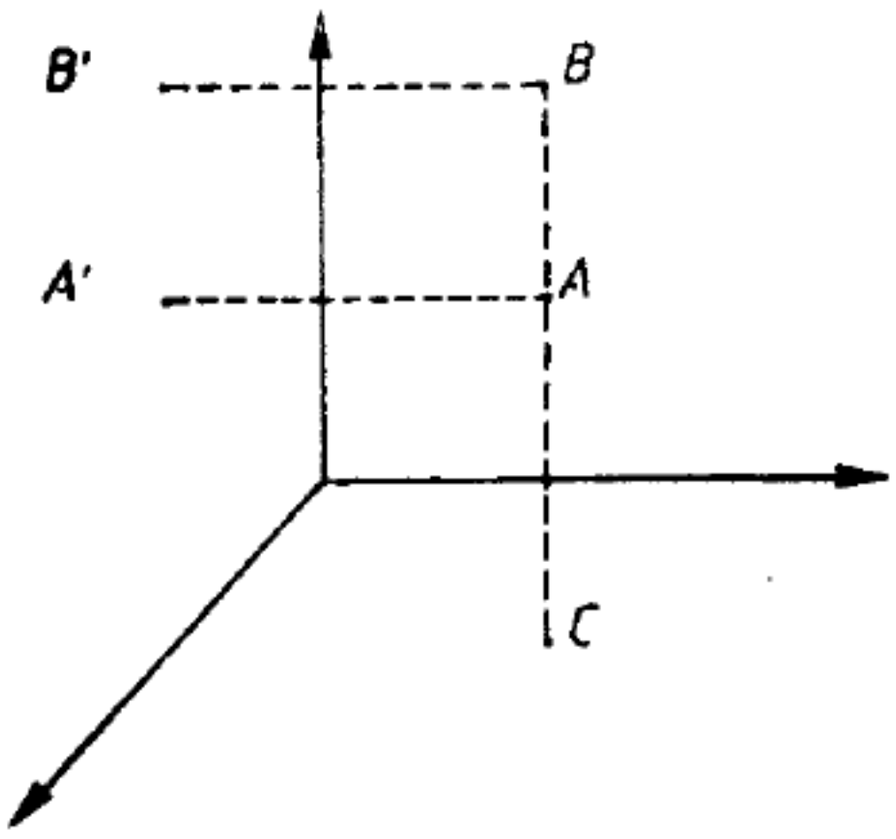
将有助于图书馆的管理者作出采集的决策：例如，他们可以决定订购那些经常被图书馆的热门期刊所引用的刊物。这些分析如果与科学计量学研究相结合，还可以确定重要的研究人员群体（不为人知晓的学院或“学派”，详细情况请见第Ⅳ编）或是一个国家内不同学科领域的分布状况。

Ⅰ.5.2.2 一种直观的主分量分析方法

我们在此面临的问题是要找出$\mathbb{R}^k$中n个点的适当显现方法。因为人类最多只能在三维空间看见目标，所以必须寻找一种能使所研究点集合的维数降低的方法。由于实际应用上的原因，通常只能用二维平面图。然而，在平面上投影将会严重歪曲原始散布图的形状。例如，将三维空间中的两个点A和B投影到垂直于AB连线的平面上就属这种情况（见图Ⅰ.5.3）。



图Ⅰ.5.3 投影

图Ⅰ.5.3这样的投影将A和B点映射在同一个点C上，说明这是A、B点集糟透了的二维表示。事实上，任何平行于直线AB的平面都能够将这一情况表现得很理想，如图Ⅰ.5.3所示，将A、B两点投影到A′、B′点。一般的散布图都要比上述两个点的简单情况复杂得多，因此无论我们取哪个平面，在实际操作中总有一些信息将会丢失。实际上，投影总会缩短点与点之间的距离（投影平面平行于所有研究点的情况除外）。我们要寻找一个平面，以使尽可

能避免这种距离的减少。

散布图内方差的定义为：任何两个不同点之间所有距离的平方和。因为有 n 个点，所以一共有$\binom{n}{2}=n(n-1)/2$段距离要考虑。

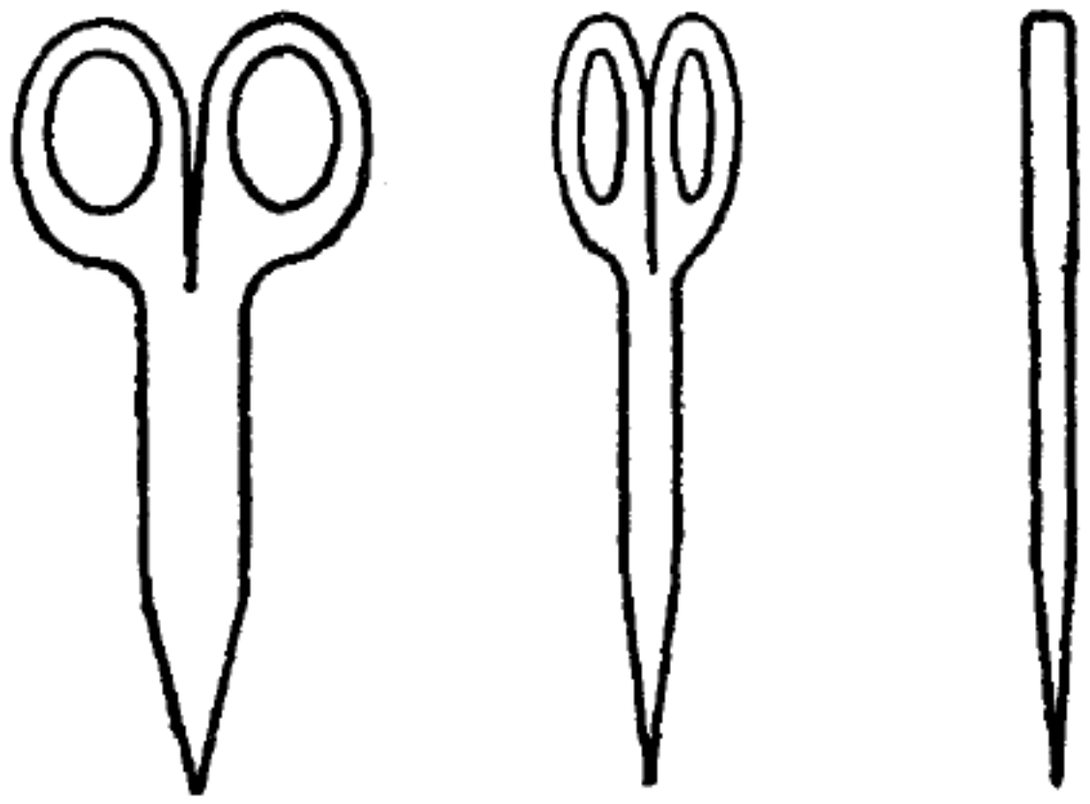
我们将试图确定一个平面，能使得投影散布图的内方差最大。

对易于识别的目标来说，哪个平面是最大平面可以一目了然。对一个很圆的球来说，任何平面都是最大平面，但是对一把剪刀来说，就没有那么多选择了（见图Ⅰ.5.4）。

应该注意的是，任何平行于最大平面的平面也是最大平面。因此，可以选择一个通过 R^k 原点$0=(0, 0, \cdots, 0)$的平面，这个平面将完全由两根互相垂直的坐标轴确定。

在实践中我们将需要一套计算机程序来满意地解决这个问题。这套程序将按以下步骤工作：首先，它要找出一根坐标轴，称其为x_1，使投影散布图的内方差相对于所有其它坐标轴来说达到最大。

接下来它要寻找垂直于x_1的第二根坐标轴x_2，使沿着垂直于x_1的所有坐标轴方向的内方差达到最大。由x_1和x_2所确定的平面就称为最大平面。按照相同的方法，程序将继续找寻第三根、第四根、第五根……第k根坐标轴，所有坐标轴都必须相互垂直。



图Ⅰ.5.4　同一把剪刀在三个不同平面上的投影

这k根坐标轴就称为散布图的主分量。在数学上，这整个过程实质就是一个寻找所谓方差—协方差矩阵C的本征值和本征向量的问题。因而第一根坐标轴就是与最大本征值相关的本征空间；第二根坐标轴是与第二个本征值相关的本征空间，依此类推。直观地说，对应于某个本征向量的本征值包含了变分总量，在观测数据的k维散布图投影到与这个本征向量相关的本征空间后，这个变分总量保持不变。

在很多情况下，第一平面（x_1，x_2）是最重要的平面。当然，在实际应用时最好还是使用几个平面（如（x_1,x_3）和(x_2，x_3)），以便获得对散布图的更好理解。这样做也有助于检测数据中的异常情况。

在实际情况下，我们将尽可能保留必要的坐标轴数（当然先从最重要的坐标轴开始），以便补偿k维散布图总方差的某一固定百分比（例如75%）。主分量是人为变量，无需具有什么具体含义或显著性。由于主分量是可度量变量的线性组合，它们本身一般不能直接度量，但是有时候可以解释（见下面的例子）。

Ⅰ.5.2.3 举例：植物学期刊的网络研究

我们在这里报告一项涉及植物学期刊引文网络的研究工作，这项研究的目标之一是证实这个引文构形是紧密的（表明植物学自身是一门强有力的学科）还是松散的（表明其它学科对植物学领域的影响）。研究中还将运用主分量分析来揭示植物学可能出现的分支学科。

根据引文判据，这项研究选择了21种期刊。虽然对“引用”与“被引”的关系都进行了研究，但是我们在这里只报告“被引”关系。因此，引用期刊是变量，而被引期刊则是目标（$\mathbb{R}^{21}$中的点）。数据取自1983年的《期刊引文报告》（有关《期刊引文报告》的详细情况请参阅第Ⅲ编）。引文矩阵见表Ⅰ.5.3。

在系统开始对主分量进行实际分析之前，要先对数据进行标准化处理，使每一个变量的平均值等于零，方差等于1。我们不打算

表Ⅰ.5.3 植物学期刊的引文矩阵C （1983年）

被引期刊	引用情况 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1. PLANT PHYSIOL	2906	382	97	682	204	1007	76	166	288	481	165	524	418	148	383	35	125	39	35	97	79
2. PHYTOCHEMISTRY	270	2115	9	119	38	134	20	23	50	48	15	82	117	20	60	8	12	13	9	18	39
3. PHYTOPATHOLOGY	42	36	1771	-	158	17	-	47	-	6	-	-	-	40	8	-	11	16	31	-	236
4. PLANTA	672	143	-	685	97	442	84	91	115	212	117	178	334	32	200	22	19	39	7	36	9
5. CAN J BOT	139	40	130	33	630	89	130	105	19	82	57	34	63	54	34	27	19	49	42	8	81
6. PHYSIOL PLANT	290	79	11	99	93	665	36	77	40	113	104	133	215	94	114	-	19	10	9	15	21
7. AM J BOT	64	26	27	40	254	60	450	73	35	47	116	37	44	21	21	25	16	105	14	-	11
8. NEW PHYTOL	118	24	18	54	312	75	63	467	13	84	79	26	54	201	20	25	-	16	12	10	14
9. ANNU REV PLANT PHYS	312	45	19	116	53	148	28	49	63	77	56	81	86	42	58	6	19	11	7	22	11
10. J EXP BOT	248	44	-	129	58	173	33	79	48	339	111	52	98	60	50	7	14	9	7	36	-
11. ANN BOT-LONDON	83	16	14	28	118	98	86	89	21	94	326	14	47	34	19	12	6	32	13	24	11
12. PLANT CELL PHYSIOL	174	62	-	65	28	141	9	6	28	28	34	466	68	-	56	6	6	14	-	13	-
13. Z PFLANZENPHYSIOL	129	63	-	60	33	183	21	16	30	64	51	60	302	16	74	7	6	9	-	-	-
14. PLANT SOIL	39	-	19	-	65	16	-	53	-	16	21	-	24	319	10	-	7	-	12	-	-
15. PLANT SCI LETT	168	44	-	80	26	102	7	18	38	32	25	44	128	10	177	-	7	-	-	10	-
16. J PHYCOL	21	13	-	9	40	-	23	36	-	-	6	-	13	-	-	132	-	-	-	-	-
17. WEED SCI	21	-	-	-	10	-	-	-	-	-	-	-	-	-	-	-	662	-	80	-	-
18. BOT GAZ	50	9	6	18	86	35	114	23	10	16	48	11	20	9	12	-	13	61	7	-	-
19. CAN J PLANT SCI	20	-	27	-	14	8	-	-	-	-	13	-	-	13	-	-	38	-	227	-	-
20. AUST J PLANT PHYSIOL	127	12	-	50	20	41	-	11	21	49	33	21	22	20	22	-	-	-	-	81	-
21. PHYSIOL PLANT PATHOL	58	24	83	8	23	8	-	13	-	-	7	-	6	-	-	-	-	-	-	-	22

在此详细叙述方法，但需要注意，程序通常会自动做这项工作或作出自由选择。

表 I .5.4　表 I .5.3中植物学期刊引文矩阵的本征值

主分量	本征值	变差%	累积变差%
1	7.71	36.69	36.69
2	2.70	12.86	49.55
3	1.87	8.91	58.46
4	1.58	7.50	65.96
5	1.33	6.36	72.32
6	1.03	4.91	77.23
7	0.95	4.52	81.75
8	0.69	3.27	85.01
9	0.67	3.17	88.18
10	0.50	2.38	90.56
.	.	.	.
21	-9.11 E-08	0.00	100.00

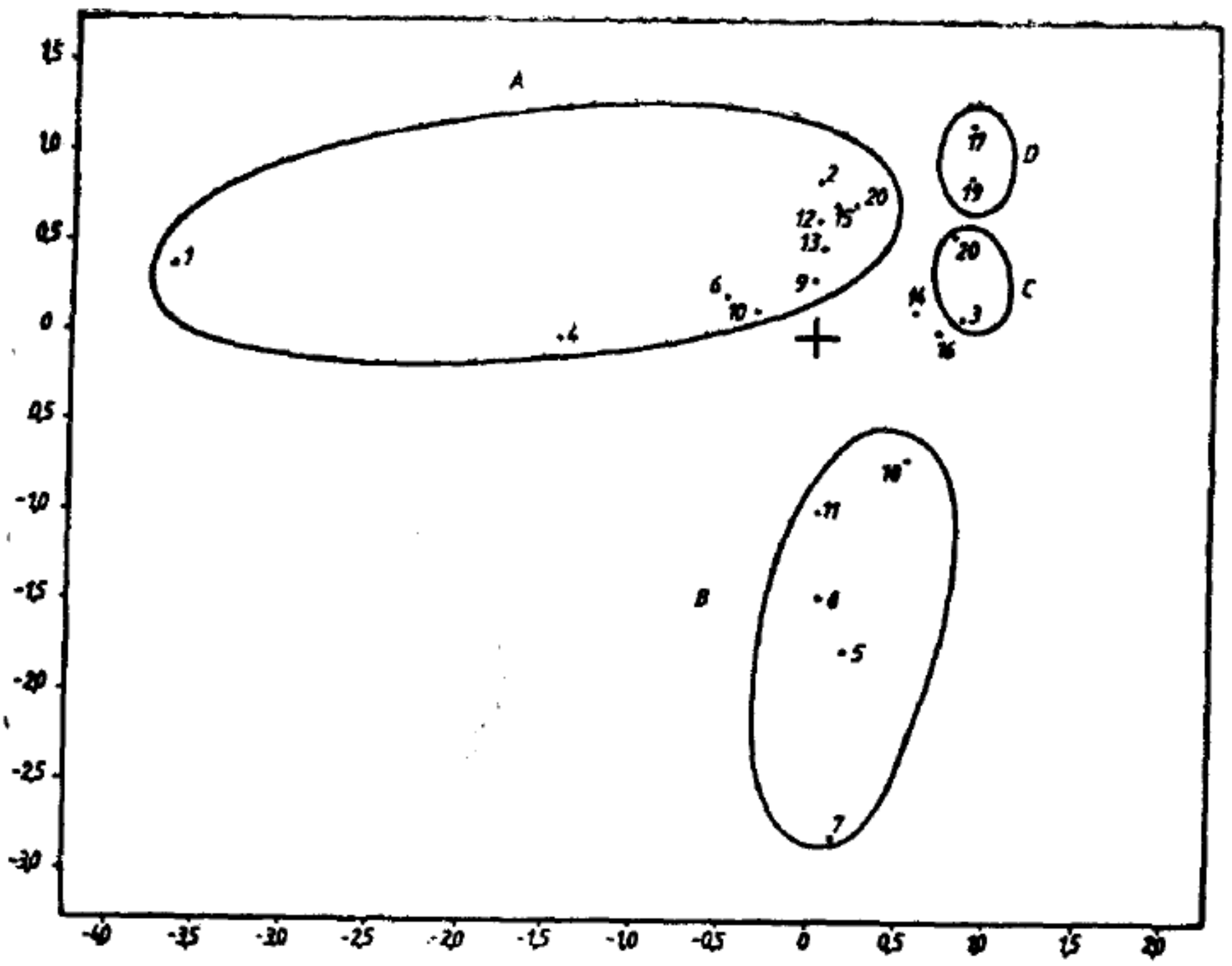


图 I .5.5　植物学期刊引文散布图在前两个本征向量平面上的投影。图中的数字所表示的意思是：

1. PLANT PHYSIOL	8. NEW PHYTOL	15. PLANT SCI LETT
2. PHYTOCHEMISTRY	9. ANNU REV PLANT PHYS	16. J PHYCOL
3. PHYTOPATHOLOGY	10. J EXP BOT	17. WEED SCI
4. PLANTA	11. ANN BOT - LONDON	18. BOT GAZ
5. CAN J BOT	12. PLANT CELL PHYSIOL	19. CAN J PLANT SCI
6. PHYSIOL PLANT	13. Z PFLANZENPHYSIOL	20. AUST J PLANT PHYSIOL
7. AM J BOT	14. PLANT SOIL	21. PHYSIOL PLANT PATHOL

表Ⅰ.5.4中是植物学期刊的本征值，这个表也列出了它们所表示的变差百分数。而且，系统绘出了在第一主分量所形成的平面上的投影图（见图Ⅰ.5.5）。

从图上直接可以看到，这个投影图只解释了变差的49.55%。但是我们从图上仍然可以得出一些有用的结论。图中被圈起来的区域表示植物学期刊的可识别类组：A＝植物生理学期刊；B＝综合性期刊；C＝植物病理学期刊；D＝应用植物学期刊。由于这些区域只是局部分离，因此可以暂时得出结论：植物学中不存在强有力的分支学科。根据这项研究工作的其它结果还可以得出结论：植物学借用了大量其它学科的研究成果（远远多于其它学科对植物学研究成果的借用）。在第Ⅰ.5.4节对聚类分析进行论述的时候，还要对相同的期刊类组进行研究。

Ⅰ.5.3 多维标度

在上节对主分量分析的研究中，我们考虑了k维空间 $\mathbb{R}^k$ 中的n个点，所研究的问题是求出“最佳”低维度表示。在情报计量学研究中，我们还会遇到更为复杂的情况，例如：

（1）我们不知道n个点的坐标，只知道它们的距离矩阵，即任意两个不同点之间的所有n(n－1)/2段距离。现在的任务与主分量分析相同：求出最佳二维表示。

（2）这个距离矩阵有时由不同方法度量的距离所组成（即所谓非欧几里德距离）。人们也经常研究相似性测度以及相应的相似性矩阵。问题仍然是相同的，但是，目标之间的相似性越大，则它们之间就显得越紧密。

处理这些问题的技巧称为“多维标度技术”。

Ⅰ.5.3.1 距离

令C = (c_{ij})是原始数据的一个（n，k）矩阵，矩阵中的n行表示k维空间的n个点C_i = (c_{i1}，c_{i2}，…，c_{ik})，i = 1，…，n。令X是这n个点的集合{C_1，C_2，…，C_n}。

度量（或“距离函数”）是一个映射 d：$X \times X \rightarrow \mathbb{R}^k$，满足下列三个条件（公理）：

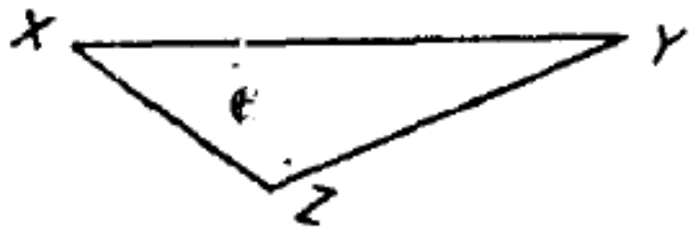
（1）对每一个x，y∈X：当且仅当x = y时，α（x，y）= 0 。这条公理表明，只有当两个点重合时，两个点之间的距离才等于零。

（2）对每一个x，y∈X有：d(x，y) = d(y，x)。

这项等式要求点x与点y之间的距离必须等于点y与点x 之间的距离。这意味着距离函数必须是对称的。

（3）对于每一个x，y，z∈X有：d(x，y)≦d(x，z) + d（z，y)。

这个不等式称为“三角不等式”，并可用图Ⅰ.5.6表示。



图Ⅰ.5.6 x与y之间的距离小于x与z及y与z之间的距离之和

配备有度量d的集合X用(X，d）表示，称为“度量空间”。

举例：

a。一般的距离函数D_0定义为：D_0(x，y) = 1(当x≠y时）和D_0(x，y) = 0(当x = y时）。虽然极为简单，但是在寻找向量理想匹配偶时（例如当检索由固定的词集所标引的文献时），这个函数是基础函数。

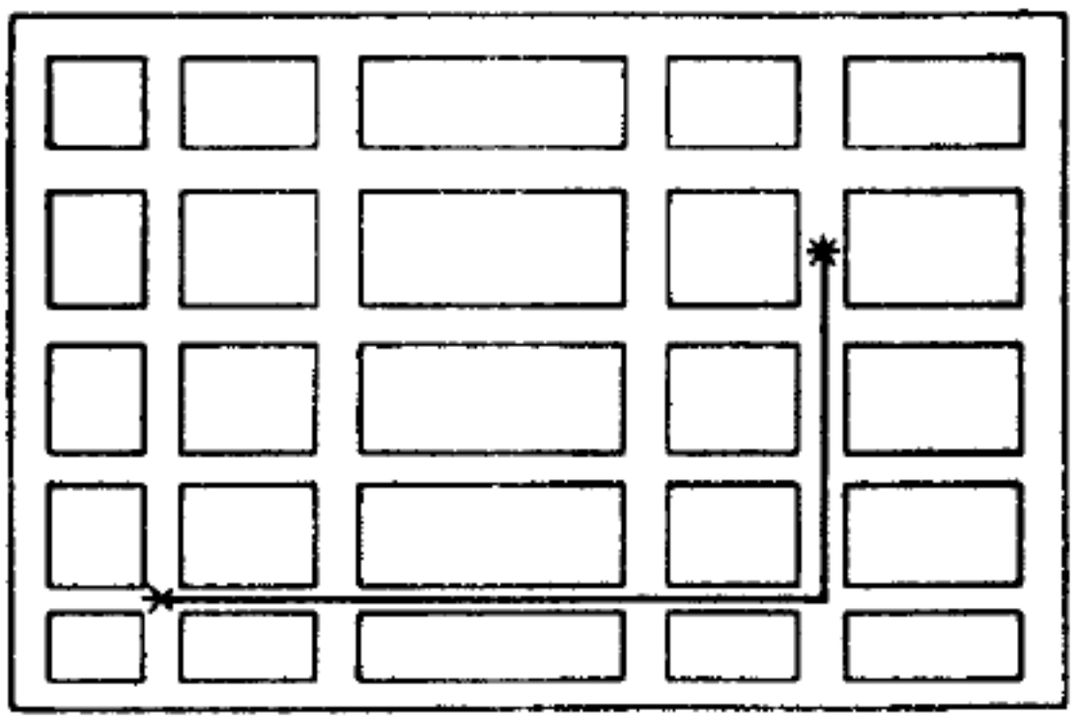
b。闵科夫斯基（Minkowski）度量d_p，p＞o。这个距离函数定义为：

$$d_p(C_i,C_j) = \left(\sum_{r=1}^{k} |c_{ir} - c_{jr}|^p\right)^{1/p} \quad . \qquad [I.5.7]$$

若取p＝2，d_p就变成了欧几里得度量。若取p＝1，就成了所谓城市街区度量：

$$d_1(C_i,C_j)=\sum_{r=1}^{k}|c_{ir}-c_{jr}|\ , \qquad [I.5.8]$$

这个度量可由图Ⅰ.5.7表示。



图Ⅰ.5.7　城市街区度量

c．契比晓夫（Chebysheff）距离，用d_∞表示：

$$d_\infty(C_i,C_j)=\max\{|c_{i1}-c_{j1}|,|c_{i2}-c_{j2}|,\ldots,|c_{ik}-c_{jk}|\}\ . \qquad [I.5.9]$$

实际上可以证明 $\lim_{p\to\infty} d_p(C_i,C_j)=d_\infty(C_i,C_j)$. 这样就解释了符号$d_\infty$。

这些广义的距离d_p和d_∞既可应用于经典布尔（Boolean）算子归纳，也可用于对情报计量学和经济计量学中集中与分散测度的研究。

Ⅰ.5.3.2　相似性与不相似性

距离测度可以认为是不相似测度。直观地讲，目标之间的距离越大，它们的不相似性也就越大。从形式上说，如果

（1）对X中的每一个x有：d(x，x)＝0；

（2）对X中的每一个x，y有：d(x，y)＝d(y，x)，则函数$d:X\times X\to R^+$称为“不相似函数”。

因为我们已经从上述公理中除去了三角不等式，这样便清楚地概括了距离的概念。此外，事实上不相同的两个项也可以使不相似性等于零。

一个相似函数$s: X \times X \to [0,1]$满足。上面第（2）条并且有$s(x,x)=1$（对于X中的每一个x）。如果s是一个相似函数，那么$1-s$就是一个不相似函数；如果d是一个不相似函数，则

$\frac{2}{\pi}\mathrm{Arctg}\left(\frac{1}{d}\right)$就是一个相似函数。

例：考虑A、B两人对问卷调查的下列答卷（见表Ⅰ.5.5，表中的Y表示“是”，N表示“非”）：

表Ⅰ.5.5　问卷调查的答卷

	问题							
被调查人	1	2	3	4	5	6	7	8
A	N	Y	Y	N	N	Y	Y	N
B	Y	N	Y	N	Y	Y	N	Y

将表Ⅰ.5.5转换为以下列联表（表Ⅰ.5.6）：

表Ⅰ.5.6　表Ⅰ.5.5中数据的列联表

		A	
		Y	N
B	Y	a = 2	b = 3
	N	c = 2	d = 1

描述A、B二人之间相似性的测度结果是：

$$s_1(A,B)=\frac{a+d}{a+b+c+d} \qquad [\text{Ⅰ.5.10}]$$

或

$$s_2(A,B)=\frac{a}{a+b+c} \qquad [\text{Ⅰ.5.11}]$$

其它的相似性测度如萨尔顿(Salton)余弦测度和捷卡德(Jaccard)指数将在引文与同引分析（第Ⅲ编）中进行研究。

标准化

不同的数据标度会得出不同的结果。下面将以表Ⅰ.5.7为例来说明这个问题。

表Ⅰ.5.7 图书馆数据

图书馆	借阅量 (X100)	图书总数 (X1000)
A	80	169
B	82	183
C	84	175

使用欧几里得距离公式［Ⅰ.5.7］(p=2)可得：$d_{AB}=14.14$，$d_{Ac}=7.21$，$d_{Bc}=8.25$，因此$d_{AB}>d_{Ac}$。这说明图书馆A与B之间不如图书馆A与C相似。但是，因为使用了不同的标度，可以得到表Ⅰ.5.8，如果用d′表示这种情况下的欧几里得距离可得：$d'_{AB}=2.005$，$d'_{Ac}=4.000$，$d'_{Bc}=2.002$，这样就会得出矛盾的结果，即$d'_{AB}<d'_{Ac}$。

表Ⅰ.5.8 图书馆数据（不同标度）

图书馆	借阅量 (X100)	图书总数 (X100000)
A	80	1.69
B	82	1.83
C	84	1.75

这种能够引起错误结果的矛盾可以通过数据的标准化处理得到解决：将某一列数据的每一个值都除以该列数据的标准偏差。在上

述关于图书馆数据的例子里可以计算出：$S_{\{80,82,84\}}=1.633$，$S_{\{169,183,175\}}=5.735$。由此可得表Ⅰ.5.9。

如果对表Ⅰ.5.8中的数据作标准化处理，也可以得到与表Ⅰ.5.9相同的结果。现在有$d_{AB}=2.73$和$d_{Ac}=2.67$，表明A和C比A和B更相似。这里想强调的是，相似性只是一个相对概念。相似性是由所研究的点（图书馆、文献、人员）的集合所确定的。

表Ⅰ.5.9 图书馆数据（标准化值）

图书馆	借阅量 (X1.633)	图书总数 (X5735)
A	48.99	29.47
B	50.21	31.91
C	51.44	30.52

Ⅰ.5.3.3 主坐标分析

主坐标分析法也称为“经典多维标度法”或“度量多维标度法”。它要解决的是第Ⅰ.5.3节概述中的问题（1）。浅显地讲，这个问题可以表述为：城市（图书馆、期刊、科学家）之间的距离是已知的，问题是要重新建立地图。更概括地说，这个问题形成于一个空间，而这个空间的维数k也是未知数。问题的部分解包括找出一个最小维数k，以便使问题在R^k中有解。

在这里我们不讨论解题的数学细节（因为这要求相当复杂的矩阵知识）。原则上我们希望解是k维空间点的构形，这样就可以利用主分量分析（PCA）来解这个构形。许多计算机程序都可以直接解决这个问题，立即得出二维表达式。

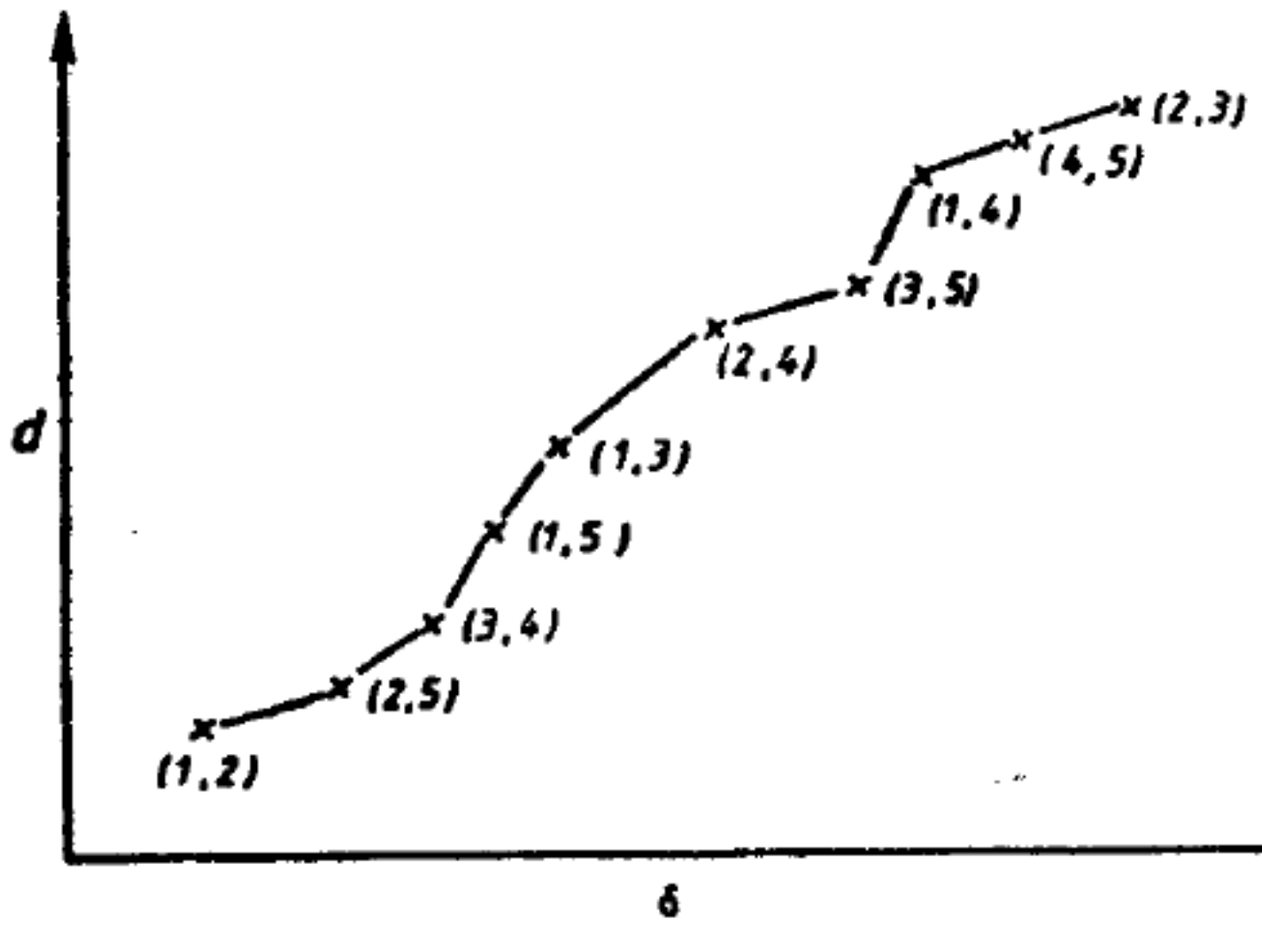
Ⅰ.5.3.4 非度量多维标度

非度量多维标度法可用来解决第Ⅰ.5.3节概述中的第二个问题。在这种情况下，我们有不相似矩阵$D=(\delta_{ij})_{i,j=1,\ldots,n}$。

所进行的一些试验和所出现的误差表明，寻找维数k以便使 R^k 中的n个点精确位于距离δ_{ij}的工作所要求的条件太多。因此，我们将试图找出某个 $\mathbb{R}^k$ 中的n个点，使每一个i，j，k，l满足

$$d_{ij} \leqq d_{kl} \Rightarrow \delta_{ij} \leqq \delta_{kl} \qquad [I.5.12]$$

这里d_{ij}是 $\mathbb{R}^k$ 中第i和第j个点之间的距离。式［I.5.12］中的要求称为“单调约束”。但是即使是这样的要求也可能太过分。如果是这样的话，那就只好尽可能满足［I.5.12］。如果［I.5.12］可以满足，我们就可以在（δ_{ij}，d_{ij}）平面上得到一个递增的图形（见图I.5.8）。



图I.5.8　距离-不相似性散布图（满足单调约束）

怎样才能获得这样的结果呢？我们先从R^k中的n个点 $(Y_i)_i = 1,\ldots,$ n着手（k实际上也是变化的，但是为了简化起见，我们使k保持不变），计算出所有的距离$d_{ij} = ||Y_i - Y_j||$，并作$d_{ij} - \delta_{ij}$图。在一般情况下，这将不会得出一个递增函数。取有相同横坐标δ_{ij}的中值，直到获得递增函数为止。（这一步实际上是运筹学方法的应用，请参阅第Ⅱ编）。这一过程通常还需要进行若干步迭代才能完成。

I.5.3.5　举例

1．麦克格拉斯（McGrath）1986年将多维标度（MDS）应用

于图书馆设计和图书馆部门划分。

2．斯莫尔（Small）及加菲尔德（Garfield）等人用多维标度法建立了文献、学科及研究人员图。数据根据同引频率确定（参见第Ⅲ编）。作为相似性的测度，麦克凯恩（McCain）还用同引数据通过多维标度法给出了显示学科的作者图。

3．宫本和中山也利用引文数据和多维标度法对工程类期刊进行了研究。

多维标度法常与聚类分析结合使用。这将是下一节的主要内容。

Ⅰ.5.4 聚类分析

聚类分析是最常用的多元方法之一，它的研究起点也是原始数据的矩阵C，目标同样是获得n个点的k维散布图的二维表示。因此，聚类分析属于降低维数技术的范畴。不过，聚类分析主要与自然群（聚合）的识别有关，而不是与k维构形本身有关。

聚类分析的结果是一种树状结构，称为“树状图”。人们常感兴趣的是构形和聚合。好在文献中有许多这方面的例子，在这些例子中，聚类分析是与主分量分析、多维标度或因素分析结合使用的。

Ⅰ.5.4.1 基本要点

虽然科学家们已经创造了一些不同的聚合方法，但是许多方法都有共同的特征。在这一小节中我们将描述这些基本要点。

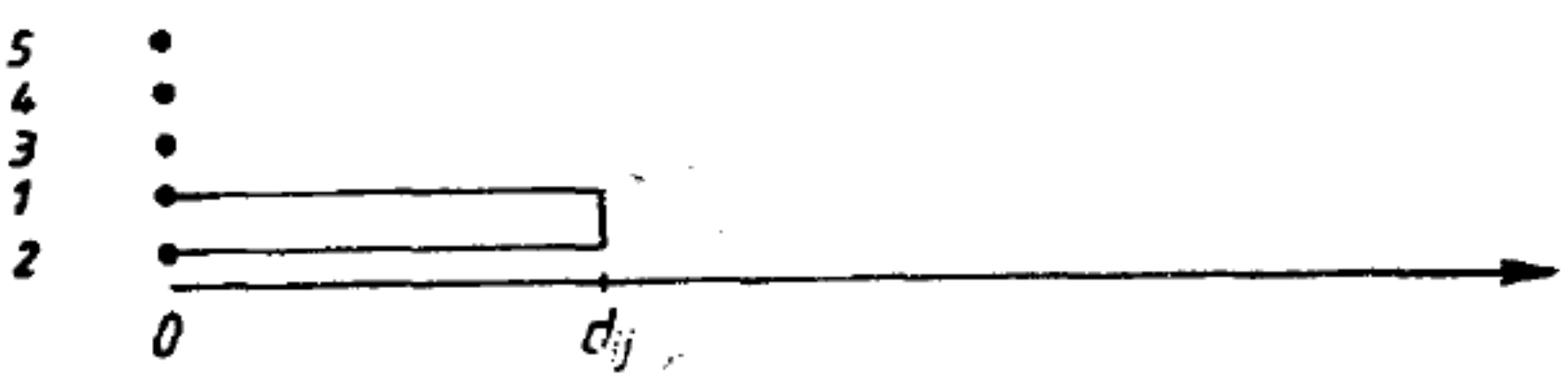
令C是一个原始数据矩阵

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1k} \\ c_{21} & & & \\ \vdots & & & \\ c_{n1} & & & c_{nk} \end{pmatrix} . \qquad [\text{I.5.13}]$$

然后将这个矩阵转换成标准矩离或不相似矩阵$D_1 = (d_{ij})$。在这

个矩阵中，$d_{ii}=0$，$i=1,\cdots,n$；$d_{ij}=d_{ji}$（D_1是对称的）。在此基础上，我们将每一个点看作是一个分离的聚类，这样可以形成更大的聚类，一个又一个，直到获得一个包含所有点的聚类。

点i和j（使d_{ij}成为D_1中的最小非零项）组合成一个聚类，用（i，j）表示，如图Ⅰ.5.9所示（这里i＝1，j＝2）。

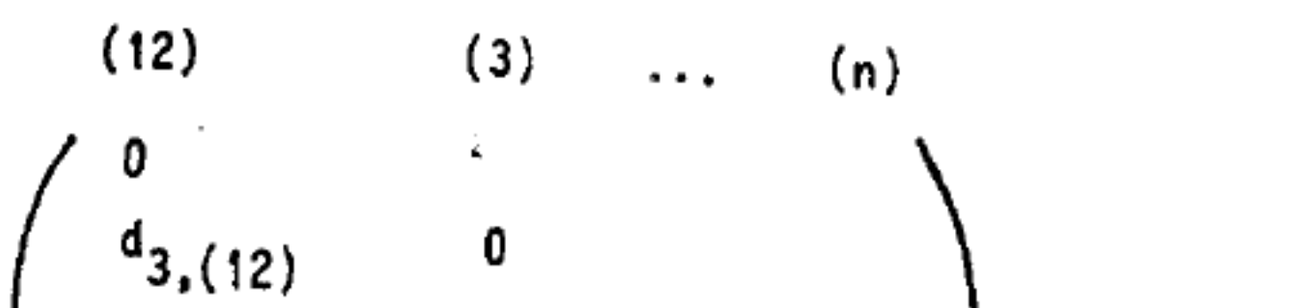


图Ⅰ.5.9　第一次聚合后的树状图

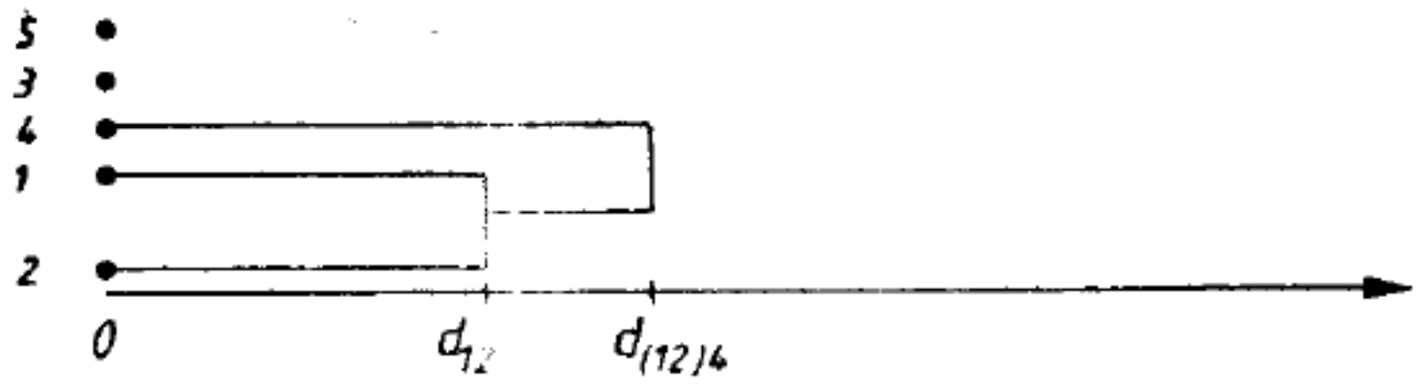
紧接着要对新矩阵$D_2=(d_{ij}^{(2)})$进行计算：对于不同于点i和j的两个点k，l有：$d_{kl}^{(2)}=d_{kl}$。从一个点到新聚类（i，j）的距离要计算新d值。这方面的一些不同计算方法将在下一小节中进行说明。假定D_2的形式如下：

$$D_2=\begin{array}{c} \\ (12) \\ (3) \\ \vdots \\ (n) \end{array}\begin{array}{c} \begin{array}{cccc} (12) & (3) & \cdots & (n) \end{array} \\ \begin{pmatrix} 0 & & & \\ d_{3,(12)} & 0 & & \\ \vdots & \vdots & & \\ d_{n,(12)} & d_{n3} & \cdots & 0 \end{pmatrix} \end{array}. \qquad [\text{I.5.14}]$$

在D_2中取最小非零值，可以得出下一个聚类。在这里可以有两种不同的选择方式：或是与（12）相邻形成一个新聚类（例如（45）），或是用一个新点（例如4）连接已经形成的聚类并产生一



图Ⅰ.5.10　树状图

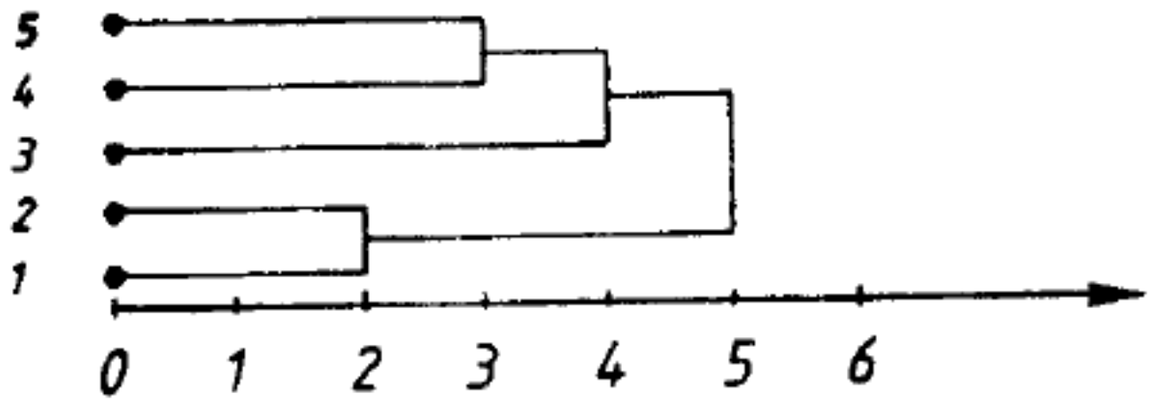


图Ⅰ.5.11　树状图

个聚类（124）。这两种选择方式分别示于图Ⅰ.5.10和图Ⅰ.5.11。

这个过程要一直进行下去，直到最后只剩下一个聚类，如图Ⅰ.5.12所示（以图Ⅰ.5.10为基础）。

最后，要对树状图进行分析，以便找出自然聚类，最好是那些可以解释的聚类，这意味着要找到一个相对长的区间，在这个区间中没有聚类形成。树状图在该点被截断，自然聚类出现。从图Ⅰ.5.12可以看到，（12），（3）和（45）是三个自然聚类。截



图Ⅰ.5.12　所有聚点集合的树状图

断树状图的方法称为“停止规则”。

Ⅰ.5.4.2　聚类法概要

本节所讨论的几种聚类法只是在定义聚类间距离的方式上有所不同。

Ⅰ.5.4.2.1　单连接法（最近邻域法）

我们将通过一个简单的（非标准化的）例子来说明这一方法。令$D=D_1$是由数据矩阵C得来的不相似（常为矩离）矩阵：

$$D_1 = \begin{array}{c} \\ (1) \\ (2) \\ (3) \\ (4) \\ (5) \end{array} \begin{array}{c} \begin{array}{ccccc} (1) & (2) & (3) & (4) & (5) \end{array} \\ \begin{pmatrix} 0 & & & & \\ 2 & 0 & & & \\ 6 & 5 & 0 & & \\ 10 & 9 & 4 & 0 & \\ 9 & 8 & 5 & 3 & 0 \end{pmatrix} \end{array} .$$

这里$d_{12}=2$是不为零的最小值。因此（1）和（2）组合成一个聚类。聚类之间的距离定义为第一个聚类的元素与第二个聚类的元素间所有距离的最小值（没有与别的点连接的点被认为是包含一个元素的聚类）。这项规则在这里可得出如下结果：

$$d_{(12)3} = \min\{d_{13}, d_{23}\} = d_{23} = 5 ,$$
$$d_{(12)4} = \min\{d_{14}, d_{24}\} = d_{24} = 9 ,$$
$$d_{(12)5} = \min\{d_{15}, d_{25}\} = d_{25} = 8 ,$$

由此可以导出以下新矩阵：

$$D_2 = \begin{array}{c} \\ (12) \\ (3) \\ (4) \\ (5) \end{array} \begin{array}{c} \begin{array}{cccc} (12) & (3) & (4) & (5) \end{array} \\ \begin{pmatrix} 0 & & & \\ 5 & 0 & & \\ 9 & 4 & 0 & \\ 8 & 5 & 3 & 0 \end{pmatrix} \end{array} .$$

在这个矩阵中，$d_{45}=3$是最小值，并且是严格的正数。这样就可以得出聚类（45）。然后再计算可得$d_{(12)(45)}=8$和$d_{3(45)}=4$。新的D矩阵变为：

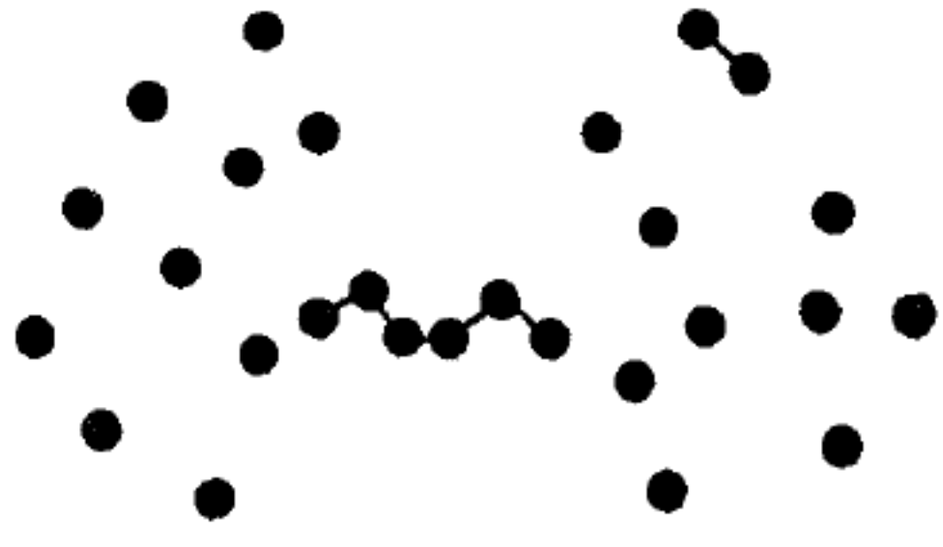
$$D_3 = \begin{array}{c} \\ (12) \\ (3) \\ (45) \end{array} \begin{array}{c} \begin{array}{ccc} (12) & (3) & (45) \end{array} \\ \begin{pmatrix} 0 & & \\ 5 & 0 & \\ 8 & 4 & 0 \end{pmatrix} \end{array} .$$

在这里$d_{3(45)}=4$是最小值，得出聚类（345）和 $d_{(345)(12)}=5$。从而有

$$D_4 = \begin{matrix} \\ (12) \\ (345) \end{matrix} \begin{matrix} \begin{matrix} (12) & (345) \end{matrix} \\ \begin{pmatrix} 0 & \\ 5 & 0 \end{pmatrix} \end{matrix} .$$

最后，（12）和（345）是可以聚合的。这种聚合过程如图Ⅰ.5.12的树状图所示。

“单连接法”的主要缺点是，聚合有时发生得太快，如图Ⅰ.5.13所示。这种形成没有什么内凝聚的松散约束聚类的趋势称为“链”。



图Ⅰ.5.13 链

Ⅰ.5.4.2.2 全连接法（最远邻域法）

全连接法与单连接法的不同之处在于：全连接法中聚类间的距离被定义为第一个聚类中的元素与第二个聚类中的元素之间所有距离的最大值。聚类本身仍然是在聚类间最短“距离”的基础上形成的，这与单连接法相同。

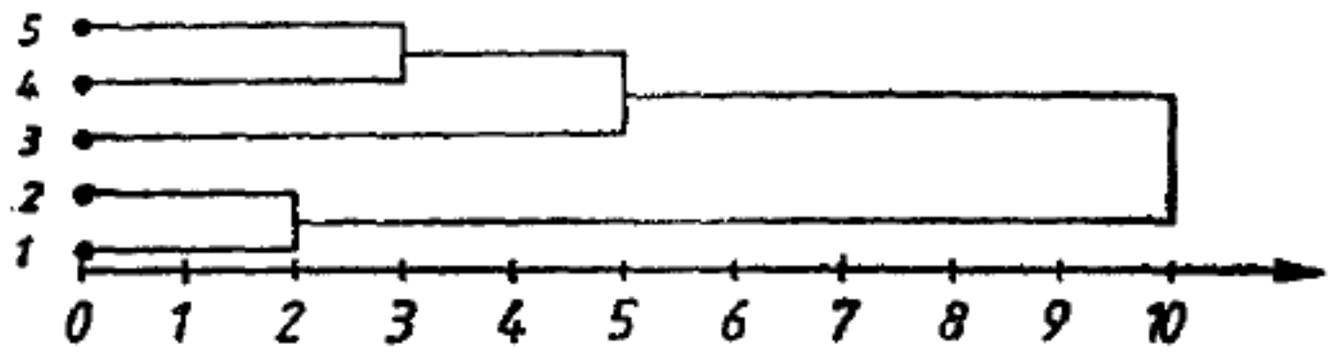
下面的一系列矩阵表示的是第Ⅰ.5.4.2.1小节中矩阵D的“全连接法”。图Ⅰ.5.14表示的是相应的树状图。

$$D_1 = \begin{matrix} \\ (1) \\ (2) \\ (3) \\ (4) \\ (5) \end{matrix} \begin{matrix} \begin{matrix} (1) & (2) & (3) & (4) & (5) \end{matrix} \\ \begin{pmatrix} 0 & & & & \\ 2 & 0 & & & \\ 6 & 5 & 0 & & \\ 10 & 9 & 4 & 0 & \\ 9 & 8 & 5 & 3 & 0 \end{pmatrix} \end{matrix} .$$

$$D_2 = \begin{array}{c} \\ (12) \\ (3) \\ (4) \\ (5) \end{array} \begin{array}{c} \begin{array}{cccc} (12) & (3) & (4) & (5) \end{array} \\ \begin{pmatrix} 0 & & & \\ 6 & 0 & & \\ 10 & 4 & 0 & \\ 9 & 5 & 3 & 0 \end{pmatrix} \end{array},$$

$$D_3 = \begin{array}{c} \\ (12) \\ (3) \\ (45) \end{array} \begin{array}{c} \begin{array}{ccc} (12) & (3) & (45) \end{array} \\ \begin{pmatrix} 0 & & \\ 6 & 0 & \\ 10 & 5 & 0 \end{pmatrix} \end{array},$$

$$D_4 = \begin{array}{c} \\ (12) \\ (345) \end{array} \begin{array}{c} \begin{array}{cc} (12) & (345) \end{array} \\ \begin{pmatrix} 0 & \\ 10 & 0 \end{pmatrix} \end{array}.$$



图Ⅰ.5.14　表示最远邻域法的树状图

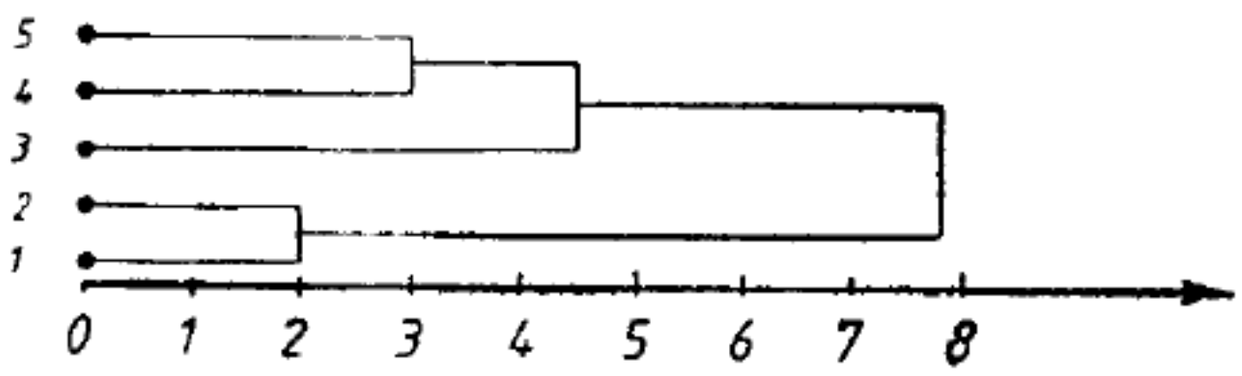
Ⅰ.5.4.2.3　群平均聚类法（平均连接）

群平均聚类法是介于上述两种方法之间的一种方法：聚类之间的距离用平均值定义。如果聚类A 与聚类B 合并，则从聚类C 到这个新聚类（AB）的距离被定义为C中的所有点和（ AB ）中的所有点之间全部距离的平均值。应用于上述例题，“群平均聚类法”可以得到以下矩阵：

$$D_2 = \begin{array}{c} \\ (12) \\ (3) \\ (4) \\ (5) \end{array} \begin{array}{c} \begin{array}{cccc} (12) & (3) & (4) & (5) \end{array} \\ \begin{pmatrix} 0 & & & \\ 5.5 & 0 & & \\ 9.5 & 4 & 0 & \\ 8.5 & 5 & 3 & 0 \end{pmatrix} \end{array} .$$

$$D_3 = \begin{array}{c} \\ (12) \\ (3) \\ (45) \end{array} \begin{array}{c} \begin{array}{ccc} (12) & (3) & (45) \end{array} \\ \begin{pmatrix} 0 & & \\ 5.5 & 0 & \\ 9 & 4.5 & 0 \end{pmatrix} \end{array} .$$

$$D_5 = \begin{array}{c} \\ (12) \\ (345) \end{array} \begin{array}{c} \begin{array}{cc} (12) & (345) \end{array} \\ \begin{pmatrix} 0 & \\ 7.83 & 0 \end{pmatrix} \end{array} .$$



图Ⅰ.5.15 群平均聚合树状图

Ⅰ.5.4.3 沃德（Ward）误差平方和法

Ⅰ.5.4.3.1 方法

沃德误差平方和法与前面的几种方法略有不同。相比之下，一步一步减少聚类数的基本原理仍然相同，但是这里所定义的点与点之间的距离（不相似性）却不是聚类之间的差。

我们将叙述如何利用沃德算法，把k个聚类降为k－1 个聚类。

第一步：取任意两个聚类并将它们合并成一个聚类，从而得到k－1个聚类，如C_1，C_2，…，C_{k-1}。

第二步：对于由点$X_{1\,1}$，…，X_{in_i}（这里$n_i = \#C_1$）所组成的聚类C_1来说，定义

$$X_1 = \frac{1}{n_1} \sum_{j=1}^{n_1} X_{1j} \qquad [\text{I}.5.15]$$

点X_1就是这个聚类的重心（通常不是C_1的点）。

第三步：对于i = 1到k − 1，取

$$ESS_1 = \sum_{j=1}^{n_1} d\,(X_{1j},\ X_j)^2 \qquad [\text{I}.5.16]$$

式中的d与确定点之间距离时的定义相同，ESS表示“误差平方和”。

第四步：取

$$E = \sum_{i=1}^{k-1} ESS_i \qquad [\text{I}.5.17]$$

第五步：对两个K聚类的每一种可能的聚合重复上述步骤。

第六步：保留使E取得最小值的聚合。要注意的是在这一步中要考虑K（K − 1）/2种可能的组合。

I.5.4.3.2 一种直观解释：“树与木”

对于完全无聚合的情况，所有的“树”都将完全呈现，但是“木”却都是未知的。在这种情况下，E = 0。

构成聚类可使“木”出现，但是“树”逐渐消失。不过，沃德法可使“树”的消失降到最低（E为最小值），从而保留信息的最大值。尽管这种方法最终可以使所有点聚合（只看到“木”），但是真正保留下来的聚类才具有良好的特性。这就直观地解释了沃德法常被认为是最优方法的原因。

I.5.4.4 聚类法的一般特性

a. 这里讨论的所有聚类法都具有这样的性质，即在 j − 1 级水平的连接距离小于在 j 级水平的连接距离。这是一个良好的特性，可以提高树状图的可评价性。

b. 在这里所提到的聚类法都有一个共同点，即对每一个j来说，n个点在j − 1级水平的划分要比在j级水平的划分细得多。

c．此外，这些方法是分级的，即一旦 j 级水平通过之后，这一水平之前的局部树状图不再会因为更进一步的内聚合活动而有任何改变。值得注意的是还有一些非分级聚类法（例如“单道”迭代聚类法），但是这些方法一般说来都不如分级法有效。

d．如果树状图中两个点之间的距离被定义为第一级水平（在坐标轴上），在这种情况下这两个点出现在相同的聚类中，则这个距离（用d′表示）满足度量的所有公理，并且对所有i，j都有：

$$d'(i,j) \leq \max_{k} (d'(i,k),d'(k,j)) . \qquad [I.5.18]$$

这一类度量称为“超度量”。从这个意义上说，聚类法可以认为是从一个度量空间向另一个超度量空间的转换。

I.5.4.5 聚类法的评价

上面所讨论的各种聚类法都能形成聚类。如果存在一个不能形成聚类的大区间，则这时的评价就很容易：见图 I.5.16。但是，事情往往并非那么简单。

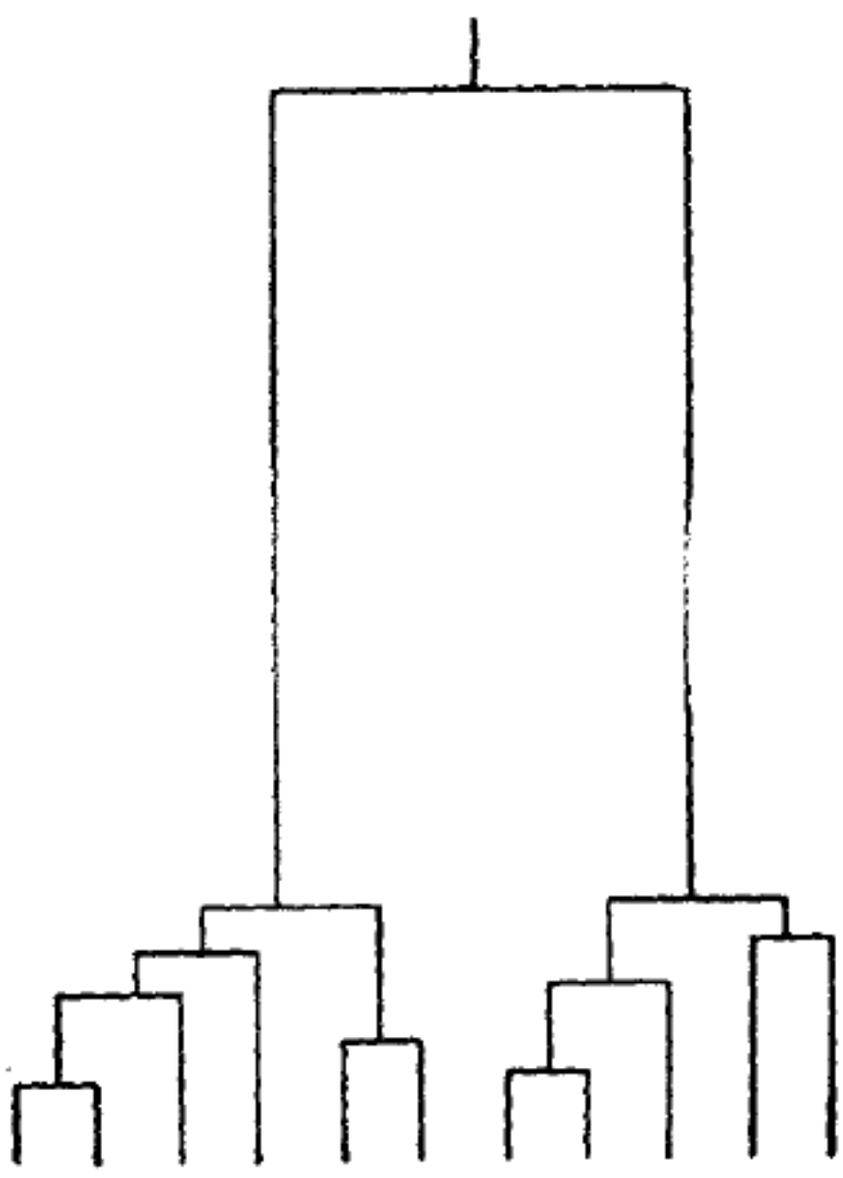


图 I.5.16 产生两个明显聚类的树状图

对聚类法的结果进行评价的主要问题涉及人工制品的出现（聚类主要是应用技术的结果）、聚类结构的稳定性以及对结果的解释。

肖（Shaw，1985，1987）等人研究了文献同引和共著者图中聚类的有效性。他们将随机图假设作为虚假设，假设图形的线条是从所有可能的线条集合中随机选择的。如果线条是随机的，则有强烈的迹象表明数据中将不存在聚合结构。杜伯斯（Dubes，1987）报告了用一些特殊结构指数估计聚类数量的蒙特-卡罗(Monte-Carlo)（模拟）研究结果。此外，他发现，用全连接聚合法判别真实聚类数显然要比单连接法好。

Ⅰ.5.4.6 举例

1．图Ⅰ.5.17是一个根据沃德法确定的植物学期刊树状图。植物学不是一门孤立的学科，其分支也并不是总能清楚定义的。当我们将树状图沿着虚线截开时，可以发现图Ⅰ.5.5所示的聚类情况。

2．阿姆斯（Arms，1978）等人用引文研究了社会科学期刊的聚类结论是：以引文为基础的聚类分析并不是设计社会科学二次服务的实用方法。多德洛夫（Todorov，1986）等人用群平均法寻找在物理学方面具有相似出版特点的国家群。鲁索（Rousseau，1989）以“重要性”为基础研究了药物学期刊的聚类问题（用单连接法）。

3．格里菲思（Griffiths）等人报告了涉及文献自动分类方面的应用。在这项研究中，沃德法得到了最成功的应用。在其他人对文献自动分类的检索所进行的研究中，主要应用的是单连接法，但是也对群平均法和全连接法进行了讨论。

4．斯莫尔等人对ISI（美国科学情报学会）的引文和同引的研究主要使用了单连接法。

Ⅰ.5.4.7 多维标度法与聚类法的组合

当将多维标度法与聚类法组合时，多维标度或主分量分析就可

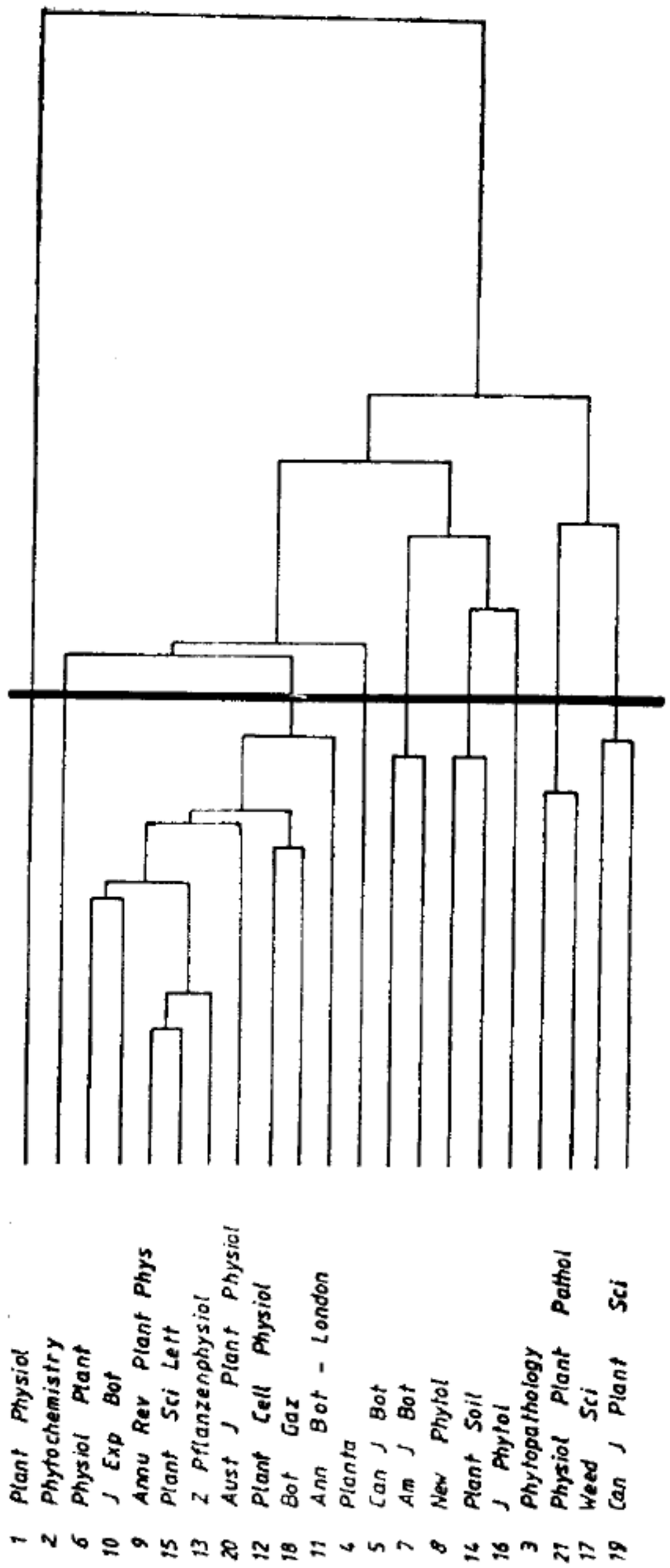
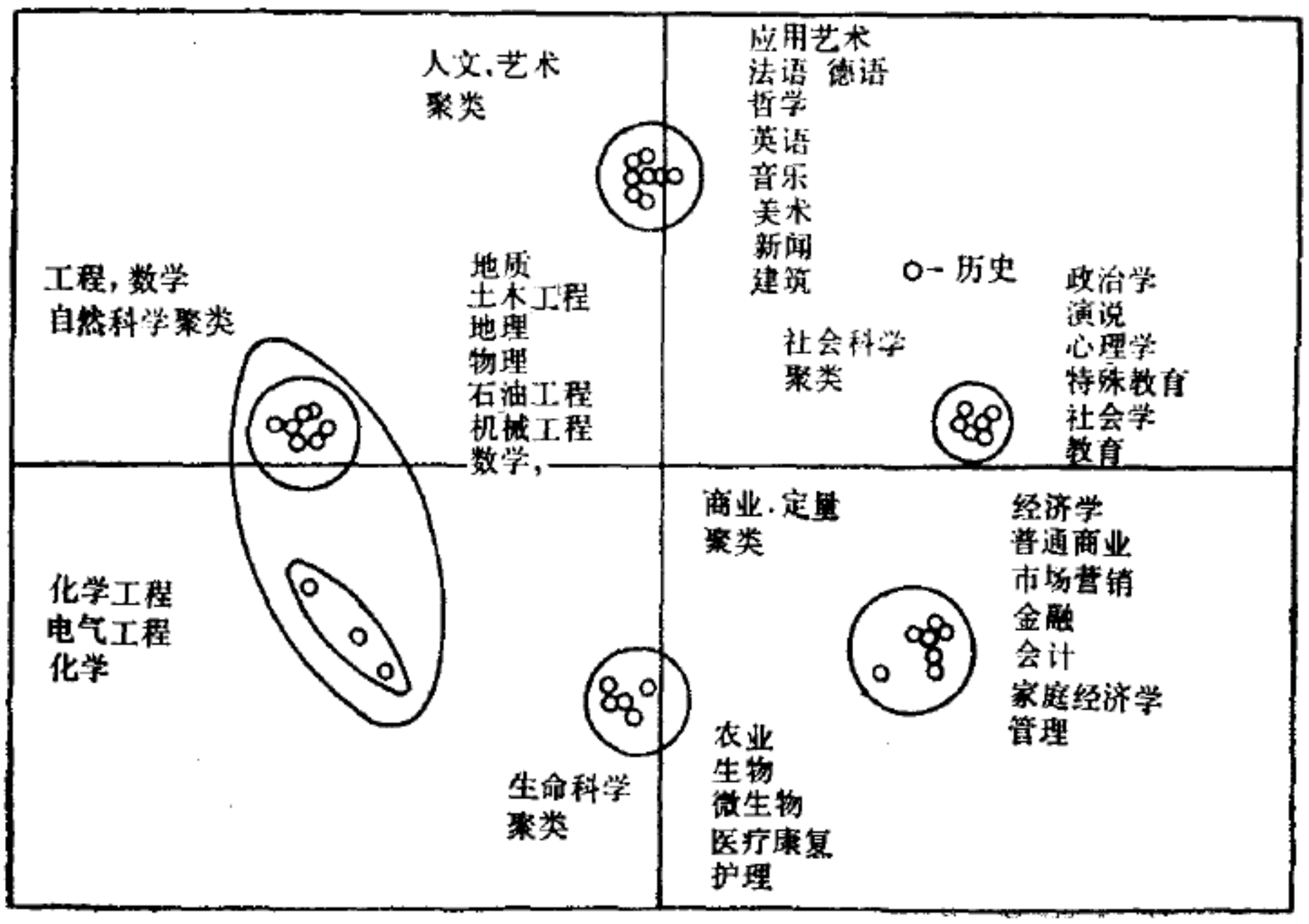


图 I .5.17　植物学期刊的树状图

以获得n维构形的二维图像。聚类法的独立应用可以产生自然分群。一些计算机程序不是显示树状图，而是以属于相同聚类的点为基础绘出二维范恩（Venn）图形，由此产生出数据的最优表示（参见图Ⅰ.5.5）。

基特莱尔（Keteleer，1986）、斯莫尔（Small，1986）、宫本和中山（1983）以及麦克格拉斯（1986）应用这种组合法研究了下列问题：校园中各个分图书馆的位置应当使每一个系都尽可能最靠近有关的分馆（即该分馆拥有最多的有关该系研究领域的图书）。当然，一个重要的约束条件是用最少的图书馆来满足这些条件。根据对37个专业学科图书的流通数据进行的多维标度分析和聚类分析结果，得出了5个有意义的聚类（见图Ⅰ.5.18）。



图Ⅰ.5.18 图书馆聚类

第Ⅱ编　运筹学与图书馆管理

概　述

图书馆的管理者和图书馆网络的规划者都必须在一定程度上了解图书馆读者的行为：他们使用某种特定参考手册的频度如何？图书馆藏书的借阅量是多少？图书应该怎样布局才能更接近读者？每年预算中的多少应该用来购买期刊？一般说来，图书馆的管理者或是图书馆网络的规划者应该能够预测不久的将来可能会发生什么情况，并将这种预测作为其决策的基础。基于这一理由，我们有必要运用统计推理、概率论以及运筹学方法来帮助决策。

图书馆运筹活动的目的可以概括如下：

（i）　帮助图书馆管理者和操作者更好地作出规划和决策，以便依靠读者来最大限度地开发利用馆藏资源。

（ii）　为检验和评估图书馆的运转情况提供依据，为有效改进图书馆工作的指导方针提供逻辑基础。

Ⅱ.1 规划问题

规划问题通常涉及有限的资源在众多的产品或活动中的最优分配。这些有限的资源可以是材料、人员、投资或是在大型、昂贵的机器设备上的处理时间。最优分配可以是一种使效益（如利润）取得最大值的分配，也可以是一种使费用取得最小值的分配。术语“线性规划”定义的是一类特殊的规划问题，在这里，选择决策变量“最佳”值的判据可以描述成一个由这些变量所构成的线性函数，控制这个过程的操作规则可以表示为一组线性方程或是一组线性不等式。

Ⅱ.1.1 双变量线性规划问题的图解

关于这个问题，我们要先从一个简单的例子着手。一家印刷装订公司要提前一天作生产计划，他们主要有两种产品需要考虑：一种是硬皮精装本（称为产品A），另一种是纸皮平装本（称为产品B）。在生产过程中，共要进行三项基本操作（这里我们不作详细说明，只是将这三项基本操作称为操作Ⅰ、操作Ⅱ和操作Ⅲ）。

表Ⅱ.1.1 加工时间（分）、生产能力和利润

	A	B	最大生产能力（每天）
操作Ⅰ	1	—	440
操作Ⅱ	1	1	1000
操作Ⅲ	1.2	0.5	620
利润（美元/件）	5	4	

产品A和产品B的加工时间见表Ⅱ.1.1。在这份表上还给出了最大生产能力和利润。这家印刷装订公司生产的产品每种应该生产

多少才能使公司的利润最多呢？请注意，为了使问题简化起见，我们只考虑供应问题而不考虑需求。

为了解这个问题，我们设

x = 1天内生产的硬皮精装本（产品 A ）的总数；

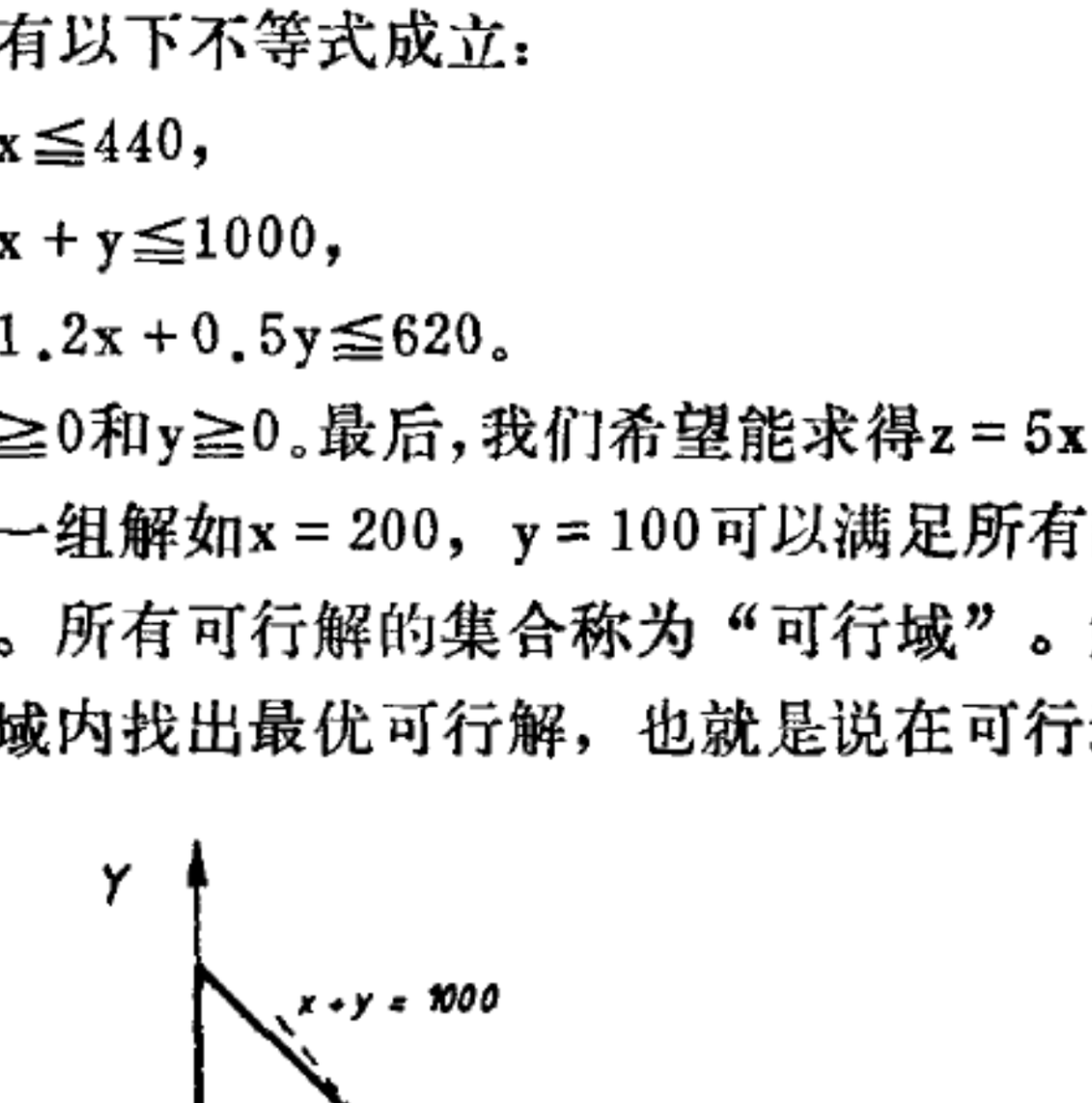
y = 1天内生产的纸皮平装本（产品 B ）的总数。

这样，必定有以下不等式成立：

$$\begin{cases} x \leqq 440, \\ x + y \leqq 1000, \\ 1.2x + 0.5y \leqq 620。 \end{cases}$$

并且还有：$x \geqq 0$和$y \geqq 0$。最后，我们希望能求得$z = 5x + 4y$的最大值。

方程的一组解如$x = 200$， $y = 100$可以满足所有的约束，被称为“可行解”。所有可行解的集合称为“可行域”。解线性问题就意味着在可行域内找出最优可行解，也就是说在可行域内要找到一个

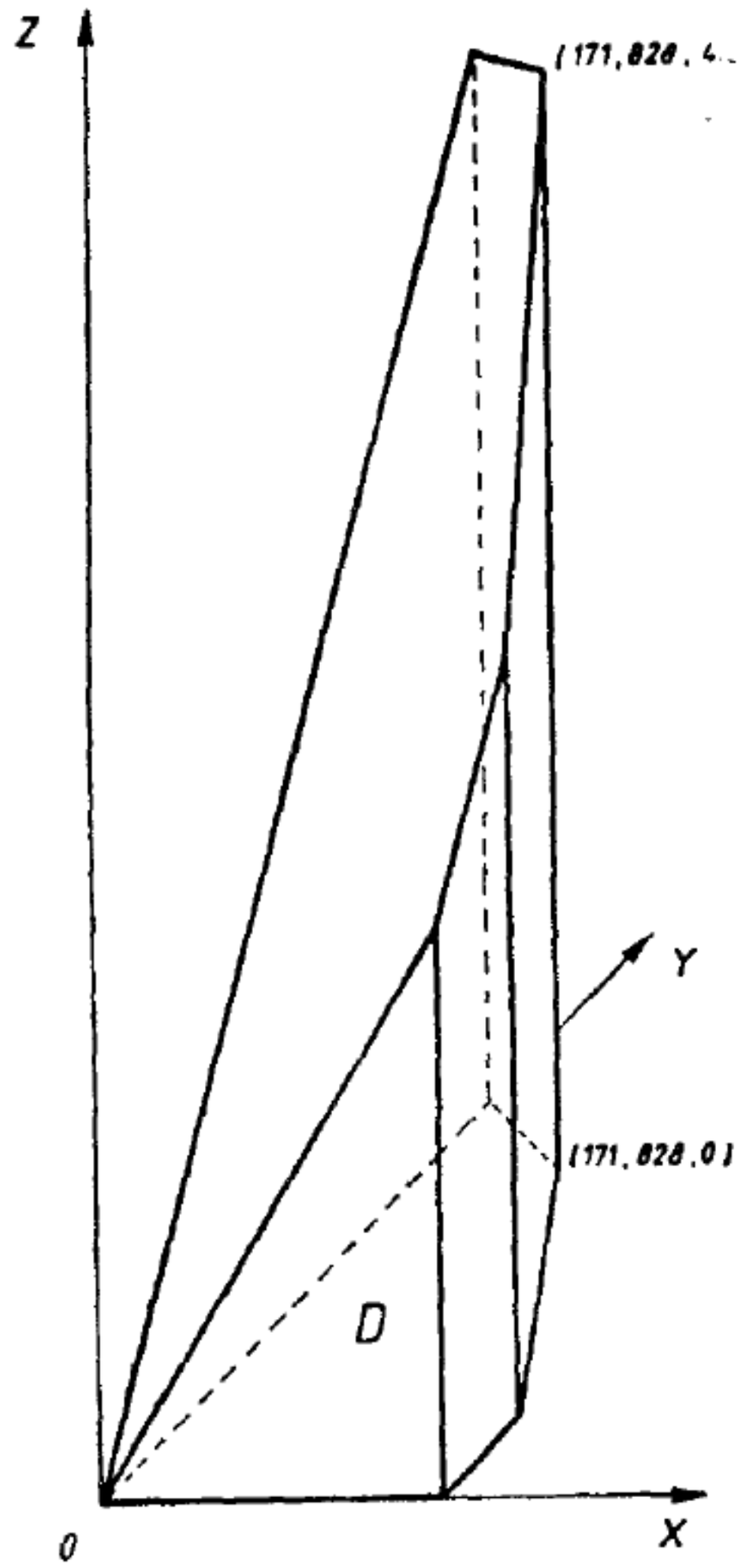


图Ⅱ.1.1　表Ⅱ.1.1的约束和可行域

点，使Z 取得最大值。

为了在二维图形上表示可行域，要将所有约束都绘在图上，满足这些约束的所有x和y的值都在图上显示（见图Ⅱ.1.1）。非负约束的意思是：双变量的所有可行值都将处于第一象限。

可行域可以由凸集 D 确定。很显然，在可行域内将有无数个可行点。我们的目的是要找出能使目标函数Z = f(x，y) = 5x + 4y取得最大值的可行点。



图Ⅱ.1.2　表Ⅱ.1.1中所列问题的可行域D和目标函数（Z）的值

如果C是一个常数，则函数5x + 4y = C代表一条直线。若改变C的值，可使整条直线转变成与其平行的另一条直线。为了找到最优解，我们为C取一个方便的值并画出目标函数线，使它通过可行域中的一个或数个点。在本例中我们取C = 0。当这条直线距离原点越来越远的时候，C的值将越来越大。唯一能限制这种增长的条件是：直线5x + 4y = C必须至少包括D中的一个点。照这样继续进行下去，我们发现坐标为（171.43，828.57）的极点作为最优可行点，可使C取得最大值4171.43。因此我们已经找到了最佳生产计划，即生产171件(数字当然要进行修约）产品A（硬皮精装本）和828件产品B（纸皮平装本），这样可得净利润4167美元。这个例子同时还表明：最优值总是在极点上获得的。

图Ⅱ.1.2表示的是三维空间上的相同例子，目标函数由Z轴表示。

与上面所提到的例子相似，最小值问题也可以在二维坐标系（双变量）上得到解决。通过研究－Z（而不是Z），最小值问题也可以简化成最大值问题。在没有太多约束的情况下，可以应用有3个变量的类似方法。但是对于实际生活中的问题来说，由于包含了更多的变量，因此需采用其它技术。这些将在下一节中讨论。

Ⅱ.1.2 线性规划问题和简化方法的形式说明

线性规划问题具有以下特征：

（1）有一个由m + n个等式或不等式组成的线性系统，在这个系统中，最后面的n是对n个变量的符号的约束。

（2）必须对线性函数（称为“目标函数”）进行优化（即最大化或最小化）。

请记住：如果对每一λ，$\mu \in \mathbb{R}$，以及对每一个

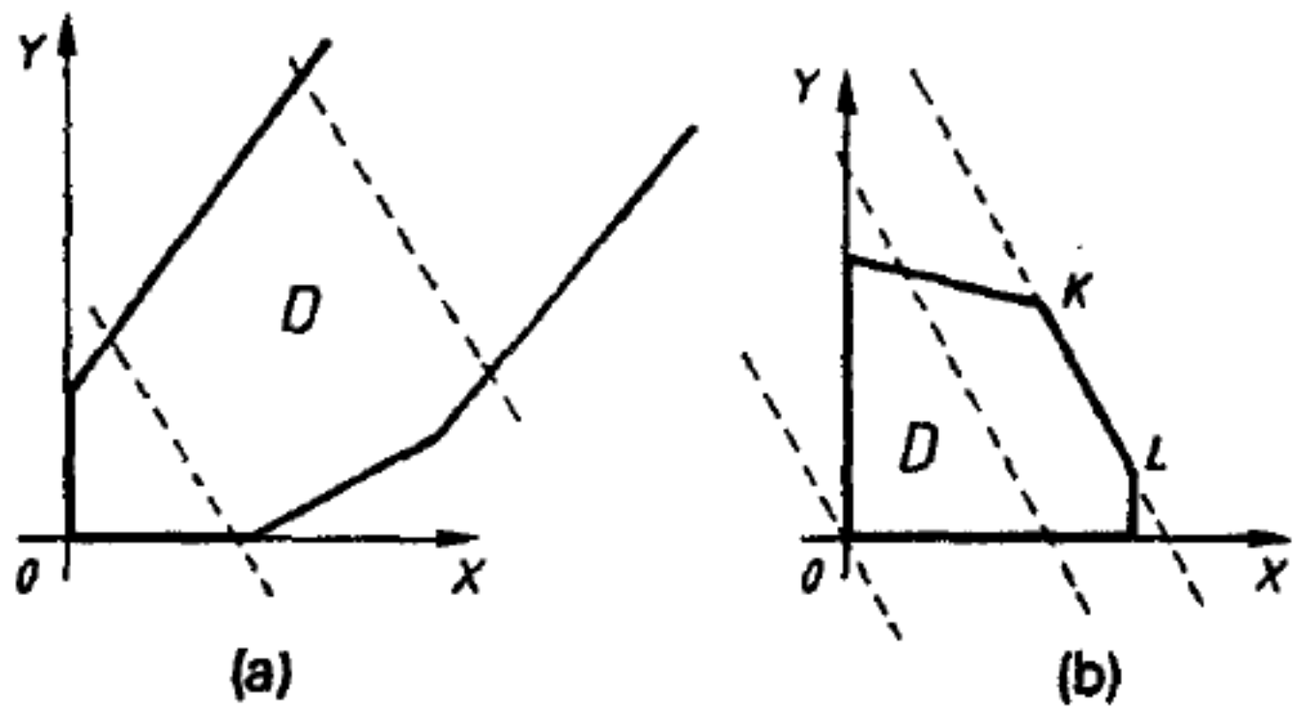
$$x,y \in X : f(\lambda x + \mu y) = \lambda f(x) + \mu f(y) \quad . \qquad [II.1.1]$$

则从一个实向量空间X到另一个实向量空间Y的函数f是线性的。如

果向量空间Y是R，我们就说这个线性函数（一次函数）是线性泛函。象 $f_1 : R^3 \to R : (x_1,x_2,x_3) \to ax_1 + bx_2 + cx_3$ $(a,b,c \in R)$ 这样的函数是线性泛函，而象

$f_2 : R^3 \to R : (x_1,x_2,x_3) \to ax_1 + bx_2 + cx_3 + d$ $(d \in R)$ 这样的函数则不是线性泛函。不过在这里可以很容易地看到，f_2的极值只不过是在f_1的极值上加了一个d而已。

各分量满足所有条件的向量$X \in R^3$被称为“可行解”。所谓“最优解”就是可使目标函数最优化的一组可行解。最优解并不一定总是存在的，而且即使它存在的话，也不见得总是唯一的。在图Ⅱ.1.3a中我们可以看到一个无限凸域D的例子，在这个凸域中，找不到目标函数的最大值（当然存在着最小解）；在图Ⅱ.1.3b所示的例子中，存在着无穷多个最优值，也就是说，线段KL上的每一个点都是最优值。



图Ⅱ.1.3a 目标函数（虚线）无最大值的可行域　　图Ⅱ.1.3b 目标函数有无穷多最大值的可行域

从形式上讲，标准原始最大值问题有如下的特征：找出$w = c_1x_1 + c_2x_2 + \cdots + c_nx_n$的最大值，并给出以下约束条件：

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \leq b_1 ; \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \leq b_2 ; \\ \qquad\qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \leq b_m , \end{cases} \qquad [\text{II}.1.2]$$

式中

$$x_1,\ x_2,\ \cdots,\ x_n \geqq 0$$

若用矩阵符号表示，以上问题变为：

求出$w = C^t X$的最大值，并且要同时满足

$$AX \leqq B \text{ 和 } X \geqq 0 \qquad [\text{II}.1.3]$$

这里

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

$$C = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix};\quad B = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix};\quad X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix};\quad 0 = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^n$$

t表示换位。并且，如果对每一个$i = 1, \cdots, k$有：$x_i \geqq y_i$，则我们说向量 $X \in \mathbb{R}^k$大于或等于向量$Y \in \mathbb{R}^k$。这个向量X常被称为“决策向量”，B常被称为“要求向量”，C称为“利润(或成本)向量”。

标准对偶最小值问题与每一个标准原始最大值问题相关联：

找出$v = B^t Y$的最小值，并且

$$A^t Y \geqq C,\ Y \geqq 0 \qquad [\text{II}.1.4]$$

这里$Y = (y_1, y_2, \cdots, y_m)^t$。

我们可以证明，如果标准最大值问题有最优解，那么它的对偶

最小值问题也有最优解，反之亦然。而且，最大值问题的目标函数的最大值等于其对偶问题的最小值。

当我们面临的问题是

使$w = C^t X$

最大化，并要满足条件：$AX = B$和$X \geqq 0$ [Ⅱ.1.5]

我们称此问题为典型原始问题。通过使用所谓松弛变量z_1，所有标准问题都可以转换成为典型问题。具体做法如下：

$$\sum_j a_{ij}x_j \leq b_i \text{ 转变成 } \left(\sum_j a_{ij}x_j\right) + z_i = b_i;\ z_i \geq 0 \ . \qquad [II.1.6]$$

因此，在实际应用中，我们只需要考虑典型问题。这里我们再一次强调，任何最小化问题都可以简化为最大化问题，只需用 - w 来代替w即可。

一种称为单纯形法的迭代方法在这里可以作为一种实用解法。这种算法的效率和通用性可以在很大程度上满足线性规划的重要作用的要求。单纯形法的重要价值在于它快速，应用广泛，并且可以回答有关对输入数据变量的解的灵敏度等重要问题。运用单纯形法，我们可以增加一定的约束条件，并且再次求解这个问题，以此来检验这些约束的影响。例如，用单纯形法可以快速计算出提供非盈利服务的费用，其目的是为了保持与用户的良好关系。

单纯形法是一个逐步计算的过程，包括两个步骤：

第一步：找出$AX = B$，$X \geqq 0$的可行解X_0，或者证明这样的解不存在。

第二步：找出最优解。

第二步可以通过在可行域D中建立一组顶点的有限序列$X_0, X_1, X_2, \cdots, X_s$用迭代方法解决，使$X_1$和$X_{1+1}$相邻接并使$w = C^t X$递增（最大值问题）。这种算法直到求出最优解或直到证明这样的解不存在为止。

我们注意到，最优解总是存在于可行域中的顶点上，这一事实

使得初始无限的问题（可行域中的注意一点都可能是最优点）简化成了大而有限的问题（有限数量的线性约束在空间产生了一个区域，在这个区域中只有有限数量的顶点）。

Ⅱ.1.3 整数规划

事实上，许多线性规划问题都要求部分或全部变量为整数解。例如，我们所出借的图书不可能是分数。我们在第Ⅱ.1.1节中所提到的图解问题仅仅产生了图书数为整数的有用结果。而术语“整数规划”所涉及的却是部分或全部决策变量都限制为整数的一类线性规划问题。但是，整数规划问题的解通常是比较困难、费时和费钱的。因此，实际的作法是将所有变量都处理成连续变量，并且用单纯形法来解此相应的线性规划，然后将所得的解修约成最接近的整数，并且还要同时满足约束条件。这种方法通常可以成功地获得近似最优整数解，尤其是在整数变量值很大的情况下（参见第Ⅱ.1.1节）。

在有些情况下，尤其是当决策变量的取值只能是0和1时，修约往往得不到最优整数解。因此，需要用一些专门技术来直接确定最优整数解。在实践中，解决整数规划问题用得最广泛的方法是所谓的“分支定界算法”。大部分用来解整数规划的商业计算机编码都是以这种方法为基础的。这些整数规划问题并不总是可以在合理的时间里得到精确答案的，在这种情况下，我们就要采用试探法来求得近似最优解。在下面的几节中，我们将要讨论一些特殊的整数规划。

Ⅱ.1.4 运输和分配问题

Ⅱ.1.4.1 运输问题

“运输问题”通常涉及的是如何用最少的费用或是在最短的时间内，将一定数量的产品从几个产地分发到许多地方的问题。

假定某地共有m家书店存有图书，而当地有n家图书馆需要图书。设书店可以供应的图书数量为a_1，a_2，…，a_m，当地图书馆

对图书的需求量为b_1，b_2，…，b_n，将图书从书店i运送到图书馆j的单位运费为c_{ij}。如果某家书店不能向某家图书馆供书，则我们取相应的c_{ij}为$+\infty$。我们希望能够作出一种最优运输计划表可以使运输总费用为最低。

这类运输问题可以表述为一种线性规划：我们将x_{ij}定义为从书店i运到图书馆j的图书数量。因为i的值可以在1到m之间任意假定，j可以在1到n之间任意假定，所以决策变量的总数等于mn。这样的运输问题可以写成公式：

在满足条件

$$\sum_{j=1}^{n} x_{ij} \leqq a_i,\ i=1,\ \cdots,\ m \text{（书店的供货限制）}$$

$$\sum_{i=1}^{m} x_{ij} \geqq b_j,\ j=1,\ \cdots,\ n \text{（图书馆的需求），}$$

$$\forall i,j: x_{ij} \geqq 0 \text{（非负约束条件）} \qquad [\text{Ⅱ}.1.7]$$

的情况下，使

$$w=\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} \text{（运输总费用）}$$

最小化。

很显然，只有当且仅当书店的总供应量至少与图书馆的总需求量相等时，图书馆的需求才能完全得到满足，这意味着

$$\sum_{i=1}^{m} a_i \geqq \sum_{j=1}^{n} b_j \qquad [\text{Ⅱ}.1.8]$$

当总供给量与总需求量相等，即$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$时，这个问题就变成了标准运输问题。其形式为：

在满足条件

$$\sum_{j=1}^{n} x_{ij} = a_i,\quad i=1,\ldots,m.$$

$$\sum_{i=1}^{m} x_{ij} = b_j \ , \quad j = 1,\ldots,n$$

$$\forall i,j : x_{ij} \geq 0 \ . \qquad [\text{II}.1.9]$$

的情况下，使

$$w = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} \qquad \text{最小化}$$

任何供求不平衡的非标准运输问题都可以转化为标准运输问题，只需要使用一个假想的书店或图书馆就可以了。例如，我们考虑一个总供给量超过总需求量的问题。为了将这个问题转化为标准问题，我们将设立一个假想的图书馆来吸收书店的剩余供应量。假定将图书从任何一个书店运送到这个假想图书馆的单位运费为0，因为事实上这个假想的图书馆并不存在，并没有发生有形的货物传递。这样就可以得到以下标准运输问题：

在满足条件

$$\sum_{j=1}^{n+1} x_{ij} = a_i \ , \ i = 1,\ldots,m$$

$$\sum_{i=1}^{m} x_{ij} = b_j \ , \quad j = 1,\ldots,n+1$$

$$\forall i,j : x_{ij} \geq 0 \ . \qquad [\text{II}.1.10]$$

的情况下，使

$$w = \sum_{i=1}^{m}\sum_{j=1}^{n+1} c_{ij}x_{ij} \qquad \text{最小化}$$

这里$j = n+1$是假想的图书馆，其需求为

$b_{n+1} = \sum_{i=1}^{m} a_i - \sum_{j=1}^{n} b_j$。对于每一个$i = 1,\cdots, m$有$C_{i,n+1} = 0$。

对于总需求量超过总供给量的情况，我们可以设立一个假想的书店，其供应量$a_{m+1}=\sum_{i=1}^{n}b_j-\sum_{i=1}^{m}a_i$。这里$x_{m+1,j}$将代表在图书馆j的图书短缺量。

一般说来，运输问题都可以用线性规划的单纯形法来解，但运输问题的特殊结构产生了特殊的解法，用这种解法所需要的计算机机时数较少。

Ⅱ.1.4.2 相关问题

我们在这里要介绍一种不属于线性规划问题的相关运输问题。假定现有m个发行中心（即前述问题中的批发商）和n个目的地（例如当地的图书馆）。设t_{ij}代表从第i个发行中心到第j个图书馆的运输时间。这个问题是要求出

$$T=\max t_{ij} \qquad [Ⅱ.1.11]$$

的最小值。上式中当x_{ij}（从发行中心i到图书馆j的运输点）为严格的正数时（i，j）有最大值。

这类问题往往出现在主要考虑时间的情况下，例如在组织馆际互借回路的时候。

Ⅱ.1.4.3 分配问题

最后，我们要讨论“分配问题”。假如一个大型综合性大学的图书馆馆长要将n位图书馆员分配给n个分馆，c_{ij}代表将馆员j分配给分馆i的效果因子。一个具有数学和图书馆学情报学学位的人可能最适合于在数学系分馆工作（效果因子高），在工程或物理系分馆工作也比较适合（效果因子中等），可能在远东问题研究分馆工作结果会很糟（如果他（她）不懂中文或日文的话）。每个分馆只能分配去一个人。这里的问题是要使馆员分配的总数量为最高。

我们定义 $x_{ij}\begin{cases}=1 & \text{如果馆员j被分配到分馆i；}\\ =0 & \text{其它情况。}\end{cases}$

因为每一个分馆只能分配到1名馆员，因此有

$$\sum_{i=1}^{n} x_{ij} = 1, \quad i = 1, \cdots n, \qquad [\text{Ⅱ}.1.12]$$

同样，每位馆员只能分配去一个分馆：

$$\sum_{i=1}^{n} x_{ij} = 1, \quad j = 1, \cdots, n. \qquad [\text{Ⅱ}.1.13]$$

我们的目标是要使

$$w = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}. \qquad [\text{Ⅱ}.1.14]$$

最大化。

从这个式子中我们可以看出，分配问题等效于有n个 发行中心和n个当地图书馆的标准运输问题，供应量$a_i = 1$ （$i = 1, \cdots, n$），需求量b_j等于1 （$j = 1, \cdots, n$）。

Ⅱ.1.5 举例

Ⅱ.1.5.1 装订问题

一个馆员会面临以下问题：她每年至少有500套期刊需要装订，装订的最高预算经费是7000美元/年。她可以在两种装订方 法 中进行选择：

1）便宜的装订方法，每件的装订费需10美元，平均使用寿命为5年。

2）较贵的装订方法，每件的装订费需要20美元，平均使用寿命为15年。

但是，合同迫使她每年至少用便宜的装订方法装订200套期刊，用较贵的装订方法装订100套期刊。那么，这位馆员到底应该 用 这两种装订方法各装订多少套期刊，才能使期刊得到的保护时间最长呢？

我们将这个问题表示成下面的数学形式：令x表示期刊低 价 装订的数量，y表示高价装订的期刊数量。由于高价装订法对 期 刊的保护时间是低价装订法的3倍，因此我们要使

$$w = 5x + 15y$$

取得最大值，并且同时还要满足下列约束条件：

$$x + y \geqq 500,$$

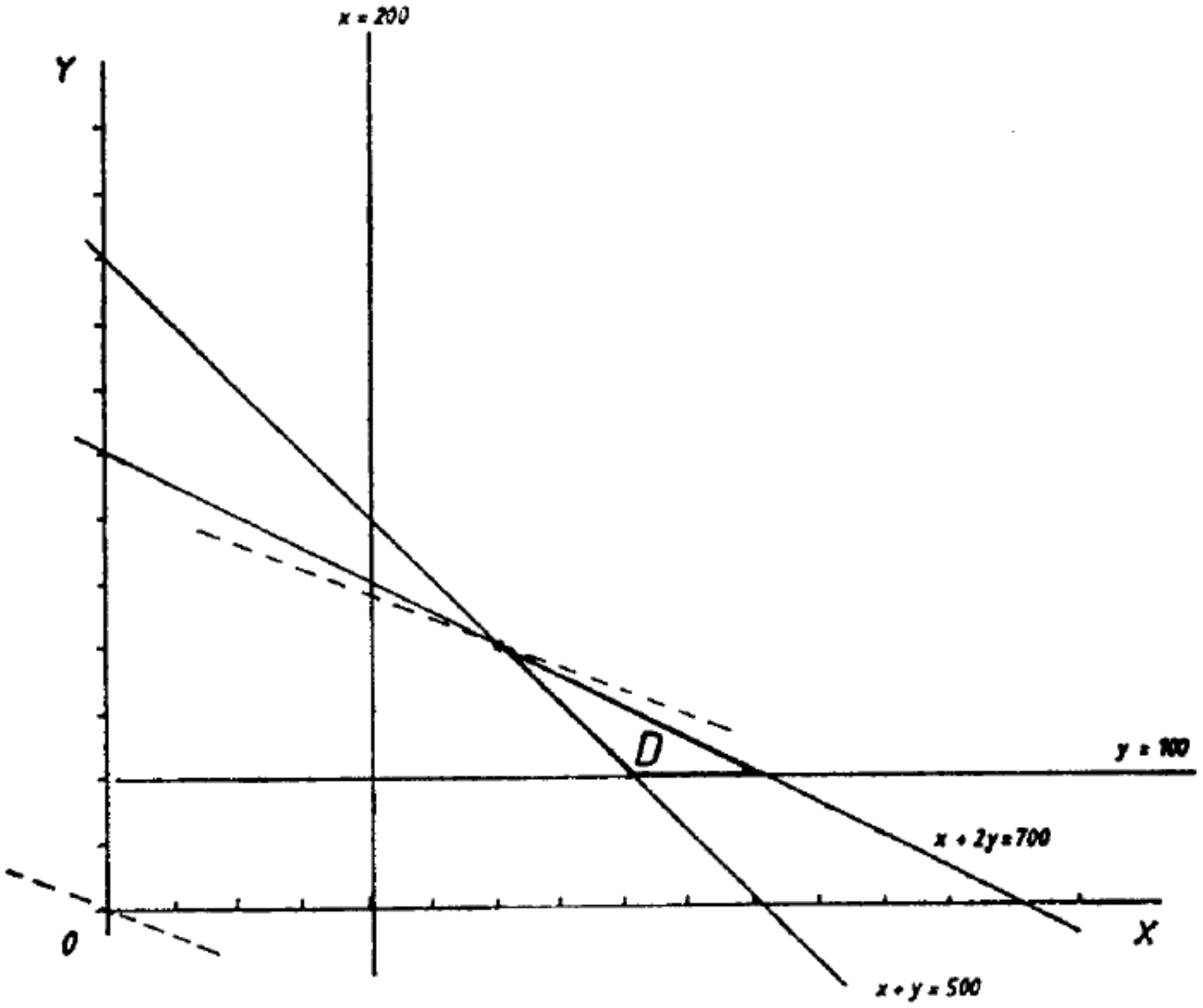
$$10x + 20y \leqq 7000,$$

$$x \geqq 200,$$

$$y \geqq 100.$$

请注意，这里自动规定$x \geqq 0$，$y \geqq 0$。这个问题的最优解（见图Ⅱ.1.4）为：300套期刊用低价但也是低质量的装订法装订，200套期刊用高价装订法装订。这样的装订比例将用完最高预算限额7000美元，并将使总的期刊保护时间长达4500年。

Ⅱ.1.5.2　大学中各系书刊采集经费的分配



图Ⅱ.1.4　第Ⅱ.1.5.1　小节中线性问题的图解法

在这一小节我们将讨论经费的分配问题。在建立目标函数时要考虑的主要因素是各系的重要性权重。我们将使用公式

$$C_1 = \left(\frac{S_1 + T_1}{2} \right) O_1, \qquad [\text{II}.1.15]$$

式中C_1表示系i的重要性，S_1是衡量社会对该系工作重视程度的因子（例如，经济系的S_1因子要大于人类学系），T_1是衡量大学对该系工作重视程度的因子（例如天主教大学会给神学系以高的T_1因子；这样的大学也希望根据对该系自然科学论文的某种引文计量，采用更加客观的因子）。O_1因子用来衡量某系规模的重要程度，这个因子的数值取决于教师数量和学生数量两方面。

下面我们将给出这个问题的公式。

设需要分配经费的系的数量为n，可用来购买书、刊的总经费数量为M，分配给系i的经费用X_1表示，系i的重要程度用C_1表示，$i = 1, \cdots, n$。我们进一步假定，分配给系i的经费有一个正值的下限L_1和上限U_1。当然，$L_1 \leqq U_1$。

到此，这个问题的公式可以表示为：

在满足条件

$$X_1 \geqq L_1 \geqq 0,$$

$$X_1 \leqq U_1$$

$$\sum_{1=1}^{n} X_1 \leqq M \qquad [\text{II}.1.16]$$

的情况下，使

$$Z = \sum_{1=1} C_1 X_1 \text{达到最大化}$$

另外，在实际的分配问题中可能还有其它的约束条件，称为“组合约束”。例如，分配给系i和系j的总经费不得超过$U_{1\,j}$，分配给系k、系l和系m的总经费不得低于$L_{k\,l\,m}$。这些附加约束条件可以表述为。

$$X_1 + X_j \leqq U_{1\,j}$$

和

$$X_k + X_l + X_m \geqq L_{k\,l\,m} \qquad [\text{II}.1.17]$$

在第Ⅱ.1.2节所述单纯形法的基础上略加变化，即可求得问题的解。实际上，在第Ⅱ.1.2节中我们只考虑了所有不等式都是同一类型的情况，而在这里，不等式类型既包括≤，又包括≥，实际上这并不成为解题的障碍，有效的方法是给不等式的两边同时乘以（－1），使它成为反向不等式。

Ⅱ.2 最短路算法

在这一章中，我们将讨论几种运筹学中的图论法。现实中有许多运输问题，无论是求时间的最小值，还是求距离的最小值，都可以用图论来解决。我们将先从图论的基础知识入手。

Ⅱ.2.1 图论基础知识

在图论中，“图”（无向图）G由顶点(结点)集合V和边(弧)集合E组成，并且每条边e∈E与一无序顶点对相连。如果一条边连接唯一的顶点对i和j，则记为e=(i，j)或e=(j，i)。在这部分内容里，(i，j)代表的是一条边，而不是一个有序点对。

有向图G由顶点集合V和边集合E所组成，并且每条边e∈E与一个有序顶点对相连。

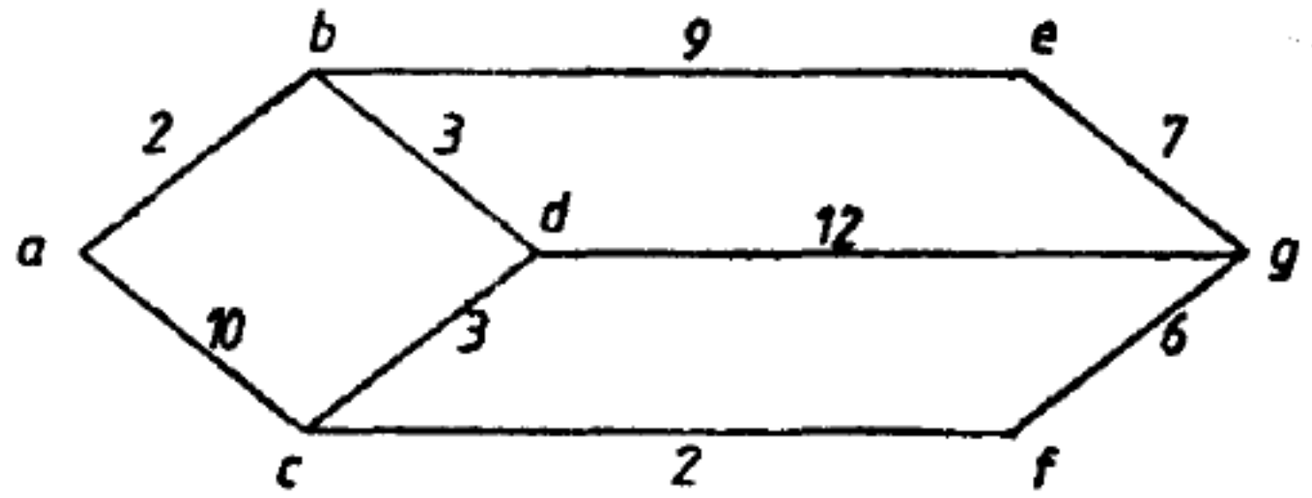
图（无向图或有向图）中的边e=（i，j）称为与i点和j点相关联。顶点i和j称为与e相关联，并且i和j是相邻顶点。如果G是有顶点集合V和边集合E的图，则记作G=（V，E）。在本书中，集合E和集合V始终假定是有限的。

从顶点i到顶点j的路就是一个从i点到j点的边的序列。这条路的长度等于相异的边数减去1。因此，如果i点和j点是相邻的，那么就有长度为2－1=1的路连接这两个点。回路（或称圈）是从i点出发又回到i点的路。对给定的任一不同点对i和j，如果有一条从i点到j点的路，则称图G是连通的。

赋权图是指内含与边相关联的数字的图。与边（i，j）相关联的值w（i，j）称为（i，j）的“权值”。例如，如果我们将图书馆表示为顶点，各馆之间的通路表示为边，并且如果我们确定了每条路的长度，就得到了一个赋权图（见图Ⅱ.2.1）。权值常常用来表示距离、时间或费用。

有一类很重要的图称为树。一棵“树”就是一个不含任何回路

的连通图。典型的树有一个特殊顶点，称为“树根”。一棵树上的任意两个顶点之间只有一条路。第Ⅰ.5.4节中的树状图就是有树根的例子，在那个图上，每个末端结点代表一个目标，非末端结点代表一个非单元素聚类，而树根则代表整个目标的集合。



图Ⅱ.2.1 赋权（无向）图

图在研究一个图书馆内空间关系时是很有用的工具。例如可以用来研究如图Ⅱ.2.2a和b所示的图书馆平面布置图（这个图是以现有公共图书馆的现实为基础虚构的）。空间关系的无向图见图Ⅱ.2.3。结点代表位置，如果可以直接从一个位置到达另一个位置，则这两个位置在图上是连通的。

图Ⅱ.2.2中所用符号的说明：

E：大楼的主入口处

L：入口处大厅

W：女厕所

M：男厕所

LE：书库入口处

CI：咨询处

CA：成人流通服务台及流通控制处

EC：供等候和休息用的安乐桌椅

DC：唱片库目录

CATA：成人图书目录

DL：唱片库及流通服务台

AF：书库：成人小说类

EM：紧急出口

S：楼梯

PM：复印机

CATJ：青少年图书目录

DJ：青少年读者可借阅的文献

CDJ：青少年读者流通服务台

J：书库：青少年读者

ANFa,b：书库：成人阅读的非小说类图书

P：现期期刊

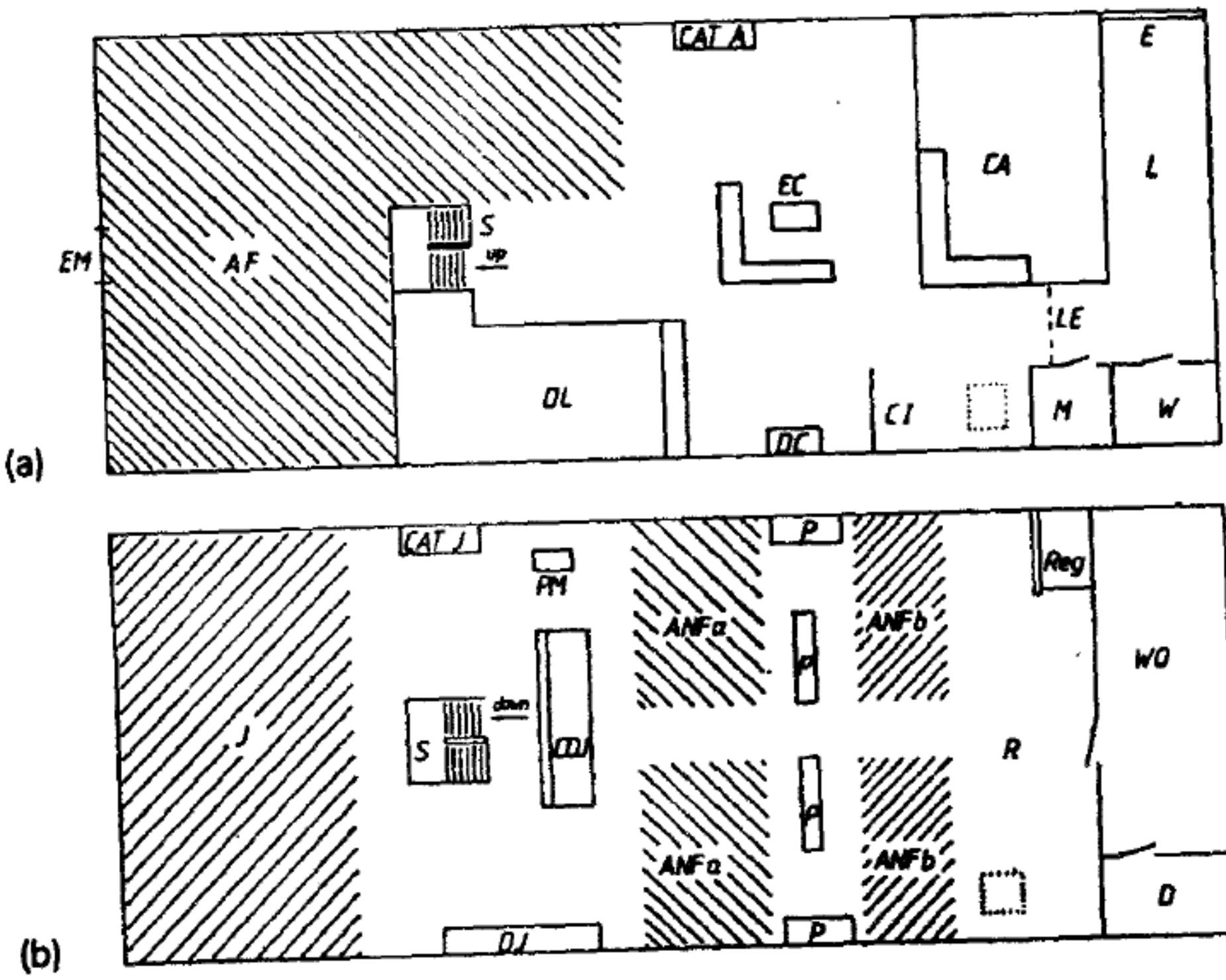
R：阅览室

REG：登记处

WO：馆内人员工作区

D：馆长办公室

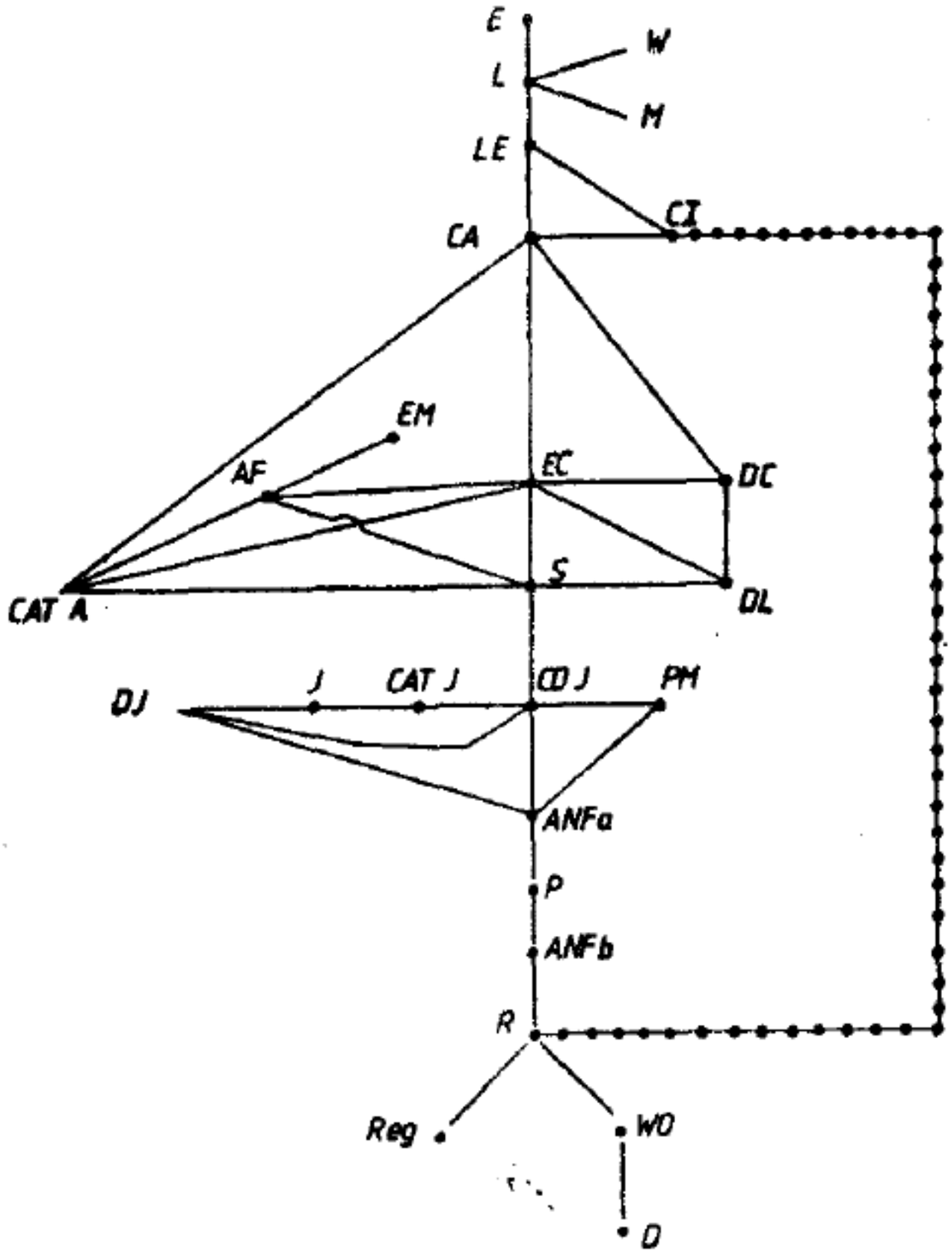
从这个图上我们能看到什么呢？首先，我们注意到总体设计是相当直线性的（一个结点连着一个结点），而不是星形的（有一个中心位置）或组合星形的。青少年读者流通服务台占据中间位置，成人流通服务台前的等候休息室中的安乐椅也是处于中间位置。



图Ⅱ.2.2　公共图书馆平面布置图

此外，馆长室不在图书馆读者易于接近的位置，而那些第一次来到图书馆并想要登记注册的读者还必须穿过所有的部门。更重要的是，在工作区内的编目人员和其他工作人员距离最近的出口至少有8个结点（路的长度为7），距离厕所有11个结点（路的长度为10），而且为了上厕所，他们还必须从图书馆外边走。更为不方便的是，成人目录与非小说类图书区相距太远，而且毫无道理地将非小说类图书区分割为两部分。

鉴于上述情况，应该要求对原方案全部重新设计，以改善这种不合理的布置。关于图书馆工作人员问题的局部解决方案是从咨询处（很容易搬开）到阅览室建造第二个楼梯，这已在图Ⅱ.2.2和图Ⅱ.2.3中用虚线标出。这样处理后，阅览室和登记处也就变得更加容易出入了。



图Ⅱ.2.3　根据图Ⅱ.2.2得出的图

Ⅱ.2.2 迪克斯特拉最短路算法

在这一节我们要研究的问题是在赋权图中找出两个已知顶点间的最小权路，这样的路称为“最短路”。“迪克斯特拉(Dijkstra)最短路算法”可以有效地解决这个问题。这也意味着我们将可以用同样的算法解“最短时间运输问题”或“最低费用运输问题”以及许多其它类似问题。在这一节中，G代表的是连通赋权图，此外我们还要假定权值是正数，我们希望能够找出从固定顶点a到固定顶点z的最短路。然后我们将介绍寻找从图上的一个固定顶点a到任一其它顶点的最短路的方法。

迪克斯特拉算法包括给顶点分配两个标号，记为L(x)和P(x)（这里的x代表一个顶点）。在任意给定的时间内，一些顶点具有临时标号，而另一些顶点具有固定标号。我们随后将证明，如果L(x)是顶点x的固定标号，则L(x)就是从顶点a到顶点x的最短路长度。在开始的时候，只有顶点a 具有固定L标号，算法的每一步迭代都可以使一对标号从临时变为固定。当z接受了固定标号时，算法结束。到这一步时，L(z)得出了从a到z的最短路长度。P标号用来寻找路本身。

现在我们来说明这种迪克斯特拉最短路算法。

第1步：初始化。

令L(a) = 0；对与a相邻的那些顶点，令L(x) = w(a, x)，P(x) = a。对所有其它顶点，令L(x) = ∞，并用T表示除a以外的所有顶点的集合。

第2步：检查算法是否已经找到了最短路。如果z∉T，则算法结束，L(z)就是从a到z的最短路长度。

第3步：考虑下一个顶点。

选择有最小L值的v∈T，如果有联系可选择任一个，但只能选择一个。令集合T = T∖{v}。

第4步：修改标号。

对每一个与v相邻的顶点x∈T

如果L(x)已经改变，则令

$L(x) := \min\{L(x), L(v) + w(v,x)\}$

$P(x) := v$

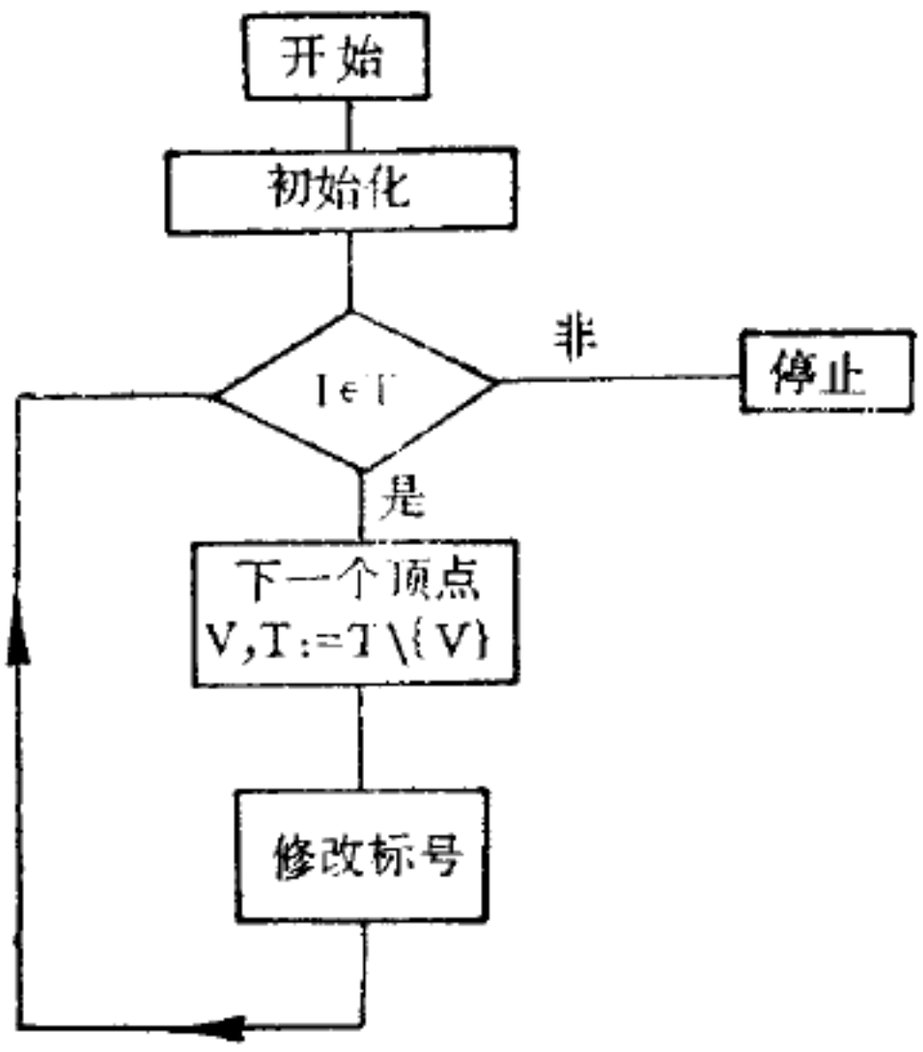
回到第2步。

如果想要求得从a到G的每一个顶点的最短路，则可以用第2′步来代替第2步。

第2′步：检查是否已经找到了所有的最短距离。

如果$T = \phi$：停止。对每一个$x \in G$，$L(x)$就是从a到x的最短路长度。

图Ⅱ.2.4表示了这种算法的流程图。



图Ⅱ.2.4　迪克斯特拉算法的流程图

作为例题，我们将把这种算法应用于图Ⅱ.2.1，以求出从a到g的最短路。每一步迭代都由一个状态图来显示。

初始状态

	T	L	P
a	-	0	
b	*	2	a
c	*	10	a
d	*	∞	
e	*	∞	
f	*	∞	
g	*	∞	

*代表属于T的元素

—代表不属于T的元素

不属于T的元素具有固定标号

第2状态

a	-	0		
b	-	2		a
c	*	10		a
d	*	5		b
e	*	11		b
f	*	∞		
g	*	∞		

第3状态

a	-	0		
b	-	2		a
c	*	8		d
d	-	5		b
e	*	11		b
f	*	∞		
g	*	17		d

第4状态

a	-	0		
b	-	2		a
c	-	8		d
d	-	5		b
e	*	11		b
f	*	10		c
g	*	17		d

第5状态

a	-	0		
b	-	2		a
c	-	8		d
d	-	5		b
e	*	11		b
f	-	10		c
g	*	16		f

第6状态

a	-	0		
b	-	2		a
c	-	8		d
d	-	5		b
e	-	11		b
f	-	10		c
g	*	16		f

第7状态

a	-	0		
b	-	2		a
c	-	8		d
d	-	5		b
e	-	11		b
f	-	10		c
g	-	16		f

因为g已经得到了固定标号，因此这是最终状态。

从a到g的最短路长度为16。

这个最短路的倒排顺序是：g－P(g)＝f－P(f)＝c－P(c)＝d－P(d)＝b－P(b)＝a，正顺序是：a－b－d－c－f－g。

在这个特定的例子中，我们已经求出了从a开始的所有最短路（因为在最终状态中，T＝φ）。例如，从a到e的最短路长度是11，顺序是a－b－e。

Ⅱ.2.3 迪克斯特拉算法的应用

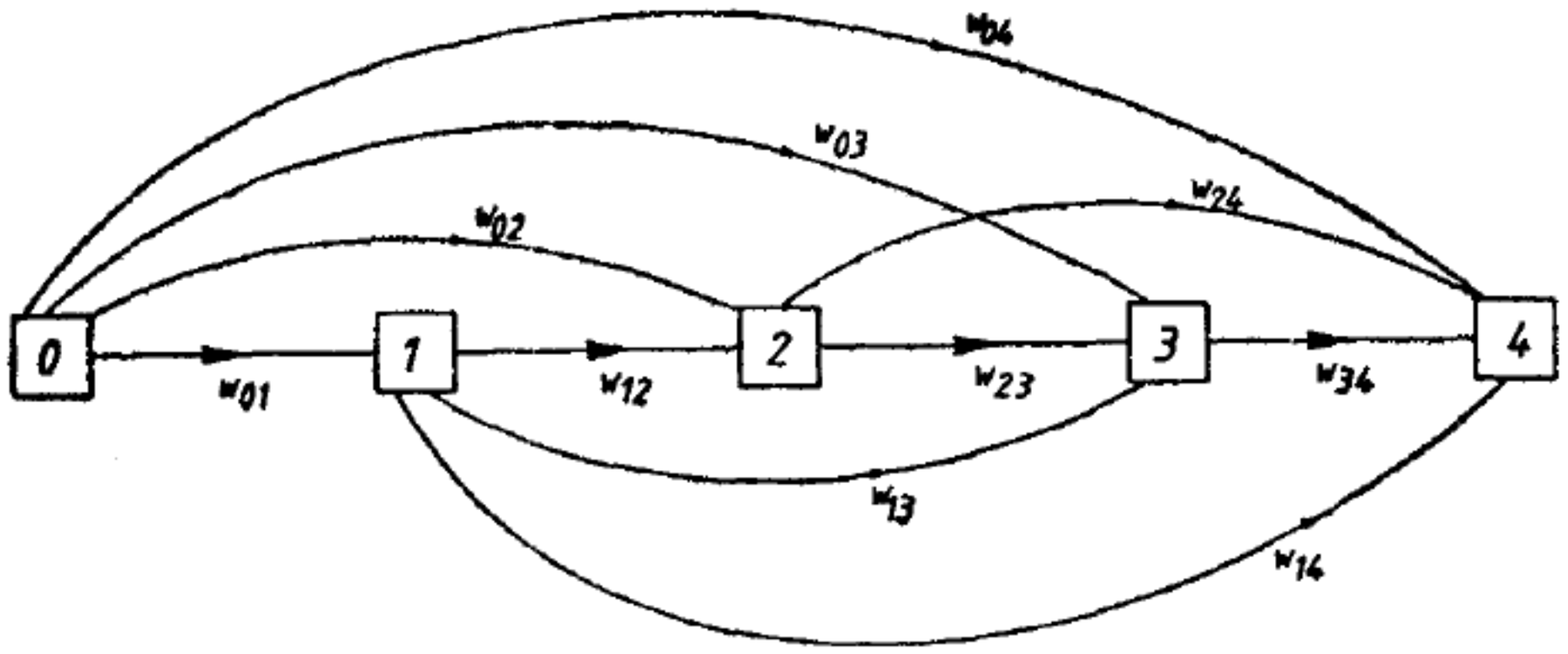
如前所述，迪克斯特拉算法的应用明显地包括了所有类型的运输问题。在本节中，我们将举两个不很明显的应用例子。

A．设备更新问题

大部分视听设备随着使用年限的增加，所需要的维修次数也越来越多。通过经常性地更新设备，可以节省相应的维修费用。但是，这种维修费用的节省是以增加每次更新设备时的基本建设投资为代价的。管理者所面临的最重要的问题之一，是要决定设备更新的时间，以便能够使总费用（包括基本建设投资、维修费以及运转费用）为最少。这个问题可以表示为有向图中的最短路问题，并且可以用迪克斯特拉算法予以解决。

假如现在有一个多功能中心计划购买设备，这台设备计划将在4年后更新。但是在这期间更新设备是否可以呢？设K_j代表在j年购买设备的价格，S_k代表使用了k年之后的折余值，设备在运转第k年间的维修和运转费用由C_k表示。由于随着设备的老化，其维修和运转的费用增加，我们假定对所有k＝1，2，3，4有$C_{k+1}>C_k$。为了将确定最佳更新方针的问题表示为最短路问题，要建立一个有向图，见图Ⅱ.2.5。

结点0和4分别代表规划周期的开始和结束，中间结点j(j＝1，2，3）代表j-1年的年终（或j年的开始），在这个时候才可能更新设备。只有当j>i时，才存在一条从结点i到结点j的连通有向



图Ⅱ.2.5　设备更新问题的有向图

弧，这就是说，由于已经在i年更新了设备，因此设备的下一次更新只能在以后几年中进行。

结点i和结点j之间的距离（权值）表达式为：

$$w_{ij}\begin{cases} = K_i - S_{j-i} + \sum_{k=1}^{j-i} c_k & \text{for } j > i \\ = \infty & \text{for } j \le i\ . \end{cases} \qquad [II.2.1]$$

权函数w等于购置设备的费用减去折余值再加上设备的维修和运转费用，而值为无穷大则表明从i到j没有弧连通。

在图中，从结点0到结点4的每一条路都代表着一种可能的设备更新方针。例如，路0-1-2-3-4相当于每年更新一次设备，其总费用为：

$$\left(\sum_{k=0}^{3} K_k\right) - 4S_1 + 4C_1. \qquad [\text{Ⅱ}.2.2]$$

另一种方针是将设备连续使用4年，这相当于路0-4。这种方针的费用为：

$$W_{04} = K_0 - S_4 + \sum_{k=1}^{4} C_k. \qquad [\text{Ⅱ}.2.3]$$

这样，确定从0到4的最短路就等于在寻找设备更新问题的最少费用方针。

这个问题怎样才能表述为线性规划问题呢？ 如果设备是在i年购买的，并且将在j年更新，我们可以取$x_{ij}=1$，对于所有其它情况，取$x_{ij}=0(i, j=0, 1, \cdots, n)$。这时的约束条件为：

$$\sum_{i=0}^{n-1} x_{ij} \leq 1 ; \quad \sum_{j=1}^{n} x_{ij} \leq 1 ,$$

$$\sum_{j=1}^{n} x_{0j} = 1 ; \quad \sum_{i=0}^{n-1} x_{in} = 1 . \qquad [\text{II}.2.4]$$

最后得到最小化的目标函数为 $\sum_{i,j} w_{ij}x_{ij}$。

B．图书馆中的图书密集存放

这里所讨论的是考虑图书尺寸的图书存放问题。假定馆藏中所有图书的高度和厚度都是已知的， 令图书按照其已知的 n 个高度 $H_1, H_2, \cdots H_n (H_1<H_2<\cdots<H_n)$ 以递增顺序排列。需要强调的是，任意一册高度为H_i的图书都可以放置在层高大于或等于 H_i 的书架上。因为每册图书的厚度是已知的，所以各个高度类别 i 所要求的长度是可以计算出来的，这里我们用L_i来表示。

如果将图书直立放置，所有的藏书只用一种层高的书架存放（考虑最高的图书所需要的层高），那么所需要的总书架面积就是图书的总长度与最高图书的高度之乘积。相反，如果将馆藏图书按照高度分成两个或多个组，则可以很容易看到，这种情况所需要的书架总面积要小于图书不分组的情况。

建造不同层高与长度的书架所需要的费用将按以下方法计算。对每一种书架的层高H_i，我们设：

K_i = 与书架面积无关的固定费用，

C_i = 单位面积的可变费用。

例如，假定馆藏图书放置在层高分别为H_m和 H_n（$H_m<H_n$）的两种不同的书架上，即将高度为等于或小于H_m的图书放置在层高为H_m的书架上，将另外的图书放置在层高为H_n的书架上。这样