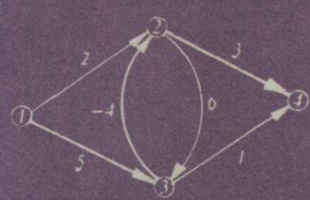
·运筹学小丛书·

动态规划简介

杨敦悌 陈斌 著



辽宁教育出版社

·运筹学小丛书·

动态规划简介

杨敦悌
陈　斌 著

辽宁教育出版社
一九八五年·沈阳

动态规划简介

杨蕴悌　陈　斌著

辽宁教育出版社出版　辽宁省新华书店发行
（沈阳市南京街6段1里2号）　沈阳新华印刷厂印刷

字数：78,000　开本：787×1092 1/32　印张：4
印数：1—1,500
1985年6月第1版　1985年6月第1次印刷

责任编辑：杨　力　责任校对：晓　舟
封面设计：周晓风

统一书号：7371·35　定价：0.63元

出版说明

运筹学是二十世纪四十年代开始形成的一门学科，是现代数学的一个重要组成部分。在科学技术迅速发展的今天，运筹学有着广泛的应用。

为了向广大读者普及运筹学知识，在中国运筹学会的关心和支持下，尤其是在徐利治教授的积极倡导和组织下，编辑出版了这套《运筹学小丛书》。

这套丛书用通俗的语音系统地介绍了运筹学中各个分支的基础知识和应用方法，其中包括规划论、对策论、排对论等方面的二十多个专题。在编写内容上，注重科学性、知识性和趣味性相结合，论述一般先从实例谈起，由浅入深，引出完整的数学理论。并且，大部分内容的引出方法都是初等的，因此，凡是具有高中以上文化程度的读者，都可以阅读。

引　言

动态规划是解决一类决策过程的最优化问题的一种有效方法，本书只限于讨论多阶段过程。所谓多阶段决策过程是指一类过程，由于它的特殊性，可以将过程（一般是根据时间或空间)分为若干个阶段或若干个子问题，而在每一阶段都需要作出决定，以便使整个过程取得最优的结果。在经济管理、工程技术、工业生产、军事以及其它领域中有许多问题都可归结为这类问题。这类问题属于规划的范畴，时间通常是个很重要的因素，这类模型称为动态模型，处理它的方法称为动态规划方法。但这种方法也可以处理一些与时间无关的静态模型，这时只要在静态模型中人为地引进“时间”因素，并把它作为多阶段的动态模型来考虑。下面举两个多阶段决策过程的例子。

例 1　从某发射站发射一枚火箭，目的是要在给定的时间内击中一个移动的目标，假设火箭的速度是能够追上目标的，问题是：如何根据目标的行动，每隔一个时间阶段来决定火箭行动的方向和速度，使之能在给定时间内击中目标。

例 2　某种有色金属的提纯装置是由若干个反应槽串联而成，前一个反应槽的输出是下一个反应槽的输入，在各反应槽内装有搅拌器和加热器，搅拌和加热时需要消耗一定的电

能，消耗的电能是反应槽的搅拌速度和反应温度的函数，而搅拌速度和反应温度又和该反应槽的提纯效率有关，问应如何控制各反应槽的生产条件使得产量为一定时，整个流程所消耗的电能为最小。

动态规划是从1951年起美国数学家贝尔曼(R.Bellman)等人所发展的，近年来我国在优化设计、自动控制、生产管理等方面也有广泛的应用。动态规划处理问题的方法和利用微分法或变分法来求极值问题的方法不同，在较简单的问题中，它通常是先提出所谓“最优性原理”，根据这个原理可导出一个动态规划递推方程，然后求其解，且通常求出的是整体最优解。在一般情况下，则须根据问题的动态规划模型建立并论证相应的动态规划基本方程，借助于它去求问题的解。然而，并不是每个多阶段决策问题都能应用动态规划来处理，用动态规划处理的问题必须具有这样的性质，即当任一阶段已经作出了决策后，下余过程构成的子问题恰巧和原先全过程的问题具有相同的结构，这样就使我们有可能在最优性原理的基础上写出递推关系式。

下面先从一个具体问题来说明动态规划方法的主要思想，并就此介绍有关的概念，然后就几类问题来说明应用动态规划方法解决问题的技巧及其在实际中的应用。

目　录

§1 最优性原理和多阶段决策过程

1.1 简单的最短路径问题

从 A 点到 B 点途中经过 $N-1$ 个中间点，由于这些中间点取得不同，所得路径也不同。最短路径问题就是从 A 点到 B 点的所有路径中寻求长度最小（或运输费用最小等）的路径。下面从一个简单的例子入手说明问题的解法。

例 1.1 设某城市的街道如图1.1所示，某人从 A 点出发要到达 B 点。图上所标的方向与数字表示各段可能的走法与里程（所标数字也可以表示走这段路所需的时间、费用，统称为支付量）。试求支付量最小的路径。

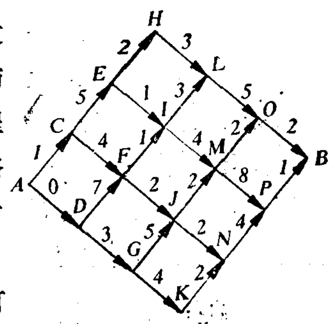


图1.1

当然，对这类问题，我们可以用“穷举法”，把从 A 点到 B 点的所有各种不同路径一一列出，分别算出各条路径的支付量，然后进行比较，从而找到最短路径。从 A 点到 B 点共有20条不同的路径，计算每一条路径的支付量需要计算 6

个数的和，即作5次加法，20条路径共需100次加法；然后比较大小，作一次比较可得出两个数中的较小者，这就共需作19次比较，总计需119次计算，才能求得需最小支付量的路径。若利用动态规划解法，则可用较小的计算量求出此问题的最优解。现在介绍它的动态规划解法。

当我们从 A 点出发时，不知道从 A 是沿着向上的方向走好还是沿着向下的方向走好，但假如我们已经知道从 C 点到 B 点的最优路径的支付量和从 D 点到 B 点的最优路径的支付量，那么我们就能作出 A 点的决策。若用 S_C 和 S_D 分别表示从 C 点和 D 点到 B 点的最小支付量，将 AC 段的支付量1和 S_C 相加，就能得到从 A 点沿向上方向出发的最优路径的支付量。类似地将 AD 段的支付量0和 S_D 相加，就得到从 A 点沿向下方向出发的最优路径的支付量，比较这二者的大小，就可作出 A 点的最优决策并得到从 A 点到 B 点的最小支付量值。这里要注意的是：要作出 A 点的决策，只需计算从 C 点、D 点出发到 B 点的最优路径的支付量，而无需计算从 C 点或 D 点到 B 点的其它9条路径的支付量，即不论 A 点的初始决策如何，从下一点出发必须沿着最优路径到达 B 点。若记 S_A 为从 A 点出发到 B 点的最小支付量，由于从 A 点出发只有两种可能的走法，因此可得

$$S_A = \min\begin{bmatrix}1 + S_C \\ 0 + S_D\end{bmatrix}$$

其中符号 $\min\begin{bmatrix}x \\ y\end{bmatrix}$ 表示取 x 和 y 两数值中之较小者。

但我们还不知道 S_C 和 S_D 的值，我们可以继续按上面讲的思想去做，也就是要求出 S_C 只需先求出从 E 点和 F 点

到 B 点的最优路径的支付量 S_E 和 S_F，即

$$S_C = \min \begin{bmatrix} 5 + S_E \\ 4 + S_F \end{bmatrix}$$

类似地

$$S_D = \min \begin{bmatrix} 7 + S_F \\ 3 + S_G \end{bmatrix}$$

S_E，S_F 和 S_G 一开始也是不知道其数值，它们的计算又取决于 S_H，S_I，S_J 和 S_K 的计算。依此继续做下去，可按类似的形式写出除 B 点外的其它每一点到 B 点沿最优路径的支付量，直到最后面的点 O 和 P。而这两点到 B 点的路径都只有一条，最小支付量直接可得，即

$$S_O = 2, \qquad S_P = 1$$

再从后往前逐步代入计算，就可得到其它各点到 B 点的最小支付量如下：

$$S_L = 5 + S_O = 7, \quad S_M = \min \begin{bmatrix} 2 + S_O \\ 8 + S_P \end{bmatrix} = 4, \quad S_N = 4 + S_P = 5$$

$$S_H = 3 + S_L = 10, \quad S_I = \min \begin{bmatrix} 3 + S_L \\ 4 + S_M \end{bmatrix} = 8,$$

$$S_J = \min \begin{bmatrix} 2 + S_M \\ 2 + S_N \end{bmatrix} = 6, \quad S_K = 2 + S_N = 7$$

$$S_E = \min \begin{bmatrix} 2 + S_H \\ 1 + S_I \end{bmatrix} = 9, \quad S_F = \min \begin{bmatrix} 1 + S_I \\ 2 + S_J \end{bmatrix} = 8,$$

$$S_G = \min \begin{bmatrix} 5 + S_J \\ 4 + S_K \end{bmatrix} = 11$$

$$S_C = \min \begin{bmatrix} 5 + S_E \\ 4 + S_F \end{bmatrix} = 12, \quad S_D = \min \begin{bmatrix} 7 + S_F \\ 3 + S_G \end{bmatrix} = 14$$

$$S_A = \min \begin{bmatrix} 1 + S_C \\ 0 + S_D \end{bmatrix} = 13$$

这样便求得从A点出发到B点的最小支付量S_A。但是，在解决实际问题时，我们更关心的是所求的最优路径究竟是哪一条。为此，我们只要在上面的反推计算中，每算出一个点到B点的最小支付量，同时标出获得该最优值的当前走向（后继点），它称为该点的最优决策。用 P_x 表示从 x 点出发沿最优路径到达的下一个点（即点x的最优决策），如$P_M=O$，这是因为$2+S_O$较$8+S_P$为小；又如$P_I=M$，这是因为$4+S_M$较$3+S_L$为小。可得各点的最优决策如下

$$P_O=B,\qquad P_P=B$$

$$P_L=O,\quad P_M=O,\quad P_N=P$$

$$P_H=L,\quad P_I=M,\quad P_J=M,\quad P_K=N$$

$$P_E=I,\quad P_F=J,\quad P_G=K\text{或 }J$$

$$P_C=F,\quad P_D=G$$

$$P_A=C$$

将上面得到的结果用图1.2表示，图中圆圈内的数字表示该点到B点最短路径的支付量。从图中可以找到各点到B点的最优路径，从A点到B点的最优路径是 $A—C—F—J—M—O—B$，在图中用粗黑线画出。沿此路径将各段上的支付量相加，得到这条路径的总支付量为

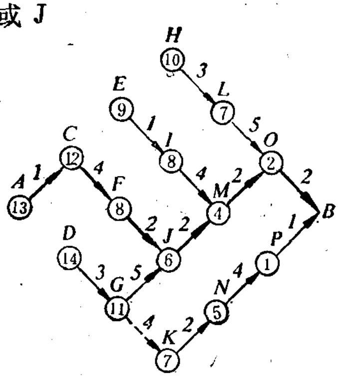


图1.2

$$1+4+2+2+2+2=13$$

和前面求得的S_A 相等。

我们也可以离开图形，从上面列出的表逐点找下去，找

到从A点到B点的最优路径。因为$P_A=C$，故先从A点到C点，而$P_C=F$，再由C点到F点，而$P_F=J$，故下一点到J点，又$P_J=M$，$P_M=O$，$P_O=B$，所以从A点到B点的最优路径是$A—C—F—J—M—O—B$。

很明显，所得到的从A点到B点的最优路径具有这样的特征：此路径上任一中间点到B点的路径（称之为由该点出发的子路径），都是该点到B点的最优路径，即最优路径的子路径也是最优的。

在上面的计算过程中，从H、L、O、K、N和P中的每一点出发到B点都只有一个决策，要计算从它们到B点的最小支付量，最多只需作一次加法，而其余九个点中的每一点都有两个初始决策，因此各需作两次加法和一次比较，这样总共需作二十四次加法和九次比较，较之穷举法的一百次加法和十九次比较要少得多。

1.2 多阶段决策过程 状态 决策

我们结合上面举的例子引入动态规划的一些基本概念。

例1.1所考虑的过程，可以分为若干阶段。从起点A出发走完第一段路程或到C点或到D点作为第一段，依此类推，从A点到B点是一个六阶段的过程。引入阶段变量来描述过程所处的阶段，阶段变量通常取非负整数值0，1，2，3，…。在每一阶段开始或是说上一阶段结束时，都面临着要作出新的决策，在最短路径问题中就是确定这一段的终点，决策不同，所得到的从A点到B点的路径也就不同。我

们把这种在每一阶段可以作出不同的决策去控制它的发展的过程称为多阶段决策过程。

多阶段决策过程的发展可以用过程各段的状态演变来描述。这些状态具有如下的性质：如果给定某一段的状态，则在这段以后过程的发展不受以前各段状态的影响，只和这一段的初始状态和状态演变规律有关，所有各段状态都确定时，整个过程也就确定。换言之，过程的每一实现可以用一状态序列表示。如上述最短路径问题中，每一段的状态就是该段路程的起始点，它同时也是前一段路程的终点，确定了这些点的序列，如前面写的 $A—C—F—J—M—O—B$，整个路径也就完全确定。又如引言中的例2，每一阶段的状态就是输入到各反应槽的待提纯物中的金属含量。

过程的上述性质意味着，过程的过去历史只能通过当前的状态去影响它的未来的发展，简单说就是“将来情况只和现在的状态有关而和过去的历史无关”。这一性质称为无后效性。如果状态仅仅描述过程的某种具体特征的话，那么并不是任何实际过程都满足上述的无后效性。但是要建立决策过程的动态规划模型，以状态描述的过程必须具有上述的无后效性。因此在构造决策过程的动态规划模型时，对状态的描述必须注意到这一点。对于实际过程，状态的某种描述方式可能导致不满足无后效性，但在适当改变状态的描述方式后，却可以得到满足无后效性的结果。例如，作为质点考虑的弹丸在外力作用下运动，假如我们要通过外力去控制在某确定时间区间内弹丸的轨迹，若从描述弹丸的轨迹这点着眼，可以弹丸在某一时刻的空间位置作为过程的状态，但这样选取

的状态却不满足无后效性。为了能达到满足无后效性的目的，我们必须取弹丸在每一时刻的空间位置以及速度矢量作为过程的状态。

在多阶段决策过程中，过程的状态通常可以用一个或一组变数来描述，称之为状态变量，即状态变量可以是一维，也可以是多维变量。在最短路径问题中，状态是各段的出发点，虽然它们不是数，但是如果把第k段的出发点编上号码1，2，3，…，如第3段的出发点可以是E，F，G三点，分别编上号码1，2，3，这样就可以用数来描述。我们可以引入变数$\boldsymbol{x}(k)$表示第k段的状态，如果是n维的，那就是状态向量$\boldsymbol{x}(k)=(x_1(k), x_2(k), \cdots, x_n(k))^T$，一组确定的数值对应着某一确定的状态。

在每一阶段，当状态给定后，我们可以选择不同决策，使以后各段的状态按不同的方式演变。在许多问题中，决策可自然而然地表示为一个数或一组数，不同的决策对应着这些变量的不同数值。描述决策的变量称为决策变量，如为m维决策变量用向量$\boldsymbol{u}=(u_1, u_2, \cdots, u_m)^T$表示。最短路径问题中，在第$k$段当初始状态确定时，所要选取的下一个点$P_k$就是决策变量，$P_k$一经确定，第$k+1$段的状态$x(k+1)$的值也就完全确定。一般地说，在第$k$段，设状态变量为$\boldsymbol{x}(k)$，决策变量为$u(k)$，如给出$u(k)$随$x(k)$变化的函数关系$u(x(k),k)$我们就确定了根据不同的当前状态作出不同决策的规则，此函数也就称为决策规则。

1.3 判别准则 最优性原理

在多阶段决策过程的最优化问题中，有一个能用以衡量由所采取的决策确定的过程好坏的数量指标，通常称为判别准则。判别准则是按问题要求的目标来确定的，如上述最短路径问题就是以总距离的长短为判别准则。当此数量指标为最小（有的问题则是最大）时可确定一最优值函数，如上例中的函数 S_M 就是从任一点 M 出发到 B 点的最短路径的长度，S 是变量 M 的函数，这里的变量 M 既表达了阶段又表达了该阶段的状态，当 $M=A$ 时，S_A 即为 A 点到 B 点的最短路长。我们也可以按另一种方式表示，用 $S(x, k)$ 表示状态为 x、阶段为 k 的点到终点 B 的最短路径长，所谓状态为 x、阶段为 k 的点是指从始点 A 到该状态为 x 的点经历了 k 个阶段。用 $d(L, M)$ 表示相邻两点 L、M 间的距离，$u(x, k)$ 是决策变量，它表示从状态为 x、阶段为 k 的点出发，在该阶段所选取的决策点。因此，前面的计算过程可以用递推关系式表示为

$$\begin{cases} S(x, k)=\min\limits_{u(x,k)}[d(x, u(x, k))+S(u(x,k), k+1)] \\ S(x, N-1)=d(x, B) \end{cases} \tag{1.1}$$

策略是指各个阶段的决策规则组成的总体，也就是决策规则序列，对应于最优决策规则的那一序列就是最优策略。设过程的始段为 0，终段为（$N-1$），对任一给定的 $0 \leqslant k \leqslant N-1$，由第 k 段开始到过程结束的过程称为原过程的 k- 子过

程。序列 $\{u(x(k), k), \cdots, u(x(N-1), N-1\}$称为$k$-子过程策略，而 $k=0$ 的序列 $\{u(x(0), 0), u(x(1), 1), \cdots, u(x(N-1), N-1)\}$就称为全过程策略，或简称策略。上述最短路径问题中，全过程策略就是从A点到B点各种不同路径上各段终点的序列，由全过程最优策略（即上述点 C—F—J—M—O—B）可得到从A点到B点的最短路径，k-子过程策略是从阶段数为 k的点到B点的各种不同路径上各段终点构成的子序列。

上面列出的递推关系式（1.1）也叫做动态规划的基本方程。可以看出，在上述最短路径中，它可用下面的“最优性原理”推导出来。

最优性原理的较一般提法如下：一个过程的最优策略具有这样的性质，即无论其初始决策如何，以后诸决策对以第一个决策所形成的状态作为初始状态而言，必须构成最优策略。拿上面的最短路径问题来说就是：从A点到B点的最优路径必须具有这样的性质，即无论A点的初始决策如何，该路径上从A点后面的任一点到B点的路径必须在该点到B点的所有可能路径中是最优的。

当选择第一个决策 $u(x, 0)$ 时，它有两种影响，其一是它直接影响$k=0$这个阶段距离的长度 $d(x, u(x,0))$，其二是它影响$k=1$-子过程的初始状态，因而也影响到该子过程的最短路径的长度 $S(u(x, 0), 1)$。最优策略的选择是根据统一考虑的结果决定的。根据最优性原理，对任一初始状态为 x的 k-子过程，不论第 k段的决策 $u(x, k)$如何，从 $u(x. k)$ 到 B 的 $k+1$-子过程必须是最优的，其最优值为