最初的原动力，却是本文要介绍的有限元方法，它是这一爆炸式技术发展过程真正的引信。40年前，有限元法作为飞机结构设计的新方法在航空工程中诞生，成为航空计算机应用技术的首次突破，推动着飞机计算机设计综合优化技术的发展，导致无纸设计这一重大成果的最终到来。尽管40年前，没有人预见到会出现这一光辉前景。

有限元分析的方法和理论在飞机结构设计、结构强度分析取得的成就，大大刺激了计算机在飞机设计其他领域的应用。与有限元方法有共同血缘关系，用作亚音速和超音速线化理论的面元法，很快成为飞机气动特性计算分析的第一代方法。由此导致一个新学科——计算流体动力学和称之为"数值风洞"的飞机粘性绕流数值模拟技术的发展。有限元的结构强度计算和飞机气动载荷分布和气动特性计算，对于飞机几何外型数据的要求和离散网格数据的自动生成需求，催促计算机辅助几何设计技术在航空高技术领域首先诞生。于是有今日CAD技术的普遍推广应用。今天，在计算机网络系统上，进行着像飞机这样复杂的系统工程的整个设计、设计分析与优化设计技术的采用、用多种试验数据的设计验证、新一轮设计的叠代循环，以及设计过程的控制与管理，甚至提供数字控制机床加工所需要的全部数控编程的数据和指令，构成十分复杂的、计算机设计制造的集成系统。

40年对于人类历史的长河不能算长，但是设计技术的日新月

异、突飞猛进的变化却十分惊人。不用制木质或者金属模型，在实物还没有被制造出来之前，能够在计算机的荧光屏上全貌展现飞机的外观和内部装置陈设，能够展现全部零件或部件的几何形状及其尺寸比例、机翼、机身的结构型式及其实现预装配的过程、起落架的收放、门窗开启、各种传动机构活动是否符合设计要求，飞机起飞降落特性、滑跑距离、操纵稳定、突风响应的阻尼、经济性、安全可靠性以及可维修性等等，所有的设计关心的方面，各种各样的问题，都在计算机上对虚拟飞机的数字模型逐一获得圆满解答，真是不可想象！这一切固然是计算机数据处理能力、数据存储能力、数字图像显示能力的伟大表现，但同等重要的，这无疑也是人类将设计计算机化这一伟大科技思想的胜利。今天，事实作出宣告：单纯依靠实验一种手段对设计进行验证确认的时代已经结束！与实验方法一起，同时采用计算方法的时代已经开始！世界上所有的设计办公室，都在用计算方法进行设计，用计算的方法进行设计分析，用计算的方法对系统进行模拟仿真，用计算的方法确认设计。有限元方法诞生之日，是这一时代的真正起点。

7.2 有限元方法——飞机设计首次应用计算机的伟大实践

航空航天飞行器提高有效载荷比重，严格限制自身结构质量增加的设计要求，呼唤新的设计技术的诞生。本世纪50年代中期，计算机在飞机结构设计上的应用，导致有限元方法及其技术的诞生和发展，是本世纪对世界经济影响最重大的科技事件之一。

航空航天飞行器在不同的、载荷谱变化很大的载荷和严酷的

环境条件下工作，经受着复杂的应力应变、振动与动力响应、加载卸载的力学过程，出现了各种多样的数学力学问题需要定性定量分析。结构的承载能力、设计要求的强度和刚度指标、断裂疲劳寿命、结构的可靠性和耐久性，皆与这些分析密切相关。分析的深度和定量数据的准确性，既决定飞行器的品质和性能，又直接影响设计与制造成本，更是关系人员和承载货物安全的根本因素。

在计算机出现之前，这些问题的解答，除了将实际问题简化再简化，将飞机机身看作一个圆柱，将几何上和受力状态十分复杂的三维问题简化成二维问题，甚至一维问题进行理论分析之外，主要依赖各种形式的结构试验。这一类实验项目很多，周期很长，有部件的和全机的静力试验、破坏实验和千百万次的疲劳实验，有常温下的试验，也有高温及其变化产生的热载荷作用下的实验。由于这些实验基本上是全尺寸的，因此实验规模之大，花费之高，周期之长是可想而知的。一架好端端的价值几百万美元的飞机在巨大的全机静力试验厂房内，通过液压传动的协调加载设备，在一声巨响之下，将其拉得支离破碎，真是十分可惜！但这种破坏试验又不得不做。

这一切在计算能力强大的电子计算机出现之后，发生了根本性的改变。实现这种革命性变革的，仅仅是一种计算机数值方法，它就是能够充分发挥计算机的能力，经十多年发展就成为强有力设计手段的有限元方法。

50年代中，美国飞机设计工程师M. J. Turner等与大学教师R. J. Clough一起合作在《J. Aero. Sci》(1956)上提出的飞机结构分析的直接刚度法和欧洲的Argyris教授创导的飞机结构分析的矩阵分析方法，被认为是孕育当代有限元法诞生的起点。作为一种解

算数学物理问题的近似方法，这一方法的原型可以在著名大数学家德国人柯朗(R. Courant)40年代发表的论文中找到，但是由于当时计算机尚未发明，柯朗的方法因计算量太大并未引起科技界的重视.

我们中国科学家对于这一方法的创导和发展具有不可磨灭的贡献.60年代初，以冯康教授为首的中国科学院的研究集体在用计算机做水坝和薄壳屋顶的应力应变分析过程中，克服了传统方法的种种缺陷，冯康教授独立于西方，创造性地提出基于变分原理的差分方法，这个方法与西方称为有限元的方法完全一样，他不仅提出方法，而且在世界上最先给出这一方法的可靠性理论，发表了世界上第一个有限元近似解收敛性的数学分析和有限元近似解的最优误差估计，开辟了有限元数学的新篇章.

7.3 有限元数学模型——模拟复杂结构受力变形过程的技术基础

飞行器结构的自身质量与结构承载能力构成尖锐矛盾，正如常识告诉我们的，结构越结实，承载能力越大，但越结实也就意味着"粗和大"，因之自身也就越重，对上天飞行的航空器，甚至要摆脱地球引力的作用，进入太空，这一矛盾就显得十分突出.因此，像鸡蛋的壳一样，薄壁结构成为航空航天飞行器结构的基本形式.除了选用或研制品质更好的材料之外，使有一定承载能力、质量尽可能轻的结构优化设计成为航空航天器设计的一项重要任务.

飞行器要能飞，必须有气动升力，前进速度要快，阻力要小，操纵性和稳定性要好，就必须有良好的气动外形，因此飞行器绝不可

能像鸡蛋那样简单，薄壁必须有加强的肋与筋，而且受力状况也复杂得不能比拟。在几何约束、应力应变多种约束的限制之下求解一个最优化问题，飞行器结构优化设计的难度也就可想而知了。

将结构设计引入新境界的就是有限元方法。就像万物多种多样，都由原子、电子和中子组成一样，有限元方法充分利用了这种复杂性和简单性对立统一的普遍原则，把复杂的结构传力受力变形统统归结为几何上简单、应力应变单纯的有限元素的研究。每一模式构成一特别的力学上理想化的有限单元体，它们一般是只受一维力的拉伸杆件，受弯矩的直梁或曲梁，只承受剪力的纯剪板件，以及也受拉应力的薄膜元素等。把复杂结构统一地理想化为这些虽然数目可能很多，但总是类型有限的单元体的一种装配。按力学应力平衡方程装配起来的有限单元体的系统，被当作为复杂结构应力应变分析的一个近似的数学模型，是有限元法化繁为简指导思想的根本，也是它取得成功的基本出发点。由于几何形状简单，受力变形单纯，每一单元模式的应力应变关系可用有限的几个节点位移通过单元刚度矩阵直接地表达出来。而且尽管结构的受力和传力变形可以十分复杂，多种多样，但是每一局部的小范围总可近似为上述模式有限的几种单元体的组装。加之这种按力平衡原理的总装数学上又等价为单元刚度矩阵的某种方式的叠加，于是一个复杂的应力应变分析能够归结为一个线性代数问题，或非线性代数问题的解算。正是上述这些特点，使我们能够明白为什么有限元法起初被称为直接刚度法或矩阵方法。

上面介绍的是结构力学家发明的有限元法的思维路线。数学家从不同的角度，不同的路线，也达到了有限元法的同一方法型态。冯康教授考虑的是一组描写物理力学问题的偏微分方程的近

似解.将这些偏微分方程,用泛函分析的方法转换为求能量极小的变分学问题.再将所有可能的场变量局部近似,在场内的每个简单的三角形或四边形区域上,假定场变量有低阶多项式的近似表达.无穷自由度的问题近似成有限自由度的 Ritz-Galerkin 问题.这样的思想几乎同时也为应用力学家所领会.

有限元方法的这种不同侧面表象,深刻地揭示了它的内在本质.结构力学家使有限元法像桁架结构一样简单明了,使有限元分析计算过程像机械动作一样,计算机程序过程统一常规,千篇一律,十分便于发展通用的有限元计算机软件系统.数学家揭露了它的普适性,它的 Ritz-Galerkin 方法的理论基础.于是不限结构力学问题,作为一种强有力的数值方法,有限元法很快在其他许多行业和学科领域得到了广泛的应用、研究和发展.

7.4 四十而不惑,有限元数学功盖四方

有限元方法从 1956 年诞生到现在,经历 40 年的研究发展,实实在在地进入了不惑之年.今天,在航空领域,用有限元方法进行着全机、部件的静力动力分析,非线性的弹塑性分析,薄壁结构的屈曲失稳分析,炮击与鸟撞的动力响应分析,着陆响应分析,机翼结构在气动力作用的气弹颤振分析等等.既用以分析金属材料结构,也用以分析复合材料的机翼结构;既用作强度评估,也用作优化设计.不仅对于飞机设计十分重要,对于发动机结构设计更是不可或缺.发动机的关键零件——转子叶片应力集中与凹角部位高应力区的强度预估、裂纹尖端的疲劳载荷作用下的扩展规律、发动机转子高速转动的动力特性分析、在千度以上高温环境下工作的

发动机燃烧室结构的温度场与热应力分析，有限元方法都在施展威力。有限元方法技术经过40年的发展，已有相当完整系统的求解策略和适合不同情况的解算方法，发展了以NASTRAN有限元程序为代表的、成熟的计算机软件系统技术，理论基础牢固，学科系统宏大，内容丰富深刻，集应用力学和计算数学之大成的科学理论已经建立，经系统性总结，像《四库全书》那样庞大的《有限元方法手册》，在一大批专家群体共同编撰之下已经在西方出版。总之，有限元方法和技术已经成为当今科学与工程分析的重要方法，计算机模拟仿真技术的基本理论、设计技术的主要手段。它早已跨出航空航天，进入机械、桥梁土木建筑、造船等行业并获得广泛应用，而且从传统的结构力学、固体力学领域向流体力学等各方面渗透，不仅用于研究物质机械运动的规律，还用于研究热运动和电磁运动的规律。在今日高科技成就辉煌的殿堂中，有限元方法已经确立其科学巨人的不朽地位。

有限元方法的科技地位，由力学家和工程师奠基，由计算数学家巩固，理论科学与实用技术共同塑造它的形象，增强它的智慧，培育它的能力，有限元方法才有今天。大科学、大群体，大协作创造出大成果的当今高科技发展的这一社会化特征，在有限元方法的发展过程中，同样表现得十分明显。毫无疑问，除其他领域科技专家的贡献之外，有限元方法之所以能取得伟大的成就，计算数学家的功劳铺盖四方。

有限元模型是理论模型的一种离散近似。近似必然产生误差，各种误差必然对有限元方法分析计算、评估预测结果的正确可靠性产生重要影响。这些误差是多种多样的。有限元模型的物理力学近似不可能没有误差，载荷施加和边界条件计算处理导致误差，计

算求解有误差.正是数学家完成了系统性的数值误差分析,揭露了有限元计算精度依赖于有限元形状函数阶次的数量关系,揭露了这些误差因素对最终结果精度产生何等程度的影响,才使有限元计算分析从经验的水平上升到理论的水平.由我国数学家冯康教授开创的这一研究,历时十年,捷克、美国、法国等许多国家的专家学者广泛参与,最终断定力学家提出的分片等参数插值多项式的有限元位移模式的逼近论性质、逼近度、有限元几何尺寸和多项式阶次的关系,使有限元方法出现质的飞跃.

为了使有限元数学模型充分刻画问题的物理与力学的特性,使有限元模型既可靠又简单,力学家从不同的力学原理出发,提出了多种富有特色的有限元模型.它们是假定位移的最小势能模型,假定应力近似的力平衡模型,既假定应力又假定位移的混合/杂交模型,以及变分犯规的非协调位移模型等.各种模型的数值试验揭示,不同模型的计算结果有不同的误差收敛与发散的性状规律.为改善协调位移模型使结构过于刚硬的缺陷,非协调位移模型采用变分犯规措施进行调节.但并不都一定奏效,有时事与愿违,得不到正确的计算结果,使计算简单的非协调有限元模型的应用蒙上了阴影.于是力学家建议,采用所谓 patch test(补丁检查)模型可靠性的检验措施予以补救.但是这些有力学依据的必要条件,最终成为保证分析一定正确可靠的充分条件,也经历了较长期的努力,由许多数学家做了大量的研究工作才告完成.在这一方面,我国数学家的工作有重要意义,石钟慈教授对非协调元方法的系统性研究就是一项世界公认的成就.

数学家为杂交/混合元模型建立理论基础,克服有限元方法处理不可压缩材料面临的收敛闭锁障碍,使有限元从固体力学进入

流体动力学等更宽广的领域，则更是大功一件．从此有限元法的发展登上了新的高峰．

基于广义变分原理的混合/杂交有限元模型，力学上直接而清楚地刻画应力、应变、位移协调各种物理与力学的关系，应该是更合理的有限元模型．但是它的鞍点问题特性，使问题复杂化了．对一个混合/杂交有限元模型的假定应力模态和假定位移模态不加限制，即便表面上是自然可取的，其结果未必是正确的．不加限制，不是刚体位移的虚假零能模态有可能出现．自然不能保证有正确的计算结果．是意大利计算数学家 Brezzi 借助泛函分析的数学理论，建立了解决上述问题的理论框架．从原则上规定了混合/杂交元模型的假定位移和假定应力模态必须满足的充分必要条件．为混合杂交元方法健康发展、广泛应用开创了一条广阔平坦的道路．自此，有限元成为处理对流扩散运动的一种有力方法，使 Navier-Stokes 方程的有限元模拟出现新局面．

在有限元模型分析方面，数学家借助广义函数和索波列夫空间理论，偏微分方程的 Hilbert 空间方法等基础理论，不仅完成了上述有关有限元近似的基础理论问题，阐明了数值上奇异的种种现象的内在依据，有十分重要不可磨灭的贡献，而且在有限元计算方面，诸如算法设计、算法可靠性、精确性分析和算得更快，有更高计算效率方面，在充分利用计算机硬件资源，更充分地发挥各部件的潜力方面，数学家更是责无旁贷而且驾轻就熟，使大规模的有限元分析计算效率不断地成倍增长．

有限元分析最基本的环节是求解结构刚度矩阵为系数的线代数方程组．对于飞机全机结构分析，它通常是上万阶的对称正定稀疏矩阵．算法的优劣，不仅影响计算结果有效数字的多少，而且计

算机 CPU 的时间长短，内存储量的大小，内外存储交换的频率差异都可能很大.这一方面的研究，经数学家的努力，已经上了两大台阶，从变带宽的矩阵分解直接消元法，发展到矩阵条件数预处理共轭梯度叠代方法，最终有今天的多重网格叠代方法.计算量从 $O(n^3)$ 的量级降低到 $O(n)$ 的量级. 今天，人们普遍认为计算机处理器速度性能，每五年提高 10 倍，是计算机发展最快的一个性能指标.对于矩阵阶数 $n = 10000$ 的问题，计算方法改进所达到的计算速度提高，似乎并不比计算机硬件速度改进慢多少.

像一个细长体悬臂梁式的飞机结构，振动模态计算是有限元动力分析的基础，是设计十分关心的一项主要数据.这一问题归结为数值代数中的广义特征值和特征向量问题的计算.计算数学在这一方面对有限元分析作出的支持，同样也是强有力的.从最小振动频率、次小频率、一个接一个精确叠代计算的一整套方法已经齐备，在实际设计分析中发挥着基本的作用.

有限元方法是计算的方法.计算数学对有限元有多方面的贡献，对所用的算法会有许多重大的改进和发现，会是十分自然的.没有有限元数学，就不会有有限元方法的今天，应该是绝对正确的.

7.5 更上层楼，有限元法迎接新挑战

“欲穷千里目，更上一层楼.”有限元方法尽管势力范围很大，新近又进入加工制造领域，在金属板材塑性成形加工中成为控制手段，为航空航天器复杂几何零部件加工发挥新的作用.但是放眼科技世界，面对新形势下新的要求，一个新的历史阶段已在面前，需要跨越.

有限元计算需要的计算网格可以是非结构性的.计算节点彼此之间的联结,不必像矩形网格那样规则呆板,可以有更大的适应性,满足数量巨大,而且非均匀不规则的网格在计算机上自动生成的要求.客观现象因几何复杂性和物理力学的复杂性产生计算场变量,在一些特殊区域剧烈变化,导致必须用非均匀不规则网格,模拟这种复杂的运动变化.正是有限元法的这一特点,为大规模科学与工程计算、复杂问题模拟仿真所青睐.因而使它肩负重任,必须迎接新的巨大的挑战.

模拟的问题越来越复杂,计算规模越来越大,粗网格高精度是自然的要求.为了显著节省计算工作 CPU 时间,在场变量变化剧烈区,计算节点密集;非剧烈变化区,计算节点布置稀疏,但网格疏密过度导致网格几何畸变.为了计算的自适应,要求有限元模型对此是几何不敏感的.对于传统的有限元模型,计算发现的另一个问题是收敛性闭锁现象发生.对于接近不可压缩材料和考虑横向剪切影响的模型参数依赖的问题,采用常规的有限元模型,表面上看似乎十分自然,实际上导致计算的发散性.这种所谓 locking 现象是有限元"鲁棒性"(robustness)差的重要表现.这些问题逼使人们作出对于新的比传统格式性能更强的有限元离散格式的追求.

基于不同变分原理构造不同的新的有限元离散格式的研究,可以追溯到有限元的童年时期,一直是有限元发展的一个重要的原动力.但是,按上述性能标准,品质优良的所谓最佳的有限元格式的研究只是新近逐渐引起关注的重要课题.

美国 Stanford 大学的 Simo 教授等发现[见 Inter. J. Numer. Meth. in Eng., Vol 29, N:8(1990)],一个通称为 Wilson 有限元的非协调格式,除了 20 年前计算上发现的高精度之外,尚有不会

locking 的优良特性. 于是开始了使用新型有限元位移插值的所谓增强格式 enhanced schemes 的探索. 这种新的有限元基函数的构造特点是着重增加单元内部自由度而不顾及基函数的协调. 协调性损失换来内部自由度经"静力凝聚"局部消除,计算效率显著提高. 因此它们对在有限元商业软件系统中至今还占统治地位的等参协调插值,提出了最大的挑战. 人们有理由怀疑:等参协调插值是有限元位移近似的合适选择吗?

尽管有关的数学理论分析仅仅是一个开始,但是 Wilson 插值法由于不顾及协调性,从内部增加自由度提高逼近度的观点,从计算效率看,是有限元插值观念的一种思想解放. 从这一方向提高有限元分析效率有着十分诱人的前景. 例如三维分析现在普遍采用 20 节点等参元,但如果 Wilson 插值的三维元素的能量误差也有 $O(h^2)$ 阶精度,那么(8/8 + 12/4) = 4 倍以上的效率提高可以实现! 问题是:Wilson 插值的星星之火,可以燎原吗? 问题的圆满回答,意味着真正意义的第二代有限元法的诞生,因此应该是有限元方法面临的一个大挑战!

这里所谓第二代有限元,是指支持 Wilson 型非协调插值的、变分不犯规的普遍可遵循的有限元格式. 究竟何种变分原理是充当大任的理性选择,尚不得而知. 但美国 Stanford 的力学家基于胡海昌 - Washizu 变分原理在探索. 包括著名美籍华人科学家、美国工程科学院院士卞学锁教授等许多国内外专家学者过去和现在从不同角度、不同方面有所考虑,所提问题绝不属于海市蜃楼虚无飘渺一类,倒是可以完全肯定的.

增加内部自由度有利于 locking 不稳定性的消除,在数学家的多种研究中,以不同形式,作为一个共同规律,有多方面的十分鲜

明的证据.无论对于粘性不可压缩流动,还是 Mindlin－Reissner 板弯问题,著名的气泡函数(bubble)就是内部自由度增加稳定性的典型措施.混合元模型在它们的配合下对 locking 不稳定性的强阻尼作用已经引起数值的和理论分析的广泛研究.因此,多个方面分析,用内部自由度丰富型非协调插值方法,取代等参协调插值在现今有限元方法中的地位,似乎是顺理成章的.只是"非协调"不是一个容易驾驭的问题.什么样的变分表达,是支持这类插值近似位移的理性选择,前景尚有一层迷雾.要证实它们的高精度、几何不敏感和不 locking,困难就更大.上述种种猜测既需要数学上严格的理论分析予以证实,又需要多种数值试验的确认,更需要专家的想像力和智慧,有许多问题等待着数学家和力学家再次携手协作、共同奋斗.

不只上述问题,当有限元方法进入"五十而知天命"的岁月时,只要是问题,个个都富有挑战性.例如在一个复杂的载荷谱作用下,飞行器结构不可能处处用线弹性的有限元模型表达,可以在不同部位同时出现性质不同的,线性与非线性的应力应变状态.线弹性的应力应变区,非线性塑性变形区和屈曲失稳组合出现客观上更为真实.这样的增量过程规模大,计算时间长,需要存储的信息多,特别需要强度试验数据的支持和配合.一个可靠的有限元模型和可行的解算策略才可能获得.诸如此类的问题,需要有限元方法变得更强,算得更快,本领更高.经过 40 年的发展,有限元法已经开辟出一条康庄大道,这"更强、更快、更高"的目标,相信一定能够实现!

8 科学计算与能源

8.1 能源问题离不开科学计算

石油是国民经济和社会发展的重要支柱.石油开发大体经历三个阶段:一次采油阶段:依靠地层的压力和能量自喷原油,采收率为10%～15%;二次采油阶段用高压泵向油藏注水,维持地层压力,将原油驱出,采收率为10%～20%;三次采油阶段依靠物理、化学和生物方法进行强化采油.目前,世界各国提出了多种三次采油方法,实验表明,采收率可提高7%～50%.

我国开发的油田均进入了二次采油期,大多数已进入注水开发中后期,特别是大庆油田和胜利油田,若继续单纯采用注水开采,产量每年将减少数百万吨.稳定石油产量的唯一方法是采用三次采油新技术,开发尚滞留在地下约50%以上已探明的储量.若能平均提高30%的采收率,即相当于再生了

同等规模的油田.

所谓油藏数值模拟,就是用电子计算机模拟地下油藏十分复杂的化学、物理及流体流动的真实过程,以便选出最佳的开采方案和监控措施.对于三次采油新技术,特别需要注意驱油剂与地下油、气、水油藏的宏观构造及微观结构的配伍性,考虑化学药剂的用量和能量的消耗.近年来,随着电子计算机计算速度和能力惊人的增长,油藏数值模拟的适用性越来越普遍,模拟结果越来越真实,即使对极其复杂的油藏情况,也获得了巨大的成功.油藏数值模拟已成为石油开采中不可缺少的重要环节.

油藏数值模拟的实现要经过四个主要阶段:第一,物理模型的建立,它能真实地反映油藏内流体流动的基本现象;第二,物理模型的数学形式,即数学模型的建立,通常为一组耦合的非线性偏微分方程组的初边值问题;第三,当研究并了解数学模型解的存在性、唯一性和正则性之后,再构造其离散格式,即所谓数值模型,并研究它的收敛性和稳定性;第四,研制高效的软件程序.一个有效的工程模拟软件并不是一次可以完成的,当程序完成后要对各种简化模型问题进行试算,其数值结果必须与物理采样进行比较,效果不好就要重新修改校正上述诸过程.

关于数学模型,当多相流体在多孔介质中流动时,流体要受到重力、毛细管力及粘滞力的作用,且在相与相之间可能发生质量交换.因此用数学模型来描述油藏中流体的流动规律,就必需考虑上述诸力及相间质量交换的影响.此外还应考虑油藏的非均质性及几何形状等.

实际描述油藏流体的流动规律:

(1)有描述油层内流体流动规律的偏微分方程组以及描述流

体物理化学性能的状态方程.

(2)给出定解条件.

建立数学模型是以下述物理原理为基础:质量守恒原理、能量守恒原理、运动方程和状态方程.根据这些原理建立数学模型的步骤:

(1)确定所要求问题的解,通常需求地层压力 p 和达西速度 U 的分布,多相渗流时饱和度 c 的分布,这是最重要的.

(2)确定未知量和其他物理量之间的关系,达西定理和流体状态方程是建立渗流数学模型所必需的方程,在考虑非恒温渗流时还需能量守恒方程.

(3)根据物理条件写出定解条件.定解条件包括:区域的几何形态、有关物理参数和系数、描述初始状态的初始条件和区域的边界条件.

油藏数值模型是首先将非线性偏微分方程组的初边值问题转换为有限元格式或有限差分格式,然后将非线性系数项线性化,从而得到线性代数方程组,再通过线性代数方程组数值解法,求得所需的未知量:压力、达西速度、饱和度、温度、组分等的分布和变化.

经数十年的迅速发展,目前油藏数值模拟的理论、方法和应用,已从油田开发发展到油气资源评价,油田勘探和环境科学等重要领域.在这里重点介绍二相渗流驱动、三次采油、油气资源评价、核废料污染、海水入侵的预测和防治、半导体器件的数值模拟的发展状况,着重介绍问题的实际背景和数学模型,并指出求解这些问题的最新数值方法、理论分析和计算机软件.

8.2 油水二相渗流驱动问题

用高压泵将水强行注入油藏，使其保持油藏内流体的压力和速度，驱动原油到采油井底，称为二次采油．可分为不混溶、不可压缩油水两相驱动问题；可混溶、不可压缩油水两相驱动问题；可压缩相混溶的驱动问题．其数学模型是关于压力的流动方程和关于饱和度的对流扩散方程．二相渗流驱动问题的数学模型、数值方法和工程应用软件是能源数学的基础．

对于考虑毛细管力、不混溶、不可压缩的油水两相驱动问题，具有重要实用价值的是特征差分方法和特征有限元法．问题的数学模型由流动方程和饱和度方程组成．

这类问题已有许多近代实用的计算方法和工业应用软件，通常二维问题多采用有限元方法，三维问题则多采用差分方法．由于它是能源数学的理论基础，这门学科发展很快，新的数值方法不断涌现，工业应用软件不断更新．

有了数学模型和计算方法后，就要建立计算机模型，也就是将各种数学模型的计算方法编制成计算机程序，以便计算机进行计算得到所需要的各种结果．工业性的计算机模型也称计算机软件，它包含图形和数据的输入和输出，各种数值解法等，可应用于各种油田开发的实际问题．

由于所要解决的问题是多种多样的，因此还要根据所要解决的问题进行历史拟合和动态预测．历史拟合即是用已知的地质、流体性质和实测的生产史输入计算机，将计算结果与实际观测和测定的开发指标相比较，若发现两者有较大的差异，则需修改输入数

据，使计算结果与实测结果一致，这就是历史拟合．动态预测是在历史拟合的基础上对未来的开发指标进行计算和预测，这里实质上需要对反问题进行研究和分析．

8.3 化学驱油（三次采油）新技术

目前国内外行之有效的保持油藏压力的方法是注水开发，其采收率比靠天然能量的任何开采方式为高．我国大庆油田在注水开发上取得了巨大的成绩，使油田达到高产稳产．如何进一步提高注水油藏的原油采收率，仍然是一个具有战略性的重大课题．

油田经注水开采后，油藏中仍残留大量的原油，这些油或者被毛细管力束缚住不能流动，或者由于驱替相和被驱替相之间的不利流度比，使得注入流波及体积小，而无法驱动原油．在注入液中加入某些化学添加剂，则可大大改善注入液的驱洗油能力．常用的化学添加剂大都为表面活性剂、醇和聚合物．表面活性剂和醇主要用于降低地下各相间的界面张力，从而将被束缚的油启动；聚合物被用来优化驱替相的流度，以调整与被驱相之间的流度比，均匀驱动前缘，减弱高渗层指进，提高驱替液的波及效率，同时增加压力梯度等．

问题的数学模型基于下述的假定：流体的等温流动、各相间的平衡状态、各组分间没有化学反应以及推广的达西定理等．据此，可以建立关于压力函数 $p(x,t)$ 的流动方程和关于饱和度函数组 c_i 的对流扩散方程组，以及相应的边界和初始条件．

多相、多组分、不可压缩混合流体的质量平衡方程是一组非线性耦合偏微分方程，它的求解是十分复杂和困难的，涉及到许多现

代数值方法(混合元、有限元、有限差方法、数值代数)的技巧. 一般用隐式求解压力方程,用显式或隐式求解饱和度方程组.通过上述求解过程,能求出诸未知函数,并给予物理解释.分析和研究计算机模拟所提供的数值和信息是十分重要的,它可完善地描述注化学剂驱油的完整过程,帮助更好地理解各种驱油机制和过程.预测原油采收率,计算产出液中含油的百分比以及注入的表面活性剂、聚合物的百分比数,由此可看出液体中组分变化的情况,有助于决定何时终止注入.测出各种参数对原油采收率的影响,可用于现场试验特性的预测,优化各种注采开发方案.化学驱油数学模型的建立、计算机应用软件的研制、数值模拟的实现,是近年来化学驱油新技术的重要组成部分,受到了各国石油工程师、数学家的高度重视.

在三次采油中,还有一类常用的方法即注气驱油,例如注天然气,注富气,注 CO_2 气体等.由于气体的进入,给数值模拟带来困难.现已有工业软件应用于生产,其中以混合元方法研制的工业应用软件最为有效.另一提高采取率的开采方法是热力采油,即向油藏供给热能,具体方式为火烧油层或注蒸气.当蒸气注入后,热量便带入油层,一方面使油的粘度大幅度下降,从而大大增加原油的流度,改善流度比,另一方面原油受热后发生体积膨胀,而减少最终的残油饱和度.火烧油层则是在注入井注入空气在井下点燃,使油层内形成一个狭窄的高温燃烧带,在前沿推进过程中、废气、水蒸气、气相烃类和凝析油会发生局部混相,产生混相驱油的作用.

8.4 油气资源评价模拟系统

为了满足对石油产品日益增长的需要，必须增加更多的油气贮量，这就要求加强石油勘探并提高其成功率，发现新的油气田和对已勘探的区域作出重新评价。盆地模拟技术就是新发展起来的对油气资源进行定量评价最有效的方法之一，它综合了石油地质、有机地球化学、快速电子计算机和计算技术等最新成果，对与生成油气有关的古温度、压力等物理量在计算机上做时空概念下的动态模拟，从而进一步研究油气的生成、运移和聚集。这对保证石油工业的稳步发展有重要的经济价值。一个完整的盆地模拟系统，由下述五部分组成：

(1)地史模型。是采用地史模型重建含油气盆地的沉积埋藏史和构造发育史，它是盆地模拟的基础，其精度直接影响盆地数值模拟结果的精度。

(2)热史模型。通过建立热史模型、重建盆地的古热流史和古温度史来确定有机质热成熟度，它是盆地模拟的关键，也是计算生烃量的主要依据。

(3)生烃史模型。在地史模型和热史模型的基础上，恢复含油气盆地的烃类成熟历史和生烃量史。油气生成理论认为，有机质成熟并转化为烃类的过程主要受温度和时间这两个参数的影响，因此可以利用反映时间和温度综合结果的时温指数 TTI 来反映有机质的热成熟度，最终确定烃类成熟历史。

(4)排烃史模型。是建立在地史、热史和生烃史三个模型基础上的，运用沉积压实原理和渗流力学方法来确定盆地排烃历史。

(5)运移聚集史模型(油、气二次运移聚集史).它是重建油气盆地运移聚集历史,也是油藏资源评价最重要最困难的部分.在油气贮量评估的基础上进一步确定油藏的位置,对寻找新的油田和在油田勘探领域具有十分重要的价值.它是建立在现代渗流力学和地质学的基础上,考虑了浮力、地下水动力、毛细管力等驱动力,并需考虑断层、通道等地质情况,来模拟油气运移聚集历史.

盆地发育资源评估部分的数学模型是由压力函数 $p(x,t)$ 的流动方程、古温度 $T(x,t)$ 的对流扩散方程和空隙度函数 $\phi(x,t)$ 的一阶微分方程组成.在盆地孔隙内的流体仅有"微小压缩"的情况下,其数学模型是由压力方程、古温度方程和孔隙方程组成的.

油气运移的过程是油气从低孔、低渗生油层运移到相对高孔、高渗的运载层,最终在储集层中可能形成一个集中的烃类聚集.初次运移是指从低孔、低渗生油层运移到相对高孔、高渗地层,其最大运移距离可达数 km.油气二次运移是指继初次运移之后,油气通过高孔、高渗运载层运移,包括在运载层运移,油气沿断层、裂缝通道运移,若遇到合适的油藏构造,油气聚集就形成油藏,其最大运移距离可达数万 m.

原油和地层水在地层中运移主要是一种渗流过程,油势场和水势场控制着石油和地层水渗流的运动方向和受力大小,它的流动速度遵循达西定律.主要驱动力有:沉积压实、排烃、流动势、浮力和毛细管力.

油、气初次运移主要发生在垂直方向,其二次运移不仅发生在垂直方向,也着重发生在横向方向,因此必须采用三维盆地模拟.问题的数学模型是一组非线性偏微分方程的动边值问题,定义区域随盆地的沉积、压实过程而变动,被模拟的盆地具有面积大、地

层厚、沉积发育时间长等特点，因此这类问题的数值分析研究也有十分重要的意义．由于盆地模拟从石油地质机理出发，是一项综合研究地史、热史、生烃史、排烃史和运移聚集史的系统工程，采用的是现代电子计算机、计算数学(有限元方法、差分方程、算子分裂法)和应用软件的最新成果，从而对石油地质的定量和计算机化起着重大的推动作用，预计在近年内将会出现重大的进展和突破．

8.5 核废料污染问题

核废料深埋在地层下若遇到地震、岩石裂隙发生时，它就会扩散，因此研究其扩散及安全问题是十分重要的．对于不可压缩、二维模型，它是地层中迁移的耦合抛物型方程组的初边值问题．问题的数学模型由四类方程组成：

(1) 压力函数 $p(x,t)$ 的流动方程．

(2) 主要污染元素浓度函数 c 的对流扩散方程．

(3) 微量元素浓度函数组 $\{c_i\}$ 的对流扩散方程组．

(4) 温度函数 $T(x,t)$ 的热传导方程．

对于不可压缩、二维模型，它是地层中迁移型耦合抛物型方程组的初边值问题：

$$-\nabla\cdot U-q+R_s=0,\quad \text{(流动方程)}$$

$$-\nabla\cdot(cU)+\nabla\cdot(E_c\nabla c)-qc-q_c-R_s=\phi\frac{\partial c}{\partial t},$$

(brine 方程)

$$-\nabla\cdot(c_iU)+\nabla\cdot(E_c\nabla c_i)-qc-qc_i$$

$$+ q_{ci} + \sum_{j=1}^{N} K_{ij}\lambda_j K_j c_j - \lambda_i K_i \phi_i = \phi K_i \frac{\partial c_i}{\partial t},$$

$$i = 1,2,\cdots,N, \quad \text{(Radionuclide 方程)}$$

$$-\nabla \cdot (HU) + \nabla \cdot (E_H \nabla T) - q_c - qH - q_H$$

$$= [\phi\rho_p + (1-\phi)\rho_R c_{pk}] \frac{\partial T}{\partial t}, \quad \text{(热传导方程)}$$

此处 $U = -\frac{k(x)}{\mu(c)}\nabla p$ 是达西速度，p 是压力函数，c 是主要污染元素的浓度函数，c_i 是微量元素浓度函数($i = 1,2,\cdots,N$)，T 是温度函数. 这些都是核废料污染问题的待求函数.

对于上述方程组系统，可以用特征有限元方法或有限差分方法求其数值解，已有工业应用数值软件应用于生产实际，并产生重要的经济和社会效益.

8.6 海水入侵的预测和防治

海水入侵沿海地区，在自然水环境条件改变和社会环境条件影响下，造成的海水向沿海地区储水层的侵入. 我国的环渤海经济区情况特别严重. 山东省莱州湾沿海地区成为海水入侵的典型区域，海水入侵给该地区的经济发展和人民生活带来极大的危害. 因此深入研究海水入侵的成因、机制、规律，有的放矢地提出防治方案，采取各种切实可行的综合治理措施，尽快制止海水入侵的发展，缓解海水入侵带来的灾害，促进资源与环境的良性循环是一项十分重要的科研和工程任务. 海水入侵这种复杂的地下水运动，具有危害大、隐蔽性强、动态变化多、难以治理等特点. 利用计算机的高速计算能力，在渗流力学、水文地质科学基础上，考虑地理

环境、地质结构等复杂条件的影响，建立合理的数学模型，进而在计算机上定量描述海水入侵过程，对认识、掌握海水入侵的机制和规律，预测海水入侵的发展趋势，是切实可行、行之有效的方法．防治海水入侵的工程要花巨大的投资，要在长时间内发挥重大的作用，因此，依靠科学与工程计算的方法，采用计算机对工程的后效进行数值模拟，对工程的作用给出定量预测是很必要的，对提出工程的调控应用模式具有更现实的意义．

海水入侵定量研究是一个较新的学科领域．国外关于海水入侵研究始于本世纪中期，自 70 年代以来，随着海水入侵危害的日益加剧，对海水入侵的研究也逐步深入．我国海水入侵的研究，起步较晚，80 年代后期才开始研究秦皇岛、大连、山东莱州湾、江苏射阳等地的海水入侵问题，做了大量的调查研究，取得了阶段性成果．

对海水入侵问题的模型与算法通常有两类：一类认为海水和淡水是互不相溶的液体，两者之间存在一个严格的突变界面；另一类则认为两者是可混溶的液体，由于水动力弥散作用，它们之间有一个从淡水密度变到海水密度的过渡带．

当前急待研究和解决的问题：

(1) 适合我国沿海地区的动态数学模型研究，以适合不同的地质条件、区域形状和环境条件，这是预测系统的前提和基础．

(2) 适合广大海岸地区的三维大范围模拟算法研究，长时间预测算法研究对实用的预测系统是至关重要的．

(3) 三维自由潜水面的准确高速算法研究，实际上海水入侵是一类自由潜水面问题，它的准确计算直接决定了海水入侵区计算的可靠性．

(4) 运移问题的算法研究，特别对盐分浓度的运移问题，数值

模拟是非常困难的，对高维长时间问题就更为突出，成为另一个关键.

(5) 各种防治工程的预测和优化调控模式问题是预测防治的重要组成部分.

海水入侵治理工程主要有：

(1) 节水工程. 加强用水管理，全面节约用水，优化地下水抽水量，控制地下水位下降，控制海水入侵.

(2) 引江河调水工程. 充分利用大江大河的丰富水资源，建设引江河调水工程，增加可供水量，缓解供需矛盾.

(3) 人工降雨工程. 自然降水量只相当于降水云体中含水量的20% ~ 45%. 采用现代技术，利用人工催化云层的方法，使云中水滴增大变为雨落到地面，增加自然降水量，缓解水资源不足.

(4) 拦蓄补源工程. 阻止地表水流失，增加地面蓄水量和入渗水量，对海水入侵地层含水量进行补足，促使地下水回升，防止海水入侵.

(5) 防潮堤和地下板墙工程. 建立防潮堤工程，减少风暴潮对沿岸地区的直接危害，建立人工地下板墙，阻止海水向内陆入渗.

过去对防治工程的后效分析，以经验的定性分析为主. 目前已能数值模拟各项工程实施后咸淡水变化运移的真实过程，定量地对各种工程后效及各项工程综合后效进行预测，以达到优化工程设计的目的.

海水入侵的对流 — 弥散模型通常由流动方程和浓度方程组成，它们以及相应的初始条件和边界条件构成一个描述海水入侵含水层的完整数学模型. 这个模型不仅可以用来描述咸淡水界面的移动和过渡带的演化和发展，还可以用来预测抽水对界面移动

(海水入侵)的影响和各种工程后效检测、调控模式.

对于这类问题已成功应用现代数值方法(特征有限元、迎风差方法、算子分裂)求出它的数值解,并广泛的应用于生产实际和环境科学领域.

问题特点和难点:

(1) 问题复杂的一面是长时间、大范围、三维问题的数值模拟.这一难点给计算机上具体实现带来意想不到的困难,虽然计算机的存贮量大,计算能力很强,速度很快,但对如此大规模的三维问题,必须提出合理的三维大区域有效处理方法,才能满足实际的要求.

(2) 强对流特征.盐分方程为对流扩散方程,且对流项起主要作用,在数值模拟中对流项的处理好坏直接关系到整个过程的数值结果的可靠性.

(3) 非线性耦合问题.由于密度 $\rho(c)$,孔隙度 $\phi(p)$ 使得整个问题变得复杂,溶质浓度方程本来仅直接依赖于渗流速度,但若转化为压力,则方程的形式与通常的溶质输运问题有很大的差别,且问题有很大的难度.

(4) 潜水面问题.潜水面是随时间改变的,给问题求解带来更大困难,潜水面的变化依赖于问题的解本身.因此实际上变成更困难的非线性 Stefan 问题,潜水面的准确计算,将直接影响整个问题求解.

(5) 各类工程的数值模拟处理.各类工程将影响海水入侵的过程,这种影响如何正确地反映到数值模拟中,只有准确合理的处理这些工程的作用,才能给出很恰当的工程后效预测.

(6) 各类工程的优化模式.防治海水入侵工程的有效性,更为重要的一方面是如何优化应用各类工程,这反映在数学模型上,将

是一类十分复杂的反问题，而这又是十分重要和困难的.

8.7 半导体器件瞬态问题的数值模拟

半导体技术的迅速发展，对其传统近似的分析方法已不再适用，需要研究通常称为扩散模型的非线性偏微分方程组的初边值问题. 对于几何形状复杂的半导体器件的高维问题，必须应用数值模拟方法求解.

半导体器件瞬时状态的数学模型由三个非线性偏微分方程组所决定：一是椭圆型的电子位势方程，另两个是抛物型的电子和空穴浓度方程. 电子位势通过其电场强度在浓度方程中出现，由电子和孔穴守恒可推出关于电场位势 φ 和电子、空穴浓度 e 和 p 的方程. 著名学者 Gummel 于 1964 年提出用序列迭代法计算这类问题，开创了半导体器件数值模拟这一新领域. 注意到在浓度方程中出现的不是电子位势，而是它的梯度(电场强度). 分别用混合元来逼近位势，用特征有限元来逼近浓度方程. 和传统的计算方法(差分、有限元) 相比，具有格式简单、截断误差小、能对时间采用大步长计算等特点，是一类很实用的半导体工程计算方法.

近年来，由于科学技术的迅速发展，在半导体器件数值模拟中，必须研究三维热传导型半导体器件的瞬态问题，要研究三维空间复杂的几何外形和结构，同时要考虑热传导对半导体瞬态问题的影响，否则模拟将会失真.

8.8 能源数值模拟的发展前景

1953 年美国 G. H. Bruce 等人发表了《孔隙介质中不稳定气体渗流的计算》一文，为用电子计算机计算油藏渗流问题开辟了一条新路. 近 40 年来，由于大型快速电子计算机的迅速发展，现代大规模工程和科学计算方法的逐步完善，大大促进了油藏数值模拟方法的发展和广泛应用. 目前，黑油、混相和热力采油模型及其软件已投入工业性生产，化学驱油模型和软件也正日臻完善. 而且，这一方法在近十年已成功应用到油、气藏勘探（油、气资源评估）、核废料污染、海水入侵预测和防治、半导体器件的数值模拟等众多领域，并取得了重要的成果. 可以预期，能源数值模拟在 90 年代以后及至下个世纪将会出现重大的进展和突破，在国民经济各部门产生重要的经济效益，并将进一步推动计算数学和工业与应用数学学科的发展，在国家的现代化建设事业中发挥其巨大的作用.

9 线性代数方程组的数值求解

9.1 多元一次方程组求解容易吗

我们在初中就学过二元一次方程组、三元一次方程组和多元一次方程组.这种方程组,在大学里称为线性代数方程组.有几个未知量的,就称几阶线性代数方程组,有 n 个未知量的,就称 n 阶线性代数方程组.

在现实生活中,常常碰到需要求解线性代数方程组的问题,例如,投入产出,大地测量,土木工程中的应力、应变计算,机械设计中的热应力分析,电路设计,油藏分析,等等.可以说在计算机进行科学计算的任务中,有很大的比例是求解这样的线性代数方程组.

在中学里学二元一次方程组,三元一次方程组求解时,往往认为求解多元一次方程组,似乎没有什么问题了.其实不然,在实际问题计算中,有很多的

问题要研究.自从本世纪40年代末期计算机产生以来,线性代数方程组的数值求解一直是计算数学研究的重要问题,吸引了许多国际上著名的专家进行这项研究工作.造成数值求解困难的原因有两个,第一个原因是:方程组的未知量个数很多,上千个,上万个,甚至更多.例如在我国大地测量问题中,碰到36万个未知量的线性代数方程组求解,大庆油田提出要解上百万个未知量的线性代数方程组.随着生产的发展,人们碰到的线性代数方程组的未知量的个数越来越多,因为阶数很高,必然使得计算量很大,就会产生计算速度问题,人们总希望求解的速度很快,有些问题必须很快求出解来,例如大庆油田的上百万个未知量的线性代数方程组,是求解一组非线性的偏微分方程中产生的,他们希望整个问题能在8小时之内获得解.再如CT切片的结果,病人希望马上能知道结果.预报明天的天气,如果后天才算出来就没意义了.

因为阶数很大,初始数据如何存放到计算机里,也是个问题.譬如二元一次方程组

$$\begin{cases} 3.16x + 5.83y = 7.91 \\ 4.53x - 6.48y = 3.06 \end{cases}$$

中,3.16,5.83,4.53,-6.48,7.91,3.06,总共有6个已知量要存放到计算机里.对于n元线性代数方程组,就得有n^2+n个数要存放到计算机里,当n很大时,计算机能不能放得下,这样就产生数据存放问题.第二个原因是:参与数值计算的数都是有限位的.每种计算机进行运算时的数都是如下形式的浮点数,

$$\pm d_0 . d_1 d_2 \cdots d_t \times \beta^c$$

β是数制的基,$\beta=10$即为10进制,$\beta=2$即为2进制,c是β的指数,它有上界U和下界L,即$L \leq c \leq U$.$d_0 . d_1 d_2 \cdots d_t$是β进制

的小数，d_0 是整数部分是小于 β 的正整数，$d_1d_2\cdots d_t$ 是小数部分，t 是位数．每种计算机运算的数是由 (β,t,L,U) 四个数决定的．下面列出几种计算机的 (β,t,L,U)，见表 9－1．

表 9－1　几种计算机的 $(\beta、t、L、U)$

	β	t	L	U
EEE 标准	2	23	－126	127
IBM3090	16	5	－65	62
CDC6000	2	47	－975	1071
CRAY－MP	2	47	－16385	8190

这样在 CRAY－MP 的一个单元中能表示的最小的正数为 2^{-16385}，最大的正数为 $(2-2^{-47})2^{8190}$，小于 2^{-16385} 的正数都表示成 0，大于 $(2-2^{-47})\cdot 2^{8190}$ 的正数，计算机的一个单元无法表示，如果硬是要在一个单元表示计算机就会溢出停机．

因为这样在计算机进行四则运算时，就会有舍入误差产生．大量的四则运算下，误差究竟对计算结果造成如何样的损害，就是一个非常复杂的问题．有时候误差很大，计算结果可以没有一位有效数字，但有时候误差很小，计算结果很好，这就是误差问题．

所以，对于变量个数非常多时，多元一次方程组求解并不是一件容易的事情．只有很好地解决计算速度、误差控制、数据存储等问题，才能有效地求解．

9.2　速度问题

n 阶的线性代数方程组写成矩阵形式是

$$Ax=b,$$

这里 $x=(x_1,x_2,\cdots,x_n)^T$ 是 n 个未知量构成的列向量，A 是由

n^2 个已知系数构成的，称为系数矩阵，$b=(b_1,b_2,\cdots,b_n)^T$ 是已知的向量，它在方程的右端，称为右端向量.

学过行列式的读者都知道，线性代数方程组有解的充分必要条件是 A 的行列式 $\det(A)\neq 0$，并且当 $\det(A)\neq 0$ 时解是唯一的，可表示成

$$x_i=\frac{\det(A_i)}{\det(A)}.$$

这就是著名的克莱姆(Cramer)法则，其中 $\det(A_i)$ 表示矩阵 A_i 的行列式，而 A_i 是将 A 中的第 i 列用 b 代替后的矩阵.

从数学上来说 x_i 已经把方程的解用已知数表示出来了，但从实际计算来说，没有人这样做，因为计算一个行列式的计算量很大.如果按最原始的办法计算，一个行列式有 $n!$ 项，每一项有 n 个数相乘，要做 $(n-1)$ 次乘法，这样计算一个行列式要做 $n!\ (n-1)$ 次乘法，按照克莱姆法则要计算 $(n+1)$ 个行列式，于是要做乘法 $(n^2-1)n!$ 次.从表 9－2 可以看出，当 $n=10,100,1000,10000$ 时，$(n^2-1)n!$ 是很大的.

表 9－2　未知量为 n 时乘法运算的次数

n	$(n^2-1)n!$
10	3592561200
100	9.3316910^{161}
1000	$4.023868576898337\times 10^{35667}$
10000	$2.8462596524454458\times 10^{35667}$

按目前最快的计算机每秒 1 亿次乘法运算的计算机来计算，按克莱姆法则计算求解 10 个未知量的方程组要用 3.59251 秒，而求解 100 个未知量的方程组要用 9.3316910^{153} 秒 -2.95906×10^{146} 年，这已是不现实的了，更何况 1000 个未知量，10000 个未知量了.

当然不能用这种方法来求解线性代数方程组.早在公元 1 世纪时,我国的《九章算术》一书的第八章方程中,就提出解线性代数方程组的消去法,后来大数学家高斯(1777－1855)把它发扬并系统化,现在国际上往往称为高斯消去法.

该方法的基本思想是首先由一个方程解出第一个变量,然后,将其代入其余的方程中,消去各方程中的这个变量.对其余的方程用同样方法消去第二个变量,如此反复进行下去,直到只剩下一个变量,即可求出该变量的值,然后逐步回代可将所有变量的值都计算出来.对于 n 个未知量的方程组,高斯消去法所需的乘法运算总数为 $\frac{1}{3}n(n^2-1)$,除法运算总数为 $\frac{1}{2}n(n+1)$,加减法运算总数为

$$\sum_{k=2}^{n}k(k-1)+\frac{n(n-1)}{2}=\frac{n^3}{3}+\frac{n^2}{2}-\frac{5n}{6},$$

乘除加减法运算合在一起为 $\frac{2}{3}n^3+n^2-\frac{2n}{3}$ 次运算.

表 9－3 表示未知量个数为 100,1000,10000,10^5,10^6 时运算量的数字.如果使用每秒 1 亿次运算的计算机进行计算,100 个未知量的方程组,1 秒不到就算出来了,而 1000 个未知量的要算 6 秒多,10000 个未知量的要算 1.85 个小时,10 万个未知量的要算 77.1617天,1 百万个未知量的要算211.399年.这项估算告诉我们

表 9－3　未知量个数为 n 时运算量的数字

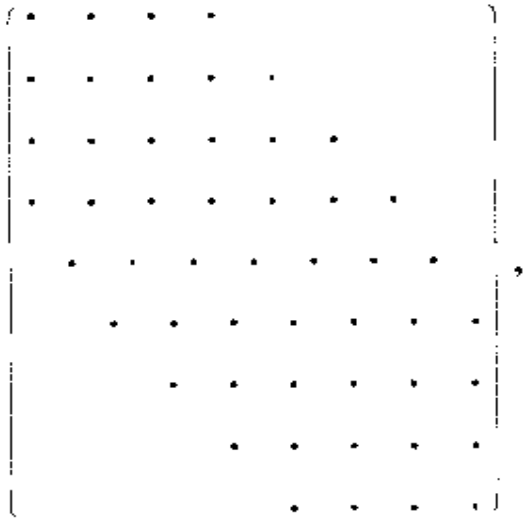
n	$\frac{2}{3}n^3+n^2-\frac{2}{3}n$
100	676600
1000	667666000
10000	666766660000
10^5	666676666600000
10^6	66666766666000000

上万阶的线性代数方程组要用高斯消去法做从速度上来讲也是有问题的.

人们发现要计算的线性代数方程组的系数矩阵中常常有许多零元素,很多实际问题有这一特点,利用这些零元素能否少作一点运算,例如4阶方程

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 = b_1, \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 = b_2, \\ a_{32}x_2 + a_{33}x_3 + a_{34}x_4 = b_3, \\ a_{43}x_3 + a_{44}x_4 = b_4, \end{cases}$$

即方程中的 $a_{31} = a_{41} - a_{42} = 0$,此时消去 x_1 时,第3,第4个方程已经没有 x_1,可以不要运算,而在消去 x_2 时第4个方程也不必计算了.对于一般的方程组来说,如果系数矩阵形状是带状的,即



带的外边的元素都为零,这样的方程组不论消去过程和回代过程都可以省去不少运算量,但如果矩阵内部的元素为零.如上述4阶方程组中 $a_{22} = a_{23} = 0$,这样在消去第二个方程中的 x_1 时,x_2 和 x_3 的系数从零变成了非零.

另外一种提高求解速度的途径就是采用迭代法来求解，对于未知量个数较多时，常常采用这种方法．迭代法先把方程组

$$Ax = b \tag{9.1}$$

等价化成

$$x = Bx + g, \tag{9.2.1}$$

这里 B 是 $n \times n$ 矩阵，g 是 n 维向量．所谓等价即二者的解完全相同．迭代法的做法是取一个初始向量 x_0，然后由

$$x_k = Bx_{k-1} + g,$$

可计算得向量序列 $\{x_k\}$，x_k 称为近似解，当然不必要把所有的 x_k 都计算出来，当 B 满足一定的条件时，x_k 收敛到(9.1)式的解，当 k 充分大时，x_k 就是很好的近似解，它与真解 x 的差相当于 ρ^k、这就是收敛速度，其中 ρ 是矩阵 B 的谱半径(即 B 的特征值的绝对值的最大值)．因此如果 B 的谱半径很小，收敛速度就很快，但往往找不到等价方程组(9.2.1)式使 B 的谱半径很小．在实际问题中，经常碰到的 B 的谱半径都在 0.9 以上．对于泊松方程

$$\begin{cases} \dfrac{\partial^2 u}{\partial x^2} + \dfrac{\partial^2 u}{\partial y^2} = f(x,y), & x,y \in \Omega, \\ u \mid \Gamma = 0, \end{cases}$$

当区域 Ω 为 $[0,1] \times [0,1]$ 时，用 5 点差分格式所化出来的线性代数方程组

$$Ax = b$$

称为线性代数方程组的模型问题．记 D 是模型问题方程组的系数矩阵 A 的对角阵，$A = D + (A - D)$，方程组可等价成

$$x = D^{-1}(D - A)x + D^{-1}b,$$

即按(9.2.1)的记法，此时 $B = D^{-1}(D - A)$．当差分的网格是由

每边 10 等分组成时，共有 $9\times 9=81$ 个未知量，此时 $\rho(B)=0.951057$，而当差分的网格是由每边 100 等分组成时，共有 $99\times 99=9801$ 个未知量，此时 $\rho(B)=0.999507$，这时要使 x_k 逼近到 10^{-6}数量级时，要求 $k\geq 28017$ 即要迭代 28000 次．按照迭代法计算，计算量也是很大的．实际问题要求精度很高，必须网格很细，所得方程组的未知量就越多，求解越不容易．为此人们要不断研究提出各种新的迭代法．例如，超松弛迭代法，块超松弛迭代法，对称超松弛迭代法，交替方向迭代法，AOR 迭代法等等，为了提高求解的计算速度，这方向的研究是没有止境的．

另一方向是从泛函极小观点来导出一些迭代法．对于方程组(9.1)式中的矩阵 A，如果是对称正定矩阵，那么(9.1)式的求解等价于求泛函

$$F(x)=(Ax,x)-2(b,x)$$

达到最小的向量．在 50 年代提出了共轭斜量法，这个方法的思想是这样的，取一个初始向量 x_0，然后有残量 $r_0=Ax_0-b$，若 $r_0\neq 0$，求 x_1，在 α 变动下得到使 $F(x_0+\alpha r_0)$达到极小的 α_0，而 $x_1=x_0+\alpha_0 r_0$，x_2 是 $x_0+\alpha r_0+\beta Ar_0$，变动 α,β 得到使 $F(x_0+\alpha r_0+\beta Ar_0)$达到极小的 α_1,β_1，$x_2=x_0+\alpha_1 r_0+\beta_1 Ar_0$；依此类推，可以导出$\{x_k\}$．如果方程组的未知量个数为 n，这个方法理论上讲 x_n 就是真解，实际计算时 n 很大，x_k 在 k 很小时就达到很好的近似．我们有估计

$$\|x_k-x^*\|\leq\sqrt{\frac{\lambda_n}{\lambda_1}}\frac{1}{T_k\left(\frac{\lambda_n+\lambda_1}{\lambda_n-\lambda_1}\right)}\|x_0-x^*\|,$$

这里 $T_k(x)$是切比雪夫多项式

$$T_k(x)=\frac{(x+\sqrt{x^2-1})^k+(x-\sqrt{x^2-1})^k}{2},$$

λ_1,λ_n 分别是 A 的最小特征值和最大特征值，$\frac{\lambda_n}{\lambda_1}=p$ 称为 A 的条件数．当 p 很大时，共轭斜量法收敛也是很慢的．这时候称方程组为病态的．像模型问题，当网格由 10 等分构成时，条件数 $p\approx 4\times 10^3$，已经很大，等分数再增加时 p 会更大．一般实际问题常常是阶数越高，方程组越病态，计算速度越慢．为了提高速度，人们又想出预处理的方法．将

$$Ax=b$$

等价化成 $PAx=Pb$，$PAP^TP^{-T}x=Pb$，令 $y=P^{-T}x$，$g=Pb$ 得

$$PAP^Ty=g\ , \tag{9.2.2}$$

P^T 表示矩阵 P 的转置，P^{-T} 表示矩阵 P^T 的逆．如果方程(9.2.2)式的条件数比(9.1)式的好，且从 $P^{-T}x=y$ 容易求得 x，而 P 又容易求逆，那就是预处理的要求．当然希望 PAP^T 的条件数越小越好．因此预处理的方案也是多种多样的，而且也可说是没有止境的．预处理不但对共轭斜量法有用，对其他任何方法都是有用的．上面讲的是对 A 是对称正定的情形，对于 A 对称而不正定，甚至 A 不对称的情形，究竟如何呢？对于 A 对称而不正定的情形，人们研究出 SYMMLQ 方法，对于不对称的情形，人们研究出正交极小化方法和改进的正交极小化方法，如 ORTHOMIN 方法，GMRES 方法，GMBACK，FGMRES，GMRESR 等方法．还有双正交化方法如：BCG，BICGSTAB，QMR，QMRCGSTAB，等等．这方面的研究正方兴未艾．

9.3 误差问题

在数值计算过程中，每一次四则运算都可能产生舍入误差，而每一步产生的误差，又会传播到以后各步计算，影响最后结果，即使简单的二元一次方程组数值求解，都会被误差所困惑．譬如方程组

$$\begin{cases} 1.00000\times 10^{-5}x_1 + 1.00000x_2 = 5.00005\times 10^{-1}, \\ 1.00000x_1 - 6.00000\times 10^{-1}x_2 = 2.00000\times 10^{-1}, \end{cases}$$

我们在(10,5,-10,10)体系中进行计算，即参与四则运算的数和计算成的结果都是形如

$$d_0.d_1d_2d_3d_4d_5\times 10^e \qquad -10\leqslant e\leqslant 10$$

样的 10 进制浮点数．按高斯消去法求上述方程的数值解．第一步用 1.00000×10^{-5} 除以第一方程的两边得

$$x_1 + 1.00000\times 10^5 x_2 = 5.00005\times 10^4,$$

第二方程减去(9.2.3)得 (9.2.3)

$$\begin{aligned} &(-6.00000\times 10^{-1} - 1.00000\times 10^5)x_2 \\ &= 2.00000\times 10^{-1} - 5.00005\times 10^4, \end{aligned}$$

在计算机里两数相加或相减时，先要对阶，即两个数的指数照大的看齐，然后在双倍位的寄存器中相加，再将和数四舍五入成所给的(β,t,L,U)体系中的数．这里的$(\beta,t,L,U)=(10,5,-10,10)$．于是

$$\begin{aligned} &-6.00000\times 10^{-1} - 1.00000\times 10^5 \\ &= -(0.000006 + 1.00000)\times 10^5 \\ &= -1.000006\times 10^5, \end{aligned}$$

四舍五入成 $\approx -1.00001\times 10^5$，

而

$$\begin{aligned}&2.00000\times10^{-1}-5.00005\times10^{4}\\&=-(5.00005-0.00002)\times10^{4}\\&=-5.00003\times10^{4},\end{aligned}$$

这样得到消去后的方程为

$$-1.00001\times10^{5}x_2=-5.00003\times10^{4}.$$

在计算机里做除法也是先在双倍位的寄存器中得到相除结果再四舍五入，求得

$$x_2=4.99998\times10^{-1},$$

代入(9.2.3)中求得

$$\begin{aligned}x_1&=5.00005\times10^{4}-4.99998\times10^{4}\\&=0.00007\times10^{4},\\&=7.00000\times10^{-1},\end{aligned}$$

但原来的方程的解是 $x_1=0.5, x_2=0.5$，计算出来的解为 $x_1=0.7$，误差达到 40%，两个未知量的方程组会产生如此大的误差，几千个，几万个未知量的方程组就不堪设想了，好在刚才的问题，是由于我们用相对很小的数 10^{-5} 除方程的两边产生的，因此可以用选主元的办法来改善这个状况．所谓选主元即是在 A 的各元素中找一个绝对值最大的，如为 $a_{i_0j_0}$，那么将第 i_0 个方程移到第一个方程的位置，将第 j_0 个未知量 x_{j_0} 记为 y_1，再令

$$x_1=y_2,\ x_2=y_3,\ \cdots,x_{j_0-1}=y_{j_0},$$

$$x_{j_0+1}=y_{j_0+1},\cdots,x_n=y_n,$$

这样对 $y=(y_1,y_2,\cdots,y_n)^T$ 有新的系数矩阵 $\tilde{A}=(\tilde{a}_{ij})$，同样右端向量

$$b_{i_0}=\tilde{b}_1, b_1=\tilde{b}_2, \cdots, b_{i_0-1}=\tilde{b}_{i_0}, b_{i_0+1}=\tilde{b}_{i_0+1}, \cdots, b_n=\tilde{b}_n,$$

得方程

$$\tilde{A}y=\tilde{b}=(\tilde{b}_1,\tilde{b}_2,\cdots,\tilde{b}_n)^T,$$

对 $\tilde{A}$ 消去 y_1，得

$$\tilde{a}_{ij}^{(1)}=\tilde{a}_{ij}-\frac{\tilde{a}_{i1}\tilde{a}_{1j}}{\tilde{a}_{11}},\quad \tilde{b}_i^{(1)}=\tilde{b}_i-\frac{\tilde{a}_{i1}\tilde{b}_1}{\tilde{a}_{11}},$$

$$i=2,3,\cdots,n;\ j=2,3,\cdots,n$$

对 $n-1$ 的矩阵 $(\tilde{a}_{ij}^{(1)})$ 再找一个绝对值最大值的元素，再仿照上面的办法得到新的方程组新的未知量，依此类推，这样作每次选主元的工作量是比较大的，因为第一次要在 n^2 个数中选一个绝对值最大的数，计算机要作很多次比较运算．第二次要在 $(n-1)^2$ 个 $a_{ij}^{(1)}$ 中选，第三次要在 $(n-2)^2$ 个 $a_{ij}^{(2)}$ 中选，…，这是很大的负担，而且每次未知量变换次序都要记录下来，才能知道最后的未知量的次序．为此有人提出了部分选主元的办法，在第一步考虑 a_{i1}，$i=1,2,\cdots,n$ 中选一个绝对值最大的 a_{i_01}，将第 i_0 个方程移成为第一个方程，未知量的次序不变动了，对于 $a_{ij}^{(1)}$，在 $a_{i2}^{(1)}$ 中挑选，依此类推．这样作还是有效果的．对于前面的方程，a_{21} 的绝对值比 a_{11} 的大，因此方程成为

$$\begin{cases}1.00000x_1-6.00000\times10^{-1}x_2=2.00000\times10^{-1},\\ 1.00000\times10^{-5}x_1+1.00000x_2=5.00005\times10^{-1},\end{cases}$$

消去后的方程为

$$(1.00000+6.00000\times10^{-6})x_2$$

$$=5.00005\times10^{-1}-2.00000\times10^{-6},$$

经过对阶四舍五入成为

$$1.00001\,x_2 = 5.00003\times 10^{-1},$$

$$x_2 = 4.99998\times 10^{-1};$$

$$x_1 = 2.00000\times 10^{-1} + 6.00000\times 4.99998\times 10^{-2},$$

$$x_1 = 2.00000\times 10^{-1} + 2.99999\times 10^{-1}$$

$$= 4.99999\times 10^{-1},$$

与真解比较误差十分小.

因此作消去法通常都要采取部分选主元的手段,但是即使部分选主元的误差还是很复杂的,有时也会很大.人们经过不断努力,终于证明了:对于方程组

$$Ax = b,$$

如果用部分选主元的消去法计算,计算机是采用 t 位二进制小数,则计算机求得数值解 $\tilde{x}$ 满足

$$(A+\delta A)\tilde{x} = b, \tag{9.2.4}$$

其中 δA 也是一个 $n\times n$ 矩阵,它的上界

$$\|\delta A\| \le 1.01(n^3+3n^2)\rho\|A\|2^{-t}, \tag{9.2.5}$$

这里 $\rho = \max\limits_{i,j,k}|a_{ij}^{(k)}|/\|A\|$,称为增长因子,这里所取矩阵范数 $\|B\|$ 指 B 的 l_∞ 范数,即 B 各行元素绝对值之和中的最大值.

利用(9.2.4)式可以估计 $\tilde{x}$ 的误差,实际上

$$(A+\delta A)x = b + \delta Ax,$$

$$(A+\delta A)\tilde{x} = b,$$

相减得

$$(A+\delta A)(x-\tilde{x}) = \delta Ax,$$

或

$$(x-\tilde{x})=(A+\delta A)^{-1}\delta A x$$
$$=(I+A^{-1}\delta A)^{-1}A^{-1}\delta A x.$$

当 $\|\delta A\|\ \|A\|<1$ 时

$$(I+A^{-1}\delta A)^{-1}\leq\frac{1}{1-\|A^{-1}\|\ \|\delta A\|},$$

记 $K(A)=\|A^{-1}\|\ \|A\|$ 为方程组(9.1)式的条件数，就得

$$\frac{\|x-\tilde{x}\|}{\|x\|}\leq\frac{K(A)\frac{\|\delta A\|}{\|A\|}}{1-K(A)\frac{\|\delta A\|}{\|A\|}}.$$

由(9.2.5)式知

$$\frac{\|\delta A\|}{\|A\|}\leq 1.01\rho(n^3+3n^2)2^{-t}.$$

这个估计式告诉我们，对坏的条件数，即大的 $K(A)$、大的 n 和大的增长因子 ρ，数值解 $\tilde{x}$ 的误差可能很大。增长因子 ρ 究竟可能达到多大，至少可以举出一个 n 阶线性代数方程组，它的增长因子 ρ 达到 $2^{n-1}/n$，例如：

$$A=\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 1\\ -1 & 1 & 0 & \cdots & 0 & 1\\ -1 & -1 & 1 & \cdots & 0 & 1\\ & & & & \ddots & \vdots\\ -1 & -1 & -1 & \cdots & & 1 \end{pmatrix}.$$

将这个 A 按部分选主元消去法，最后可得

$$\max_{i,j,k}|a_{ij}^{(k)}|=2^{n-1}.$$

但在大量实际问题计算中，ρ 一般不会大于 8，当然估计式(9.2.5)式也是不够理想的，还有待于更好的估计式。

考虑到整体估计 $\|x-\tilde{x}\|$ 往往将小的 x_i 忽略掉，为此要研

究数值解 x 的每一个分量的相对误差，即研究

$$\frac{|x_i - \tilde{x}_i|}{|x_i|}, \qquad i = 1, 2, \cdots, n$$

的上界.

对于迭代法也有误差问题，计算每一步迭代误差有多大？特别从 x_k 计算 x_{k+1} 时，因为计算中的误差，或 x_k 的误差，影响 x_{k+1} 是否很大？如果影响比较大就为不稳定，此时会影响求解的速度. 为了克服这种不稳定现象，对迭代法要作改进. 譬如：在 9.2 中提到的，因为 BCG 方法不够稳定，所以改进成 BICGSTAB，同样 QMRCG 改进成 QMRCGSTAB. 这些都太专门了，不便在此作进一步的介绍.

9.4 存储问题

对于前面提到的方程组(9.1)式如果把矩阵 A 和向量 b 的元素全部存放在计算机内存中得有 $n^2 + n$ 个单元，这是很重的负担. 利用很多矩阵是稀疏的即有很多元素为零的特点，研究将那些零元素不存放在计算机的内存中. 这一设想对于作迭代法求近似解时是可以采用的，只要将那些元素不为零的列足标告诉计算机. 实际上在迭代法中矩阵是用在与向量的乘法上 $y = Bx$，y 的第 i 个分量 $y[i] = \sum_{j=1}^{n} B[i,j]x[j]$，如果计算机知道 B 中第 i 行的 $j_1, j_2, j_3, \cdots, j_k$ 个元素 $B[i, j_l] \neq 0$，$l = 1, 2, \cdots, k$，其余元素都为零，那么

$$y[i] = \sum_{l=1}^{k} B[i, j_l]x[j_l].$$

因此只要在计算机中存放非零元素 $B[i, j_l]$，如果 B 中有 m 个元

素非零，那样要实现 Bx，除了这 cn 个元素要存放外还得存放 j_1，$j_2, j_3, \cdots, j_k$，要增加 cn 个单元。如果这些列足标排成一个一维数组，从第一行的列足标开始排，接着为第二行的，依次类推，于是还需要告诉计算机第 i 行的列足标从哪里开始，这样还得要 n 个单元，尽管如此，存放的量比起 n^2 来可能还是很小的，因而对于非常高的线性代数方程组常常采用迭代法求解。

对于消去法来说，光有存放矩阵的非零元素是远远不够的，因为原来矩阵中的零元素经过消去过程可以变成非零的，这种现象英语中称为 fill - in，如：4 阶方程组矩阵 $a_{22}=a_{23}=0$，但 a_{12}, a_{13}, a_{21} 都不为零，就有 $a_{22}^{(1)}=a_{22}-\frac{a_{21}a_{12}}{a_{11}}$ 可以不为零，$a_{23}^{(1)}=a_{23}-\frac{a_{21}a_{13}}{a_{11}}$ 也可以不为零，如果原来没有存放 a_{22}, a_{23}，新产生的 $a_{22}^{(1)}$、$a_{23}^{(1)}$ 就要重新安排存放的地方。整个消去过程中会产生多少这种 fill - in 也是很难预料的，而且可以发现，当方程组中的方程次序和未知量的次序变化时，这个 fill - in 数也会变化，即对矩阵 A 左乘置换阵 P 和右乘置换阵 Q，可以改变 fill - in 数，于是可以考虑找 P, Q 使得 PAQ 在消去过程中有较小的 fill - in 数。人们已经研究出一些寻找这种 P, Q 的算法。

在 9.2 节中提到矩阵 A 为带型时，可以节省计算量，当然也可以节省存贮量。但更重要的是这带型的宽度（称为带宽）也是随着未知量的次序变化而变化的，即对于置换矩阵 P，P^TAP 的宽度随 P 而变化，因此就必然提出如何找一个 P，使得 P^TAP 的宽度最小的问题，对于这个问题也已经研究出一些算法，已找到较好的 P，尽管它不一定是最佳的，但却是比较实用的。

10 最优化计算方法

10.1 从黄金分割法到瞎子爬山

最优化问题也许是我国最广为人知的数学问题.著名数学家华罗庚教授在全国各地推广优选法,此项工作得到了毛泽东和周恩来的重视.在全国各地广泛地开展起来,华罗庚教授在其著作《优选学》中提到:"优选法首先应用在化工、电子工业,然后逐步推广到各行各业.连炸油条省油,做豆腐省豆都用上了.在医学上,外科医生用了优选法能够在病人肠子里找到出血点."可见优选法在当时的普及范围是非常之广.

当时优选法的试点与推广工作主要是介绍和应用黄金分割法,使人们尽可能少做试验而尽快找到生产最优化方案.黄金分割法是求一维的单峰函数极大值的计算方法.设函数 $f(x)$ 在区间 $[0,1]$ 是单

峰函数，黄金分割法是在 $w=\frac{\sqrt{5}-1}{2}$ 处和对称点 $1-w$ 处计算函数值，然后割去不含最大值的区间，反复这样进行下去，直到余下的区间小到可接受的误差范围即终止计算．w 被称为黄金分割比例式或黄金分割因子，它被认为是最美的比例，常常应用于建筑学和造型艺术．“黄金分割”这一术语是由著名艺术家达·芬奇引入的．由于 $w=\frac{\sqrt{5}-1}{2}\approx 0.618$，所以黄金分割法也常称为 0.618 法．

考虑爬山问题．设经度与纬度是两个变量，则对应于给定经度与纬度点的海拔高度是关于经度与纬度的函数．爬到山顶也就是求该函数的极大值．正因为爬山问题是一个二维优化问题，华罗庚教授把求解优化问题的坐标轮流搜索法称为“瞎子爬山法”．该方法是依次沿着各个坐标轴的方向搜索，此方法在“爬山”问题表现为先沿着经线方向往前或往后迈一步（看那个方向能向上走），再沿着纬线方向往前或往后迈一步．一直这样反复走．如果沿经线和纬线迈步都不能上升，则缩小步长再继续走．如果步长已小于给定的精度要求，则认为已达到山顶．很显然，任何一个瞎子爬山不会走一步转九十度弯再走．可见“瞎子爬山法”比瞎子爬山要笨得多．

10.2 最优化问题及其计算方法

最优化就是在众多的决策里挑选“最优”的，如寻求最大利润、最短路径、最佳弹道、最少下料、最少试验次数等等．它在化学反应、工程设计、交通运输、城市规划、油田开发、自动控制、经济市场、空间技术、生命科学、大气海洋等方面都有着广泛的应用，是运筹学、计算数学、应用数学、力学以及众多的其他应用领域的交叉．

“运筹帷幄之中，决胜千里之外”，从这句话可见我国古代就已很重视决策的最优性．但最优化作为一门学科被系统深入的研究只有几十年的历史．事实上，运筹学的英文名称 operational research 的直译是“军事行动研究”，运筹学的产生与迅速发展的主要原因是第二次世界大战中出现了大量的兵力布置、兵种配置、武器装备等的优化问题，以及战后计算机的飞速发展．

最优化问题在数学上是满足于一些约束条件（如等式、不等式等）求函数的极大值或极小值．由于判断一个点是不是极值点要知道该点附近所有点的函数值，而这样的点往往有无穷多个，故要直接验证一个点的最优性是几乎不可能的．所以研究最优点应满足的条件（必要条件）以及能保证该点为最优点的条件（充分条件）是非常重要的．对于无约束优化问题，不难看出，可微函数的极值点必定是稳定点（即导数为零的点）．如果有约束，则极值点是拉格朗日函数的鞍点．对于特殊类型的函数，最优性条件的形式不同．而且对于高阶最优性条件，非光滑问题的最优性条件等仍是从事最优化理论研究的专家十分关注的问题．

研究最优化计算方法主要是寻求构造求解最优化问题的数值计算方法；对算法进行理论分析，如研究收敛性、收敛速度、算法复杂性等；进行算法的数值试验，在计算机上对算法比较与分析，揭露算法的表现与特点；利用所构造的算法求解实际应用部门的优化问题等．算法的构造、方法的提出、技巧的设计等都是最优化计算方法研究的重要方向．最核心的是如何构造好的计算方法．

当前最优化问题的特点是规模大，变量个数几万个甚至几百万个．例如，在全球天气预报的计算中，如果网格是以 20km 乘 20km 划分的，则一共有几百万个网格点，再将大气层分为二十

层，则节点个数就达到几千万个甚至上亿个．在陆地上方节点的初始数据（气温、风速、压力）都可通过直接测量（如放高空探测气球）得到，在海洋上方的初始数据无法用直接测量得到，而是利用求最小二乘来计算，这就是一个求几千个变量的最小二乘问题．另一个大规模优化问题是分子动力学计算，每个分子对应3个变量，由于分子的个数往往是非常多的，故从分子动力学导出优化问题常常是有几千几万个变量．大规模问题需要计算量大，故算法的效率是至关重要的．这时算法效率的提高可带来巨大的经济效益．规模大带来的另一个困难是存储量需要很大，这一点在第9章已经提到．例如用牛顿法求解一个 n 个变量的无约束优化问题需要存储 $n\times n$ 海色阵，这就需要 $\frac{1}{2}n(n+1)$ 个存储单元（考虑对称性）．如果 n 是4千万时，则 $\frac{1}{2}n(n+1)>\frac{1}{2}n^2=8$ 百万亿！这已超过了目前超级计算机的存储量．可见，对于大规模优化问题，一定要研究特殊形式的方法，使得算法尽可能少用存储但仍保证收敛较快．

最优化问题倍受重视，是由于它的应用能解决大量重要的工程实际问题以及产生巨大的经济效益．例如，80年代美国著名数学家P. Lax教授在向美国政府上书呼吁科学计算对美国国民经济的重要性时列举了三个重要的计算问题．其中之一就是在全美国电力线路设计中成功地应用了非线性优化的逐步二次规划方法，得到了最优化方案，取得了很好的效果．所以一定要重视对最优化计算方法的研究，使其在我国国民经济建设中发挥作用．

10.3 Karmarkar 算法

优化界乃至整个科学计算领域在 80 年代最轰动的是 Karmarkar 算法的提出. Karmarkar 算法是求解线性规划的一个利用投影的内点算法,它在 1984 年被提出后立即受到了广泛的注意,《纽约时报》、《华盛顿邮报》等美国重要报纸都在醒目的位置报道了这一成果.

线性规划问题是在线性等式或线性不等式的约束下求一个线性函数的极小值,它相当于求一个凸多面体的最高的顶点.传统的求解方法是单纯形法,沿着棱边从一个顶点爬到更高的一个相邻顶点,直到达到最高点.单纯形法在大量实际应用中是非常有效的,但可以构造例子,使单纯形法计算所有的顶点后才能找到最优解.由于顶点的个数是随着问题的规模增长而指数增长的,因此单纯形方法的复杂性是指数的.也就是说,利用单纯形方法求解线性规划问题,在最坏的情形下所需时间将随着问题规模增长而以指数增长的.是否存在和能否找到计算时间仅以多项式速度增长的方法,自然就引起了不少优化专家的注意.1978 年前苏联数学家 Khachian 提出了一个多项式时间的椭球方法,正面回答了多项式时间算法的存在性,当时引起了优化专家的高度重视.但是,由于该算法在实际计算中表现太差,以至于昙花一现后就很快被人们忘记了.

1984 年 Karmarkar 提出的算法是多项式时间的.该方法不仅具有很好的理论性质,而且在实际计算中也表现非常好.由于线性代数方程组求解也可看成是一个特殊形式的线性规划问题,几乎

所有的科学计算问题最内层的迭代都是求解线性规划问题，当时国外一些报道声称世界上所有计算机的运算有三分之二是用来求解线性规划问题。正因为如此，Karmarkar 算法一提出就受到研究单位、工业部门、金融机构的广泛重视。

Karmarkar 算法的提出，掀起了研究线性规划内点算法的热潮，目前内点算法已成为线性规划的研究主流。如何探讨将单纯形方法与内点技巧有机的结合，可能是线性规划方法突破的途径。另外，从理论上可证明 Karmarkar 算法在一定意义下是一个特殊的基于 log 罚函数的罚函数方法。从事非线性优化研究的人都知道，利用罚函数方法求解约束优化问题并不是很好。Karmarkar 方法的成功告诉我们，一些在一般情形下不好的方法可能对某类特殊问题非常有效。因而不要忽视那些“不好”的方法。内点算法也促使人们重新认识传统的罚函数方法，进一步深入研究罚函数。

10.4 拟牛顿方法

牛顿方法是求解无约束优化最简单的方法之一，它具有二次收敛性。但是如此一个既简单又收敛速度快的算法，并不是在实际计算中常见的方法，这是因为它有一个致命的缺点：需要计算函数的二阶导数。在变量个数为 n 时，二阶导数矩阵是一个 $n \times n$ 的对称矩阵，它也称为海色阵。牛顿法需要计算海色阵带来的问题有三：第一，计算量非常大，即每次需要计算海色阵的 $\frac{1}{2}n(n+1)$ 个元素；第二，海色阵如果不正定的话，牛顿法得到的方向可能不是目标函数的下降方向，从而会导致方法失败；第三，如果海色阵几

乎奇异，则会产生一个非常长的牛顿步，使牛顿法在实际计算中出现数值困难．但早先的不利用海色阵方法（如最速下降法）收敛非常慢．研究人员十分关心能否构造出既不需要计算海色阵但又能具有超线性收敛的方法？拟牛顿方法正是这样一类方法，但这类方法不是由著名数学家而是由一个当时在美国 Argonne 国家实验室工作的青年人 Davidon 所发现的．

Davidon 并不是基于数学推理提出该方法，他构造的方法主要是根据他对优化问题的几何理解．由于 Davidon 的结果是发表在 Argonne 国家实验室科研报告上而不在正式学术杂志上，故他的方法 1959 年提出后几年都没引起人们的注意．1962 年 Powell 利用 Davidon 的方法算了一些例子得到令他惊喜的结果．所以当他被邀请在一个会议上报告时他介绍了 Davidon 的算法以及他得到的惊人计算结果．报告完后 Flefecher 说他也在研究 Davidon 的方法．后来他们进行合作，于 1963 年在《Computer Journal》上发表了关于 Davidon 的方法的整理、分析以及计算结果．这篇文章是最优化影响最大的文章之一，被引用次数仅 SCI 记录的就已超过了千次．从此以后这个方法就被称为 DFP 方法．

DFP 方法与牛顿法的差别就是不用海色阵而用一个对称矩阵代替，为了减少计算量，该对称矩阵从一次迭代到下一次迭代的变化是一个秩 2 的矩阵，也就是说，第 k 次迭代时的对称阵加上一秩为 2 的修正项就得到了第 $k+1$ 次迭代所需的对称阵．由于这一性质，我们称该矩阵为修正矩阵．每次修正都使得到的矩阵满足一个“拟牛顿方程”或“拟牛顿公式”，它要求修正矩阵乘以最后两个迭代点之差等于这两点的导数差．从台劳公式可知，两点的导数差应近似于海色阵乘以这两点的差，所以拟牛顿方程要求修正

矩阵在上一个迭代方向上近似于海色阵.在几何上,牛顿法用在方程求根(最优化可等价于求导数为零)是切线法,而拟牛顿方法是割线法.也正因为这一性质,DFP 方法以及其他所有满足于拟牛顿修正的方法都称为拟牛顿方法.如果目标函数是严格凸的,则只要初始的修正矩阵为正定时所有的修正矩阵都是正定的.修正矩阵为正定,则算法产生的方向可理解为基于此修正矩阵的范数意义下的最速下降法.正因为如此,Davidon 将他的方法称为变尺度算法,如今,人们把所有的修正矩阵都正定的拟牛顿法称为变尺度方法.

DFP 方法在 1961 年就可解决变量为 100 个的优化问题,它比原有方法(如梯度法、单纯形法、坐标轮流搜索法等)要快的多.由于它不计算二阶海色阵,但收敛是超线性且具有二次终止性质.即对于二次函数可在 n 次迭代后求得精确解,许多优化专家在 60 年代都重视拟牛顿法的研究,在 70 年代涌现了一批突破性的研究成果.拟牛顿法是非线性优化研究在 70 年代的最热点.在 60 年代末,Broyden 等人提出了对称秩 1 算法(SR1),它是形式最简单的对称拟牛顿修正公式.1970 年 Powell 将非对称秩 1 修正公式对称化,导出了一个秩 2 修正公式,被称为 PSB 方法.另一个与 DFP 齐名的拟牛顿法是由 Broyden,Fletcher,Goldfarb,Shanno 在 1970 年独立提出,故被称为 BFGS 方法.后来,更多的拟牛顿修正公式被导出,著名的有 Broyden 簇和 Huang 簇.DFP 方法、BFGS 方法和 SR1 方法都属于 Broyden 簇.

拟牛顿方法有广泛影响是由于它速度快,能解决实际问题,它已成为当今求解中小规模(也可大到几百甚至几千个变量)问题最常用的方法.拟牛顿方法之所以有吸引力是关于它的收敛性分析.

不少收敛性结果的证明都是难度很大、技巧性很强的.例如,Powell给出的利用修正矩阵追迹以及行列式值来分析收敛性的技巧已被广泛应用,后来被 Byrd 和 Nocedal 推广称之为 ψ 函数的技巧.拟牛顿法的收敛性分析也造就了一大批著名的优化专家:如 Powell, Broyden, Dennis, More 等, Powell 于 1982 年获得了首届 Dantzig 奖(Dantzig 是发明单纯形法的著名数学家),可见拟牛顿法受到了优化界的高度重视,拟牛顿法的研究成果受到了高度评价.

拟牛顿法虽已有了许多收敛性结果,但目前仍有两个著名的问题尚未解决.第一是 DFP 方法对于凸函数在非精确搜索下是否收敛? 第二个问题是对于非凸函数 BFGS 方法能否一定找到稳定点?

关于 DFP 的全局性收敛问题,我们提出了利用修正矩阵的平方阵来估计它的增长速度,导出了一些能保证收敛的条件,证明了在许多情形下(如梯度单调变化等)DFP 方法都会收敛.我们的结果虽尚未彻底解决 DFP 的收敛性问题,但已推进了一大步,为进一步研究提供了新的技巧.

我们还从逼近的角度分析了拟牛顿算法,在一维问题上,发现拟牛顿法实质等于利用了当前点的函数值、导数值和在前一个迭代点的导数值这三条件来进行二次插值.直观看来,这样一种插值显然不如将上一点的导数插值换成在上一点的函数值插值.利用这一小的改动,我们得到了一改进的 BFGS 方法.基于这些进展,我们提出了利用三次 Hermite 插值函数的二阶导数来构造模型,在一维优化中导出了不需计算二阶导数但能保证二阶收敛的算法.将这些结果推广到高维,我们给出了一类新的计算方法——非

拟牛顿算法，为寻求新的有效求解无约束优化计算方法开辟了一条新路，做出了有意义的尝试，这方面的工作还有待进一步深入。

拟牛顿法的收敛性研究是公认的非线性优法算法理论分析中最难的、最具技巧性的。如果我们能够解决上面提到的两个遗留难题或其中之一，则会极大提高我国非线性优化研究队伍在国际上的地位，也将促进我国优化算法理论研究的进一步发展。

10.5 信赖域方法

瞎子爬山与计算机求最优值传统的优化计算方法是线搜索方法。每次迭代给出一个初始值，然后根据某些规则求出一个搜索方向，再沿该方向寻求一个较好点或最优点，将找到的点作为新的迭代点，这样反复下去直到求得一个可接收的点。线搜索法用在爬山的表现是这样的：设想有一个机器人在山脚下，它通过某种判断（利用它脚底下这一点的信息，如高度、坡度等来作出判断）决定一个前进方向，然后它有一个无限长的手在该方向上找到最高点，于是它跳过去，再重复进行选方向、决定步长，直到达到山顶。我们有理由相信，机器人爬山比起瞎子要慢得多。这个例子也说明，无论目前的超级计算机多先进，速度多快，它和人相比还不知差多远！

瞎子爬山比计算机快是因为他用拐杖一打就把脚下以及周围一米左右的情况都弄清楚了，但计算机计算函数值、导数值仅是脚底下高度、坡度。它只能利用这些信息来决定方向进行搜索。线搜索方法可能碰到的困难有以下三种：一种困难是由于了问题变态产生的搜索方向是非常大的向量，它或者导致数值上溢或搜索失败。另一种困难是由于目标函数的非线性特性很强，即便求得的搜

索方向是下降方向但却在该方向上找不到(数值上)更好的点.还有一种困难是数值误差可能使理论上应为下降方向的搜索方向是一个上升方向,这样算法无法进行下去.

信赖域方法就是为克服线搜索方法可能出现的困难而提出的一类新的计算方法.它有着很强的逼近背景.信赖域方法的基本思想是不进行线搜索,每次迭代在一信赖域内找一个试探步,然后决定该试探步是否可接受.信赖域法只在当前迭代点的一个邻域内考虑模型子问题,这是基于任何模型的近似函数(如台劳展开等)都只是在局部才与原函数很好地吻合.这可理解为我们只在这一邻域内相信近似模型,因此该邻域被称为信赖域.利用信赖域的方法也就称为信赖域方法.

信赖域方法的研究起源于 Powell 在 1970 年的工作.他提出了一个求解无约束优化的算法,该算法在每次迭代是强制性地要求新的迭代点与当前迭代点之间的距离不超过某一控制量.对步长进行控制是因为上面已提到了传统的线搜索方法常常由于步长过大而导致算法失败,当问题病态时尤为如此.正因为如此,在早期人们也把信赖域方法叫为限制步长方法.后来,研究人员发现控制步长在模型函数是二次函数时,等价于对模型函数的海色阵增加一个纯量矩阵.而这种对海色阵增加一纯量阵的技巧在著名的 Leveberg－Marquardt 方法中已被用到,这个用于求解非线性最小二乘的方法是 Levenberg 于 1944 年提出的.Marquardt 于 1963 年再发现的.也正由于这种隐含的等价关系,人们提到信赖域法的历史总要追溯到 Levenberg - Marquardt 方法.

关于信赖域法的研究是进入 80 年代才多起来的,而且很快就成为非线性规划的热点.信赖域方法虽然很新,但它已和线搜索方

法并列为目前求解非线性优化问题的两类主要数值方法.信赖域法思想新颖、算法可靠,具有很强的收敛性质.它不仅能很快地求解良态问题,而且也能有效地求解病态问题.信赖域方法是近年来非线性优化研究领域的一个重要方向,是当今寻求如何构造新的优化计算方法的主要途径.

信赖域方法推广到约束优化问题的一个主要困难是线性化的约束在信赖域内无解,于是著名的 SQP 方法(也就是拟牛顿法推广到约束优化)就不能直接与信赖域方法结合.解决这一问题的作法有三种:其一是将线性化约束的可行点集往当前点平移,然后利用零空间技巧可将子问题转化为低维空间的无约束信赖域子问题.其二是将所有线性化约束转化成一个最小二乘约束,然后要求该二次约束满足一定的下降,这样得到的子问题是在一个球与一个广义椭球的交求一个二次函数的极小值,这类子问题最早由 Celis、Dennis、Tapia 建议,故常称为 CDT 子问题.其三是考虑 SQP 的罚形式,这样导出的子问题是在信赖域内求一个非光滑函数的极小值.约束优化的另一个困难是点的好坏不是很好判断.在无约束优化评价一个点的好坏就以函数值来确定.但有约束时,点的好坏不仅要看函数值的大小而且要看约束条件满足的好坏.而且,一个出乎人们意料之外的现象,是迭代点从一个离最优点远的点移到靠近的点时,函数值变大而且约束也变坏了.这在 SQP 方法中已被发现,被称之为 Maratos 效应,是 Maratos 于 1978 年做博士论文时最先发现的.

CDT 子问题的研究得到了一些十分有意义的结果.对于凸的子问题,利用对偶规划,可将问题转化为两个变量的最大值问题,可用牛顿方法求解,对于非凸的 CDT 子问题无论是理论还是算法

方面仍有许多工作需要做.关于CDT问题的研究还使我们得到了一个有趣的关于矩阵对的结果:如果两个矩阵产生的二次型的较大者在空间任一点非负,则必存在这两个矩阵的凸组合,使得该组合得到的矩阵是半正定的.这个结果推广到多个矩阵是不成立的,但是目前尚不知道是否有弱形式的关于多个矩阵的类似结果.

信赖域方法在试探点不可接收时,完全不考虑试探步的信息似乎不太妥当,目前人们已开始对不成功的试探步进行分析与利用.最直接的一种技巧是利用该试探步上的函数值来修正逼近模型,由于在坏的试探点一般不计算导数值,所以拟牛顿修正公式不能用,因而用非拟牛顿类修正是十分合适的,另一种技巧是在不成功的试探步方向上进行线搜索.我们提出了用信赖域方法和线搜索相结合来构造新的计算方法,并依此给出了一个利用信赖域以及回溯技巧的数值计算方法.这是综合两大类方法之优点的一个大胆试探,为寻求新算法的研究闯了新路.

10.6 共轭梯度法

共轭梯度法是除了最速下降法外最简单的优化计算方法,它具有程序简单、内存小等优点,是当前求解特大规模问题的主要方法.

共轭梯度法有很好的理论性质,对于二次函数,在精确搜索下共轭梯度可以经过 n 步迭代后求到最优解.共轭梯度法的基本原理是利用产生的共轭方向将一个 n 维的优化问题转化为等价的 n 个一维问题.

但是,当线搜索不精确时,共轭性却产生不好的作用.由于共

轭性使后面的方向与前面的方向无关，于是前面迭代时因非精确搜索产生的误差将不可能在后面的迭代中修正．基于这一观察，我们提出利用二维子空间模型来修正共轭梯度法，这样就克服了由于步长不精确引起的问题．

共轭梯度法对于非二次函数有许多不同形式．著名的共轭梯度法有 FR 方法、PRP 方法、CD 方法等．关于这些方法的收敛以及计算比较一直是人们关心的问题．1983 年 Powell 给出一个漂亮的例子说明 PRP 方法对于非凸函数即使是精确搜索也会出现不收敛，但 FR 方法在同样条件下可证明是收敛的．这一结果再一次说明，有些计算方法虽然有很强的收敛性结果，但实际计算并不一定好．而有些算法虽然只有较弱的收敛结果，但在计算中非常有效．因为大量的实际计算已经表明 PRP 方法要比 FR 方法好．

由于 Powell 的工作，共轭梯度法的收敛法分析再次引起科研人员的关注．1985 年 Fletcher 的学生 Al－Baali 证明了当非精确的参数满足一定条件下 FR 方法一定收敛，这推广了 Powell 关于 FR 收敛性分析的结果．通过深入的分析，我们得到了 FR 法收敛的非精确搜索应满足的充分必要条件，并且提出了将每两步迭代结合起来考虑的分析技巧，这使共轭梯度法的收敛性分析更加简单和灵巧．对于其他著名的共轭梯度法如 PRP，CD 等方法．我们也得到了一系列收敛性结果．这些结果完善了共轭梯度法的理论性研究．

共轭梯度法的各个不同公式之间的联系是一个有趣的问题．我们将共轭梯度法与拟牛顿法相比．1959 年提出 DFP 方法，到 1970 年产生 BFGS 方法，拟牛顿法就有了带参数的一族方法．DFP，BFGS，SR1 等方法联系在一起．有了一个带参数的公式，分

析收敛性便有了统一的工具，而且也容易发现它们的内在联系与差别所在.但是，共轭梯度法的各个算法目前尚未找到统一的带参数的公式，这使得分析共轭梯度法的收敛性局限于一个一个孤立的考虑，而且很难找到这些方法之间的相互关系，特别是很难判断它们的优劣以及如何去解释它们的差别.例如 PRP 方法比 FR 方法为何有效至今尚没有令人信服的理论来证实.所以，我们认为，寻求共轭梯度法的统一公式是一件很有意义的事情，希望能早日使该问题得到圆满的解答.

10.7 直接方法

求解最优化问题的直接方法是指不需要计算导数的方法.举一个例子，设想我们要找到一个湖的最深点，但给的工具是一条小船以及一根足够长的竹竿.将船划到一处能利用竹竿测量到此处的深度，但并不知道湖底在这一点的坡度.在数学上就是只有函数值但不知道导数值.这种情况在实际中经常出现，或者是由于导数值非常难求；或者是目标函数根本就没有解析表达式从而不能计算导数.

最简单的直接方法是坐标轮流搜索方法，即沿着每个坐标轴方向依次搜索.这个方法容易编写程序，但收敛常常是非常慢的.另一个直接方法是单纯形法，这个方法与线性规划的单纯形法根本不是一回事，虽然它们有着同一个名字.求解 n 维非线性优化问题的单纯形法是利用 $n+1$ 个点组成一个多面体(称为单纯形)，然后通过多面体的翻转、扩大与缩小，逐步使该多面体移至目标函数的极小值附近.目前最有效的直接方法是共轭方向法，这方法由

Powell 在 1964 年提出，后经 Zangwill 于 1967 年以及 Brent 于 1973 年进一步改进.共轭方向法的基本思想是利用 n 次一维搜索找到一个共轭方向.在目标函数是凸二次的情形下，共轭方向法具有有限终止性.

无论是从理论上还是从实际计算效果上来看，直接算法从 60 年代共轭方向法提出以后没有本质的进展，由于大量的实际问题是不能计算导数或导数十分难求，目前的直接算法收敛速度还不是太快.所以十分有必要对直接算法进行深入的研究，争取能够早日构造出比共轭方向法更有效的直接算法.

后　记

“大规模科学与工程计算的方法和理论”是攀登计划项目(A)之一，立项运作5年，已通过国家验收．此项目共分六个课题，主要参加人员63人，在“八五”期间取得了巨大的成就，为推动我国科学计算事业的发展，为培养我国在该领域的青年人才等方面起了重要的作用．

本书只是部分地讲述了这个项目的一些内容、意义及应用于实际的情况．知识性、可读性、可用性都很强．由首席科学家主持的专家委员会讨论决定了本书的内容．全书由中国科学院院士石钟慈、研究员袁亚湘主编，由项目秘书赵静芳负责组织编写．主要供稿人有：王兴华、秦孟兆、黄明游、丁培柱、刘林、张关泉、张宇、李开泰、李德元、季仲贞、傅德薰、周天孝、袁益让、蒋尔雄、袁亚湘等．

本书有不周全、不恰当之处恳请指正．

编　委

1997年12月