运筹学讲义

动态规划

中国科学院数学研究所运筹室编

中国人民大学经济信息管理系印

一九七九年二月

前 言

近三十年来，由于实际上和理论上的需要，最优化领域中的各种方法和理论蓬勃发展，无论在理论、方法和应用方面都取得了许多出色的成就。这些成果的取得，使最优化的武库不断增添新的内容。动态规化作为最优化的一个分支近年来也不断发展着。1951年，美国数学家R. Bellman等人根据一类多阶段决策问题的特性，提出了解决这类问题的"最优化原理"并研究了许多实际问题，从而建立了最优化的一个新分支——动态规划。动态规划的问世，在工程技术、经济、工业以及军事部门中引起了广泛的兴趣，许多问题，利用动态规划的思想来处理，也得到了良好的效果，因此吸引许多人在各个方面作新的尝试。在某些情况下，利用动态规划来处理问题，不但对问题能作出定性分析，而且也能给出利用电子计算机求其数值解的具体方法，这是动态规划的优点。现在，有关动态规划的书籍不断出版，许多最优化的书中也常常包括动态规划的内容。然而，它也有一些带根本性的弱点，一是利用最优化原理得出函数方法后，必须根据问题的各种性质结合其他数学技巧来求解，而没有一种统一的处理方法。二是变数的个数（维数）不能太大，常常有一些问题由于维数较大或涉及的函数没有理想的性质而使问题只能用动态规划来描述，而不能用动态规划的方法来求其解。可是，对于离散性的问题，由于解析数学无法施其技，而动态规划方法又成为一个有用的工具。

R. Bellman等人早年主要研究一些多阶段决策问题的函数方程解的存在性、唯一性问题以及一些应用。近年来，许多人则将动态规划应用于最优控制问题。

这些所讲的内容只是动态规划的简介，主要是通过一些问题来说明动态规划处理问题的基本思想和方法，请批评指正。

赖炎连

1979.2

于中国科学院数学研究所

目录

动态规划

1. 基本概念与最优化原理

1.1 多阶段决策问题及例

所谓多阶段决策问题，是指一类活动过程，它可以分为若干个互相联系的各个阶段，在每一个阶段都需要作出决策。每一个阶段的决策确定以后，就完全确定了一个过程的活动路线。各个阶段的决策确定以后就构成一个决策序列。我们称为一个策略。由于每一个阶段都有若干个决策可以选择，因而就有许多策略可供我们选取，对应于一个策略可以确定活动的效果，这个效果我们假定可以用数量来确定。策略的不同，效果也就不同。多阶段决策问题，就是要在容许我们选择的那些策略中间，选取一个最优策略，对应于这一策略，其效果在预定的标准下是最好的。

这一类问题是很广泛的，它形成数学问题以后或者是数学规划的问题，对时间连续的情形，则属于最优控制问题。动态规划处理这些问题的思想和方法是不一样的。"动态"的意思也就是指这类问题中，时间是个很重要的因素，在各个时间阶段采取的决策是与时间有关的。也有一些问题，表面上看来与时间无关，也可以变为一个多阶段决策问题。

多阶段决策问题常常可以碰到，现举几个例子：

1、最短路线问题

如图1.1，给出一个线路网络，A为起点，E为终点，两点之间的连线可以表示道路、管道、水渠或电路，称为弧。每一条弧根据实际情况与一个数值相对应。今要求从A到E的一

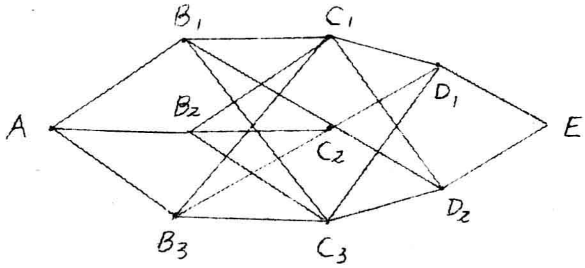


图 1.1

条连通弧，使其总弧长最小。

2、多阶段资源分配问题

设有数另 x 的某种资源，将它投入两种生产 A 与 B，若以数另 y 投入生产 A，剩下的另 $x-y$ 投入生产 B，则可得到收入为

$$g(y)+h(x-y)$$

其中 $g(y)$ 与 $h(y)$ 为已知函数，且 $g(0)=h(0)=0$，显然，对不同的 y 值，收入不同。现设以 y 与 $x-y$ 投入 A、B 生产后可以回收再投入生产，设其回收率分别为 $0\leq a<1$ 与 $0\leq b<1$，因此在第一阶段生产后回收的总资源为

$$x_1=ay+b(x-y).$$

我们将 x_1 再投入生产 A、B，和第一阶段一样，若以 y_1 与 x_1-y_1 分别投入生产 A 与 B，则又可得到收入

$$g(y_1)+h(x_1-y_1)$$

因此，两个阶段的总收入为

$$g(y)+h(x-y)+g(y_1)+h(x_1-y_1)$$

如果上面的过程进行 n 个阶段而且我们希望选择 $y, y_1, y_2, \cdots y_{n-1}$ 使 n 个阶段的总收入最大，则我们的问题便成为

$$\max_{y, y_1, \cdots y_{n-1}} \{ g(y) + h(x-y) + g(y_1) + h(x_1-y_1) + g(y_2) + h(x_2-y_2) + \cdots + g(y_{n-1}) + h(x_{n-1}-y_{n-1}) \}$$

满足条件

$$\text{(a)} \quad \begin{aligned} x_1 &= ay + b(x-y) \\ x_2 &= ay_1 + b(x_1-y_1) \\ &\vdots \\ x_{n-1} &= ay_{n-2} + b(x_{n-2}-y_{n-2}) \end{aligned}$$

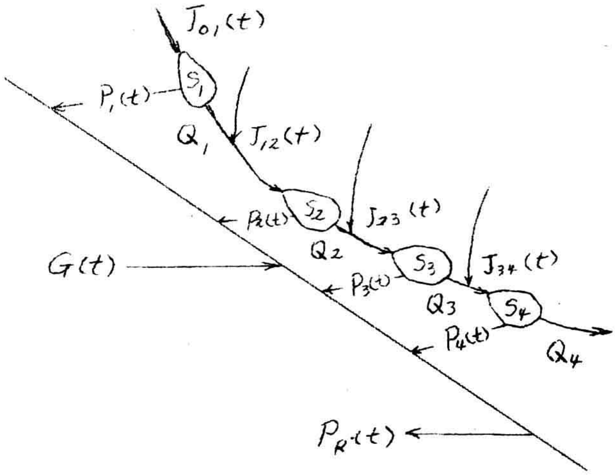
(b) $0 \le y \le x$，$0 \le y_i \le x_i$，$i = 1, 2, \cdots n-1$

3、连轧机最佳压下制度问题

连轧机的工作过程，也是一个多阶段决策过程。设连轧机有 6 个轧辊，按反向编号 6、5、4、3、2、1。当坯料来了之后，先通过 6 号轧辊轧一次，若入口板厚为 H_6，轧制之后压下率为 r_6，然后进入 5 号轧辊轧制，压下率为 r_5，依次往下一直经过最后一道轧辊轧成我们所需要的厚度 H。每次的压下率 r_i 是可以选择的，但有一定的工艺条件限制，我们的目的就是要在一定的操作条件下，求出最好的压下率 r_i，使得在保证钢板质量的情况下使确定的 $\{r_i\}$ 成为最佳压下制度

4、水库调度问题

一条河上有四个水电站 S_1、S_2、S_3、S_4，另有一个火电站，它们构成一个电力系统向周围地区提供电力。假定他们提供的电力能够满足用户要求，火电厂机四起动较坊也希望校定地进



行工作，水电厂控制发电比较容易，但下游电厂受上游水库蓄放水量的影响，因此调度的重点放在水电站上面：

令

$P_R(t)$ —— 时刻 t 用户要求的电力；

$G(t)$ —— 时刻 t 火电厂的发电量；

$P_i(t)$ —— 时刻 t 水电站 i 的发电量； $i=1,2,\cdots,4$，

$S_i(t)$ —— 时刻 t 水库 i 的蓄水容量；

$Q_i(t)$ —— 时刻 t 水库 i 的发电用水量。

以 $J_{01}(t), J_{12}(t), J_{23}(t), J_{34}(t)$ 记 S_1, S_2, S_3, S_4 的天然流量，$\tau_{12}, \tau_{23}, \tau_{34}$ 分别为上游水库放水后流至下游水库的时间。因此可得关系式：

$$\frac{d}{dt}S_1(t)=J_{01}(t)-Q_1(t)$$

$$(5)\qquad \frac{d}{dt}S_2(t)=J_{12}(t)+Q_1(t-\tau_{12})-Q_2(t)$$

$$\frac{d}{dt}S_3(t)=J_{23}(t)+Q_2(t-\tau_{23})-Q_3(t)$$

$$\frac{d}{dt}S_4(t)=I_{34}(t)+Q_3(t-\tau_{34})-Q_4(t)$$

若 $f(G(t))$ 为火电厂提供电力 $G(t)$ 时的燃料等费用，并略去水电耗损，则我们希望在时段 $[0,T]$ 内，在电力平衡的情况下，使火电厂的费用最小。即求，

$$\min\int_0^T f(G(t))\,dt$$

满足：

$$P_R(t)-\sum_{i=1}^{4}P_i(t)-G(t)=0 \qquad (1)$$

$$\underline{G}\le G(t)\le\overline{G}\ ;\quad \underline{Q_i}\le Q_i(t)\le\overline{Q}_i\ ,$$

$$\underline{S_i}\le S_i(t)\le\overline{S}_i \qquad (2)$$

由于 $P_i(t)$ 是 $S_i(t)$，$Q_i(t)$ 的函数，利用(1)式消去 $G(t)$，则问题变为求 $\{S_i(t),\ Q_i(t)\}$，满足(5)及(2)，使

$$\int_0^T f(S_1(t),\ S_2(t),\ S_3(t)\ S_4(t),\ Q_1(t),\ Q_2(t),$$
$$Q_3(t),\ Q_4(t),\ t)\,dt$$

最小。

5、转炉顶吹炼钢问题

一个转炉，在一定的操作规程下生产某种钢，若在顶吹氧气冶炼过程中温度 T 和脱碳速度的变化规律为

$$\frac{dT(t)}{dt}=P(C(t),\ T(t),\ u(t))$$

$$\frac{dC(t)}{dt}=Q(C(t),\ T(t),\ u(t))$$

其中：

$C(t)$ —— 时刻 t 时单位体积铁水的含碳量；

$T(t)$ —— 时刻 t 时的铁水温度；

$u(t)$ —— 时刻 t 时的吹氧流量。

若已知入炉时单位体积铁水的含碳量 $C(0)=C_0$，铁水温度 $T(0)=T_0$，若冶炼时间为 20 分钟，以 $\overline{C}$、$\overline{T}$ 分别表示出钢时的合格含碳量与温度，问题是如何控制 $u(t)$，使在给定的出钢时间 $\overline{t}$，使

$$(C(\overline{t})-\overline{C})^2+(T(t)-\overline{T})^2$$

最小。

6、生产与储存问题

设在时段 $[0,T]$ 内，在时刻 t，$x(t)$ 为某种产品的实际生产率，$r(t)$ 为对产品的需求率。我们生产该产品以满足需求，但当 $x(t)>r(t)$ 时，需要把剩余产品存入仓库，设其存储费用为 $F(x(t)-r(t))$；若增加生产即 $\frac{dx}{dt}>0$ 时，由于增加生产而引起的费用率为 $G(\frac{dx}{dt})$，其中 F、G 为已知函数，我们的问题是在满足需求下求生产与储存的费用最小，即

$$\min\int_0^T\left[F(x(t)-r(t))+G\left(\frac{dx}{dt}\right)\right]dt$$

满足

$$x(t)\geqslant r(t)$$

1.2 最优路线问题与最优化原理

现在，我们先用一个最短路线的问题来说明动态规划方法

的基本思想。

如图1.2，我们从A出发经过各点至目的地F，问沿什么路线经过那些点而使总路程最小？这是一个五个阶段的问题。由A至B（B_1、B_2、B_3）中的一点是第一阶段，B中的点至C

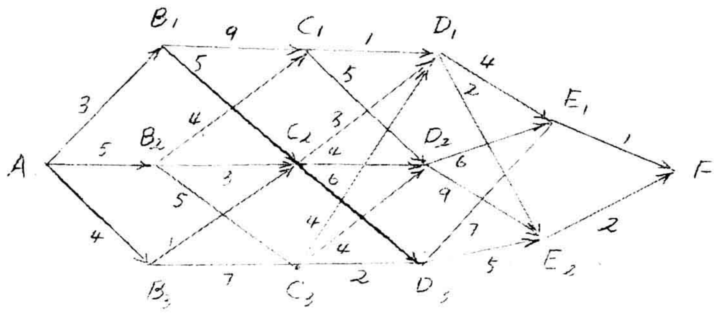


图 1.2

中的点是第二阶段，C至D是第三阶段，D至E是第四阶段，E至F是第五阶段。

现在我们引进几个符号与概念。

令n表示由某点至终点F之间的阶段数，例如，由B_2至F是四个阶段。n称为阶段变量。

令S表示在任一阶段所处的位置，S称为状态变量。

令$x_n(S)$为决策变量，它表示当状态S还有n个阶段要走时所要选取的下一个点，策略是指各个阶段的决策组成的总体。

令$f_n(S)$表示现在处在状态S（即处在S点）还有n个阶段要走，由S至终点F的最小距离。

令$d(S, x_n)$表示S点至$x_n(S)$的距离。

我们从最后一个阶段开始计标，由定义，$f_1(E_1)$表示由E_1至F的最小距离，故$f_1(E_1)=1$，同样，$f_1(E_2)=2$。现计标

最后二个阶段，$n=2$，我们从 D_1 出发可以有二个选择，一个是到 E_1，一个是到 E_2，故有

$$f_2(D_1)=\min\left\{\begin{array}{l} d(D_1,E_1)+f_1(E_1) \\ \\ d(D_1,E_2)+f(E_2) \end{array}\right\}=\min\left\{\begin{array}{l} 4+1 \\ 2+2 \end{array}\right\}$$

$$=4.$$

上式说明，由 D_1 至终点 F 的最小距离是 4，其路线是 $D_1 \to E_2 \to F$，且相应的决策变量 $x_2(D_1)=E_2$。同理，

$$f_2(D_2)=\min\left\{\begin{array}{l} d(D_2,E_1)+f_1(E_1) \\ \\ d(D_2,E_2)+f_1(E_2) \end{array}\right\} \min\left\{\begin{array}{l} 6+1 \\ 9+2 \end{array}\right\}=7.$$

因此，由 D_2 至 F 的最小路线是 $D_2 \to E_1 \to F$，距离是 7
$x_2(D_2)=E_1$，同理可得 $f_2(D_3)=7$，路线是 $D_3 \to E_2 \to F$，$x_2(D_3)=E_2$。

现在再看 $n=3$ 的情况，起点有 C_1，C_2，C_3

$$f_3(C_1)=\min\left\{\begin{array}{l} d(C_1,D_1)+f_2(D_1) \\ \\ d(C_1,D_2)+f_2(D_2) \end{array}\right\}=\min\left\{\begin{array}{l} 1+4 \\ 5+7 \end{array}\right\}=5$$

$x_3(C_1)=D_1$，上式表示由 C_1 出发可有两种选择，一个是到 D_1，一个是到 D_2（C_1 至 D_3 无路可通），如果选择 D_1，则以后应显然是选择由 D_1 至 F 的最小路线，如果选择 D_2，则以后也必然应该选择由 D_2 至 F 的最小路线，然后再在两条路线中选择一条小的，即 $C_1 \to D_1 \to E_2 \to F$。同理，若从 C_2 出发此时有三个选择。

$$f_3(C_2)=\min\begin{Bmatrix} d(C_2, D_1)+f_2(D_1) \\ d(C_2, D_2)+f_2(D_2) \\ d(C_2, D_3)+f_2(D_3) \end{Bmatrix}=\min\begin{Bmatrix} 8+4 \\ 4+7 \\ 6+7 \end{Bmatrix}=11$$

故由 C_2 至 F 的最优路线是 $C_2 \to D_2$（由 $D_2 \to E_1 \to F$），距离为 11，$x_3(C_2)=D_2$，同理，$f_3(C_3)=8$，$x_3(C_3)=D_1$．

对 $n=4$，同理可得

$f_4(B_1)=14$，　　$x_4(B_1)=C_1$．

$f_4(B_2)=9$，　　$x_4(B_2)=C_1$．

$f_4(B_3)=12$，　　$x_4(B_3)=C_2$．

当 $n=5$ 时，出发点只有一个 A 点．

$$f_5(A)=\min\begin{Bmatrix} d(A, B_1)+f_4(B_1) \\ d(A, B_2)+f_4(B_2) \\ d(A, B_3)+f_4(B_3) \end{Bmatrix}=\min\begin{Bmatrix} 3+14 \\ 5+9 \\ 4+12 \end{Bmatrix}=14$$

$x_5(A)=B_2$，因此最优路线已经求得：

$$A \to B_2 \to C_1 \to D_1 \to E_2 \to F.$$

最short距离是 14．

从上面的计算过程中，我们可以看出，在求解的各个阶段，我们利用了 n 个阶段与 $n-1$ 个阶段之间的如下关系．

$$\begin{cases} f_n(S)=\min\limits_{x_n(S)}\{d(S, x_n(S))+f_{n-1}(x_n(S))\}, \\ f_1(S)=d(S, F). \end{cases}$$

我们将会看到，动态规划就是根据这种递推关系求得问题的解的。上述方程称为动态规划的函数方程。而函数方程是根据动态规划的最优化原理而推导出来的，它可以述之如下：

动态规划最优化原理

“一个过程的最优策略具有这样的性质，即无论其初始状态及初始决策如何，其以后诸决策对以第一个决策所形成的状态作为初始状态的过程而言，必须构成最优策略”。

对函数方程

$$f_n(S) = \min_{x_n(S)} \left\{ d(S, x_n(S)) + f_{n-1}(x_n(S)) \right\}.$$

来说，当第一个决策 $x_n(S)$ 选定时，它有二个影响，一是直接影响第一个阶段的距离的长度 $d(S, x_n(S))$，一是影响后面 $n-1$ 个阶段的过程的初始状态，因而也影响到后面 $n-1$ 个阶段的最小距离。上述方程表明，最优策略的选取是两者统一考虑的结果而决定的。

从上述例子中可以看到，一个多阶段决策过程的极值函数可以看作是过程的初始状态与阶段数目的函数，现在我们将用更一般的方式来描述这类多阶段决策过程。

设有一个物理系统，在任何时刻系统的状态可以用向量 p 来表示，p 的分量称为状态变量。所谓一个决策就是将系统由状态 p 变到 p'，状态的变化也可以看作是经过一个变换而得到的，即一个决策可以看作是一个变换 T：$T(p) = p'$。与此同时，系统由于状态的变换而得到一定数量的效益，所得效益的多少依赖于状态 p 与所作的决策。

在每一个阶段，可供选择的决策，有时是有限个，也可以是无限个，一般我们用向量 q 表示一个决策，$q \in \Omega$，Ω 是决

策集合，不同的阶段（或不同的时刻）有不同的决策集合。令$T_q(p)=p'$表示由状态p采用决策q后变为状态p'，$R_q(p)$表示当初始状态为p使用决策q时所得到的各阶段的效益。

我们的目的是要使n个阶段的过程的总效益最大。由于最大总效益只是初始状态p与阶段数目n的函数，若将它记为$f_n(p)$，则由上述的最优化原理，可得函数方程

$$
(2)\qquad
\begin{cases}
f_1(p)=\max\limits_{q\in Q} R_q(p) \\
f_k(p)=\max\limits_{q\in Q}\left[R_q(p)+f_{k-1}(T_q(p))\right],
\end{cases}
$$

$$k=2,3,\cdots n.$$

综合上面的例子，我们可以看到这组函数方程有如下特点：

1、已知阶段总数为n的过程，则阶段的编号是依n，$n-1$，$\cdots$ 2，1从左到右编排的（见图1-2）。这种编排常常是与实际的问题的阶段序号相反，因此，从左到右，我们顺次得出递推关系：

$$
(3)\qquad
\begin{cases}
f_n(p)=\max\limits_{q\in Q_n}\left\{R_q^{(n)}(p)+f_{n-1}(T_q^{(n)}(p))\right\} \\
f_{n-1}(p)=\max\limits_{q\in Q_{n-1}}\left\{R_q^{(n-1)}(p)+f_{n-2}(T_q^{(n-1)}(p))\right\} \\
\qquad\vdots \\
f_2(p)=\max\limits_{q\in Q_2}\left\{R_q^{(2)}(p)+f_1(T_q^{(2)}(p))\right\} \\
f_1(p)=\max\limits_{q\in Q_1} R_q^{(1)}(p).
\end{cases}
$$

2、在求函数方程的解的时候，从求 $f_1(P)$ 开始，逐次求解最后求出 $f_n(P)$，所以求解顺序是与过程的发展顺序相反的。而且 $f_1(P)$ 它是作为状态 P 的函数而得到的，在求 $f_2(\hat{P})$ 的时候，对于 $f_1(P)$，我们要求对于 $f_1(T_q^{(2)}(\hat{P}))$，当给定 $\hat{P}$ 以后，对于一切状态 $P=T_q^{(2)}(\hat{P})$，$f_1(P)$ 均应是已知的。对于上述例子而言，若 $\hat{P}=D_1$，则对从 D_1 出发采取所有可能的决策后得到的状态集合而言，$f_1(P)$ 均应是已知的。但在过程的最后一个阶段，我们已经求得了 $f_1(E_1)$，$f_1(E_2)$，因此满足了求解的要求。同时，我们也可注意到，不同的阶段上的最优决策是该阶段的初始状态的函数，当我们最后求得 $f_n(P)$ 以后，我们便得到依赖于 P 的各个阶段的最优决策函数 $r_n(P)$，$r_{n-1}(P)$，…… $r_2(P)$，$r_1(P)$。

3、在求 n 个阶段的最优策略的时候，由于初始的状态是已知的，例如 P_0，则在阶段 n，由 $h_n(P)$ 得出 $r_n(P_0)$，它也就是剩下的 $n-1$ 个阶段的初始状态，因此又可得出 $r_{n-1}(r_n(P_0))$ 为第二次变换后的状态，同时它也是剩下的 $n-2$ 阶段的过程的初始状态，等等。于是逐次变换便可求得最优过程的路线。

1.3 函数方程的求解

现在我们讨论例2的问题与更一般的最优路线问题。

对于例2，当 $K=1$ 时，

$$f_1(x)=\max_{0\le y\le x}\left[g(y)+h(x-y)\right]$$

当 $K=2$ 时，由于头一个阶段分别以 y、$x-y$ 投入 A、B 生产以后可以回收 $x_1=ay+b(x-y)$ 作为下一个阶段开始时可以投入 A、B 生产的资源数量，若采取最优分配的方式生产，则得

到时剩下阶段的总收入仅是 x_1 的函数，若以 $f_2(x)$，$f_1(x_1)$ 分别记二个阶段与一个阶段各以 x，x_1 投入生产时采取最优策略后的总收入，则由最优化原理，我们有：

$$f_2(x)=\max_{0\le y\le x}\left[g(y)+h(x-y)+f_1(ay+b(x-y))\right]$$

当 $k=n$ 时，应用上面的分析，便有

$$\begin{cases} f_n(x)=\max\limits_{0\le y\le x}\left[g(y)+h(x-y)+f_{n-1}(ay+b(x-y))\right] \\ f_1(x)=\max\limits_{0\le y\le x}\left[g(y)+h(x-y)\right] \end{cases}$$

现在对 $n=3$，$a=0.3$，$b=0.6$，$g(y)=0.8y$，$h(y)=0.5y$ 来求解

于是

$$f_1(x)=\max_{0\le y\le x}\left[0.8y+0.5(x-y)\right]=\max_{0\le y\le x}\left[0.3y+0.5x\right]$$

由于 $0.3>0$，故不论 x 为何值，当 $y=x$ 时，$0.3y+0.5x$ 取得最大值，所以

$$f_1(x)=0.8x$$

$$f_1(ay+b(x-y))=f_1(0.3y+0.6(x-y))$$

$$=f_1(0.6x-0.3y)=0.8(0.6x-0.3y)$$

$$=0.48x-0.24y.$$

将上式代入 f_2 中，则得