动态规划及其应用

张之驳　李建德　编著

国防工业出版社

动态规划及其应用

张之骅　李建德　编著

国防工业出版社

(京)新登字106号

图书在版编目(CIP)数据

动态规划及其应用/张之𬳶等编著.—北京:国防工业出版社,1994

ISBN 7-118-01211-4

Ⅰ.动…

Ⅱ.张…

Ⅲ.①动态规划-应用数学　②应用数学-动态规划

Ⅳ.O221.3

动态规划及其应用

张之𬳶　李建德　编著

责任编辑:张戬山

*

国防工业出版社出版发行

(北京市海淀区紫竹院南路23号)

(邮政编码　100044)

新华书店经售

北京大兴兴达印刷厂印装

*

开本 850×1168　1/32　印张 5¼　134千字

1994年6月第1版　1994年6月北京第1次印刷　印数:1　2000册

ISBN　7-118-01211-4/O·94　定价:5.50元

前　　言

本书是根据编者多年来对动态规划的教学和科研的有关讲义和资料整理而成的。可作为国防科技、军事院校和工科院校的研究生和师资进修班的通用教材，也可供从事最佳控制、运筹学、系统工程、经济管理等专业的科技人员参考。

本书详细系统地叙述了动态规划的基本概念、基本原理和基本方法，也比较详细地介绍了多方面的实际应用问题(如“气象导航”、“武器系统的火力分配”等)。本书紧密结合多阶段决策过程的控制与优化，系统地叙述了确定性、离散性、随机性和连续性的动态规划模型，以及微分动态规划模型，其内容比较全面、系统实用。

本书把五重组$\{X,U,T,v,V\}$、最优化原理以及函数迭代法和策略迭代法作为决策过程的基本概念、基本原理和基本方法的主线贯穿始终，我们认为，这有助于增强本书的系统性和深度。

本书努力贯彻“理论联系实际”的原则，如注意从具体问题引出概念和原理，并着重对实际应用问题的分析和介绍等。同时，力求文理通顺，文字易懂，以便读者阅读理解。

在编写过程中参考了许多有关著作，限于篇幅不能一一列举，在此一并表示感谢。

由于编者水平有限，错误和不妥之处在所难免，敬请读者批评指正。

编　者

1990 年 8 月于大连舰院

内容简介

本书介绍动态规划的基本概念、基本方法并详细地介绍了多方面的实际应用。

本书结合多阶段决策过程的控制与优化，叙述了确定性、离散性、随机性和连续性的动态规划模型，以及微分动态规划模型。本书把五重组、最优化原理以及函数迭代法和策略迭代法作为决策过程的基本概念、基本原理和基本方法的主线贯穿始终。

目　录

第一章　最短行军路线问题 …… 1
§1.1　最短行军路线问题 …… 1
§1.2　最短行军路线问题的数学模型 …… 4
§1.3　最短行军路线问题的顺序解法 …… 10
第二章　动态规划的基本概念和基本原理 …… 12
§2.1　动态规划的基本概念 …… 12
§2.2　动态规划的基本定理 …… 13
§2.3　最优化原理 …… 14
§2.4　动态规划的基本方程(数学模型) …… 16
§2.5　可逆过程及其顺序解法 …… 17
第三章　动态规划的应用 …… 19
§3.1　气象导航问题 …… 19
§3.2　武器指挥决策系统的火力分配问题 …… 22
§3.3　搜索力的最优分配问题 …… 44
§3.4　战斗时间的优化问题 …… 48
§3.5　最优装载问题 …… 54
§3.6　复合系统的可靠性问题 …… 56
§3.7　多级火箭的最优设计问题 …… 61
§3.8　生产计划问题 …… 63
§3.9　海图成图周期的优化问题 …… 68
§3.10　实验仪器使用问题 …… 70
第四章　不定期与无期动态规划 …… 74
§4.1　不定期最优路线问题 …… 74
§4.2　函数迭代法 …… 76
§4.3　策略迭代法 …… 83
§4.4　不定期动态规划 …… 90

§4.5 无期动态规划 …… 93
第五章 随机性动态规划 …… 97
§5.1 随机性决策过程的基本概念 …… 97
§5.2 马尔可夫决策过程 …… 98
§5.3 (MDP)的有限阶段目标函数模型 …… 99
§5.4 (MDP)的 F 有限折扣目标函数模型 …… 109
§5.5 独立干扰的决策过程 …… 118
第六章 连续型动态规划 …… 123
§6.1 连续型动态规划模型 …… 123
§6.2 最优性的必要条件——HJB 方程 …… 126
§6.3 应用举例 …… 128
第七章 微分动态规划 …… 132
§7.1 引言 …… 132
§7.2 策略局部微调的微分动态规划(DDP—CM) …… 134
§7.3 策略全局调整的微分动态规划(DDP—CJ) …… 137
§7.4 算法的收敛性 …… 140
练习题 …… 143
练习题答案 …… 155

第一章　最短行军路线问题

在这一章中，我们将通过一个具体而典型的例子（最短行军路线问题），引出有关动态规划的一些名词和记号，进而得到动态规划的数学模型。

§1.1　最短行军路线问题

问题　图1.1.1中给出一个行军路线网络，从 A 点要走到 G 点，中间要经过 B、C、D、…… 等很多站，各站间的距离如图中所示，今要求选择一条由 A 点到 G 点的最短行军路线。

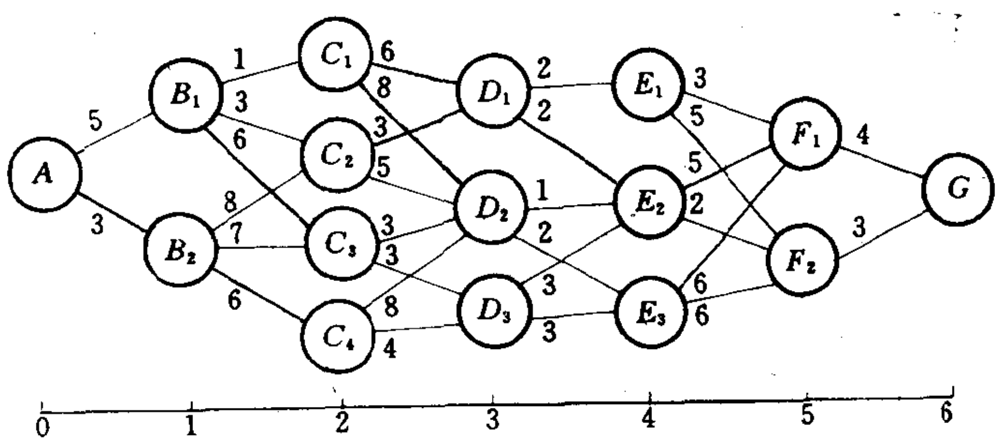

图　1.1.1

这是一个多阶段决策问题。从 A 到 G 可以分为6个阶段，从 A 站出发到 B 站为第一阶段。这时有两个选择：一是到 B_1 站；一是到 B_2 站。若我们选择到 B_2 站的决策，则 B_2 站就是第一阶段决策的结果，它既是第一阶段的终点，又是下一阶段（第二阶段）路线的始点。在第二阶段，再从 B_2 站出发，对应于 B_2 点就有一个可供选择的终点集合 $\{C_2,C_3,C_4\}$。若选择由 B_2 走到 C_2 为第二阶段的决策，

则 C_2 就是第二阶段的终点，同时又是第三阶段的始点。类似地可以递推下去，直到终点 G 站。我们看到：各个阶段的决策不同，所走的路线也就不同。现在要求：在各个阶段中选取一个恰当的决策，使由这些决策所决定的一条路线，其总距离最短。

下面我们介绍"标号法"的求解方法。

首先要注意到下面一个明显而重要的事实：如果某一条路线，如 $A \to B_1 \to C_2 \to D_1 \to E_2 \to F_2 \to G$ 是最优路线，那么无论从该路线中的哪一站开始(如从 D_1 站开始)到达终点 G 站的那一段路线，仍然是从该点(即 D_1 点)到达终点 G 的所有可能选择的不同的路线的最优路线(称为由 D_1 出发的最短子路线)。这一事实，以后我们称之为"**最优化原理**"。因为如果不是这样，从 D_1 点到终点还有另一条更短的子路线存在，那么把它和原来最短路线由始点 A 到达 D_1 点的那部分连接起来，就会形成一条比原来最短路线更短的路线，而这是不可能的。

根据上面的事实，我们可以从后段开始，逐段往前求最优子路线，从而得到全过程最优路线。在逐段求最优子路线时，采用"标号法"，见图 1.1.2。

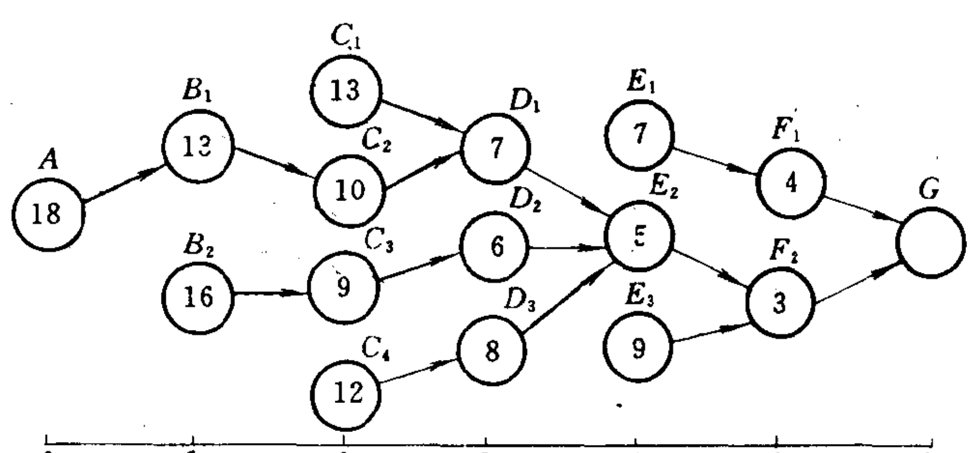

图 1.1.2

(1) 先标第五段：因从 F_1 或 F_2 到 G 点，仅各有一条路线，因而也分别是它们的最短路线，于是将其最短距离值 4 和 3 分别标写在 F_1、F_2 的圆圈内，并划上路径 $F_1 \to G$ 及 $F_2 \to G$。

(2) 再标第四段：从 E_1 点出发有两种决策可供选择：一是从 E_1 过 F_1 到 G 点，其距离值为 $3+4=7$；另一是从 E_1 过 F_2 到 G 点，其距离值为 $5+3=8$。这样，从 E_1 出发过 F_1 到 G 点是最短路径，将其距离值 7 标写在 E_1 的圆圈内，并划上路径 $E_1 \to F_1$。同样，从 E_2 出发，也有两种决策可供选择，即 $E_2 \to F_1 \to G$ 和 $E_2 \to F_2 \to G$，而最优点为 $E_2 \to F_2 \to G$，其值为 5，将 5 标写在 E_2 的圆圈内，并划上路径 $E_2 \to F_2$。同理，E_3 到 G 的最优路径为 $E_3 \to F_2 \to G$，其值为 $6+3=9$，将 9 标写在 E_3 的圆圈内，并划上路径 $E_3 \to F_2$。

(3) 标第三段：从 D_1 出发，也有两种决策可供选择：一是过 E_1 去 G 点，其最优距离为 $D_1E_1 + E_1$ 到 G，最优距离 $= 2+7=9$；另一是过 E_2 去 G 点，其最优距离为 $D_1E_2 + E_2$ 到 G，最优距离 $= 2+5=7$。因此，从 D_1 出发，过 E_2 去 G 的路线是最优路线，将其距离值 7 标写在 D_1 圆圈内，并划上路径 $D_1 \to E_2$。同理，从 D_2 出发，过 E_2 去 G 点是它的最优路径，将其距离值 $1+5=6$ 标写在 D_2 圆圈内，并划上 $D_2 \to E_2$。同理，在 D_3 的圆圈内标写上从 D_3 到 G 的最优距离值 $3+5=8$，并划上最优路径 $D_3 \to E_2$。

(4) 继续仿上述按逆推过程一直标写到始点 A，在 A 点也有两种决策可供选择：一是过 B_1 去 G 点，其最优距离为 $5+13=18$；另一是过 B_2 去 G 点，其最优距离为 $3+16=19$。因此，应选过 B_1 的路线，将其最优距离值 18 标写在 A 圆圈内，并划上其最优路径 $A \to B_1$。

于是，从图 1.1.2 中易知，我们从 A 出发，顺推回去，即得最优路径：

$$A \to B_1 \to C_2 \to D_1 \to E_2 \to F_2 \to G$$

其最优距离值为 18。

上述标号法要比穷举法优越得多，表现在 ① 计算量少：穷举法要对 48 条路径进行比较，即进行比较运算 47 次，即使用逐段累加方法也要进行加法运算 $6+12+24+48+48=138$ 次。而标号法共进行比较运算 $3+3+4+4+1=15$ 次，每次比较运算相应有两次加法运算，再去掉中间重复的两次（即 $B_1 \to C_1$，$B_2 \to$

C_4 各多算了一次)，实际只有 28 次加法运算。② 丰富了计算结果：标号法计算结果，使我们不仅得到由 A 出发到达终点 G 的最短路线及其相应的最短距离值，而且得到了从所有各中间站出发到达 G 点的最短路线和它相应的最短距离，这对许多实际问题来讲是很有用的，有利于帮助分析所得结果。

§1.2　最短行军路线问题的数学模型

一、几个名词

求解最短行军路线问题的标号法，事实上就是动态规划方法的具体体现，因此，我们从该问题的标号法解法中，可以看到动态规划方法的几个主要特征：

(1) 阶段(Stage)：最短行军路线问题是一个多阶段决策过程。它共分为六个相互联系的阶段。常用K表示阶段变量，$K=0,1,\cdots,5$。对有 n 个阶段的问题，K 的值为：$K=0,1,\cdots,n-1$。

(2) 状态(State)：最短行军路线问题中，各阶段都有若干个站，这些站，它是该段以后某路径的出发点，也是前一段路径的终点，这些站叫做状态(各阶段的站叫做该阶段的状态)。如第二阶段有四个状态，即点集合$\{C_1,C_2,C_3,C_4\}$。

如果给定过程某一阶段的状态，那么在这段以后过程的发展要受到该段给定状态的影响，而不受该段以前各段状态的影响。这就是说：过程的发展，只受当前的状态的影响，过去的历史只能通过当前的状态去影响它的未来，而不能直接影响它的未来。这种特性叫做无后效性。在建立实际问题的动态规划模型时，其状态的描述，要求其满足无后效性。

描写状态的变量，叫状态变量。第k阶段的状态变量x_k的取值集合(称为状态集合)为 $X_k=\{1,2,\cdots,i,\cdots,r\}$，或表示为

$$x_k\in X_k=\{x_k^{(1)},x_k^{(2)},\cdots,x_k^{(r)}\}$$

例如最短行军路线问题的第二阶段的状态集合可记为

$$X_2=\{x_2^{(1)},\cdots,x_2^{(4)}\}=\{1,2,3,4\}=\{C_1,C_2,C_3,C_4\}$$

(3) 决策(Decision)：决策就是某阶段状态给定后，从该状态演

变到下一阶段某状态的选择。描述决策的变量，称为决策变量。常用 $u_k(x_k)$ 表示第 k 阶段处于第 x_k 状态时采取的决策。显然 $u_k(x_k)$ 是状态 x_k 的函数，它的取值范围称为允许决策集合，通常以 $D_k(x_k)$ 表示第 k 阶段处于状态 x_k 的允许决策集合，即

$$u_k(x_k) \in D_k(x_k)$$

例如，在最短行军路线问题中，第一阶段的状态集合是 $X_1 = \{B_1, B_2\}$，则从 B_1 站出发，它可能有三种决策，即 x_k 取值 B_1 时，其决策集合为 $D_1(B_1) = \{C_1, C_2, C_3\}$，如选择从 B_1 到 C_2 的路径，则 $u_1(B_1) = C_2$；如选择从 B_1 到 C_3 的路径，则 $u_1(B_1) = C_3$。同理，

$$u_1(B_2) \in D_1(B_2) = \{C_2, C_3, C_4\}$$

(4) 状态转移：在最短行军路线问题中，当给定第 k 段状态变量 x_k 的值 $x_k^{(i)}$，如果决策变量 $u_k(x_k^{(i)})$ 的值一经确定(如选取从第 k 阶段第 i 点到第 $k+1$ 阶段第 j 点的路径)，则第 $k+1$ 阶段的状态变量 x_{k+1} 的值也就完全确定，即 $x_{k+1} = x_{k+1}{}^{(j)}$，这时，

$$x_{k+1} = u_k(x_k^{(i)}) = x_{k+1}{}^{(j)}.$$

由此可见，一般来说，x_{k+1} 的值随 x_k 和 u_k 的值变化而变化，这种变化关系可用函数表示为：

$$x_{k+1} = T_k(x_k, u_k)$$

上式表示了由第 k 阶段到第 $k+1$ 阶段的状态转移规律，称为状态转移方程。T_k 依赖于 u_k 表明状态的转移与所选取的决策有关。

(5) 策略(Policy)：假设给定问题可分为 n 个阶段，$k=0, 1, \cdots, n-1$，那么由第 0 阶段开始到第 $n-1$ 阶段终点为止的过程称为问题的全过程。由每段的决策函数 $u_k(x_k)$，$k = 0, 1, \cdots, n-1$，组成的决策序列，就称为全过程策略，简称策略，记为 P_{0n}。即

$$P_{0n}(x_0) = \{u_0(x_0), u_1(x_1), \cdots, u_{n-1}(x_{n-1})\}$$

如最短行军路线问题中决定最优路线的策略是：

$$P_{06}(A) = \{B_1, C_2, D_1, E_2, F_2, G\}$$

由第 k 阶段开始到全过程的终点为止的过程，称为原过程的后部子过程(或称为 k 子过程)。其决策函数序列称为 k 子过程策

略，简称 k 子策略。记为：

$$p_{kn}(x_k) = \{u_k(x_k),\cdots,u_{n-1}(x_{n-1})\}$$

在实际问题中可供选择的策略有一定的范围。所有可供选择的策略所组成的集合，称为允许策略集合，用 P 表示，即

$$p_{0n}(x_0) \in P_{on}(x_0) \text{ 或 } P_{kn}(x_k) \in P_{kn}(x_k)$$

从允许策略集中找出使问题达到最优效果的策略称为最优策略，记为 $p_{0n}^*(x_0)$。

(6) 报酬函数：当过程处于状态 x_k，并采取决策 u_k 而得到的报酬（或费用）。显然，它是定义在 $X_k \times D_k$ 上的函数，称为第 k 段的报酬函数，记为 $v_k(x_k,u_k)$。在最短路线问题中，它表示第 k 段由点 x_k 到第 $k+1$ 段点 x_{k+1} 之间的距离，即 $v_k(x_k,u_k) = C_{ij}$，其中 $x_k = i$，$u_k = j$。

(7) 目标函数：在决策过程问题中，用来衡量所实现过程的优劣，定义在全过程和所有后部子过程上的确定的数量函数，叫做目标函数（或称指标函数）。

若我们考虑的过程是从第 k 段开始到过程终点的 k 子过程，则目标函数可表示为：

$$\begin{aligned} V_{kn} &= V_{kn}(x_k,x_{k+1},x_{k+2},\cdots,x_n) \\ &= V_{kn}(x_k,u_k,u_{k+1},\cdots,u_{n-1}) \\ &= V_{kn}(x_k,p_{kn}(x_k)) \end{aligned}$$

其最优目标函数值可表示为：

$$\begin{aligned} f_k(x_k) &= \underset{(u_k,\cdots,u_{n-1})}{\text{opt}} V_{kn}(x_k,u_k,u_{k+1},\cdots,u_{n-1}) \\ &= \underset{p_{kn}(x_k)\in P_{kn}(x_k)}{\text{opt}} V_{kn}(x_k,p_{kn}(x_k)) \\ &= V_{kn}(x_k,p_{kn}^*(x_k)) \end{aligned}$$

这里“opt”是英文 optimization（最优化）的缩写，可根据题意而取 min 或 max。$p_{kn}^*(x_k)$ 表示初始状态为 x_k 的后部子过程所有子策略中的最优策略。

在很多实际问题中，目标函数还满足以下递推关系：对任意的 $k(0 \leqslant k < n)$ 有

$$V_{kn}(x_k,u_k,\cdots,u_{n-1}) = \Phi(x_k,u_k,V_{k+1,n}(x_{k+1},u_{k+1},\cdots,u_{n-1}))$$

或 $$V_{kn}(x_k,p_{kn}(x_k)) = \Phi(x_k,u_k,V_{k+1,n}(x_{k+1},p_{k+1,n}(x_{k+1})))$$

在不同的问题中，目标函数也不同，可能是距离、利润、成本、产品的产量或资源的消耗等等。例如，在最短行军路线问题中，目标函数 V_{kn} 就是第 k 段由点 x_k 到达终点 G 的距离：

$$\begin{aligned} V_{kn} &= d_{kn}(x_k,x_{k+1},\cdots,x_n) \\ &= d_{kn}(x_k,u_k,u_{k+1},\cdots,u_{n-1}) \\ &= \sum_{j=k}^{n-1} v_j(x_j,u_j) \end{aligned}$$

其中 $v_j(x_j,u_j)$ 表示第 j 段的报酬函数，也即第 j 段的距离。显然，它满足递推关系：

$$\begin{aligned} V_{kn} &= v_k(x_k,u_k) + \sum_{j=k+1}^{n-1} v_j(x_j,u_j) \\ &= v_k(x_k,u_k) + V_{k+1,n}(x_{k+1},\cdots,x_n) \end{aligned}$$

或有 $$\begin{aligned} V_{on}(x_0,p_{on}(x_0)) &= \sum_{j=0}^{k-1} v_j(x_j,u_j) + \sum_{j=k}^{n-1} v_j(x_j,u_j) \\ &= V_{ok}(x_0,p_{ok}(x_0)) + V_{kn}(\tilde{x}_k,p_{kn}(\tilde{x}_k)) \end{aligned}$$

其中 $\tilde{x}_k$ 是由前一段子过程在子策略 p_{ok} 下确定的。

二、最短行军路线问题的动态规划数学模型

在最短行军路线问题中，用标号法求得最短路径和它的最短距离值。如果我们引用上节所介绍的概念和记号，将求解过程逐步写下来，那么就可归纳出该问题求解的动态规划的数学模型。具体如下：

(1) 第 5 段：$f_5(F_1) = 4,\quad F_1 \to G$

$f_5(F_2^*) = 3,\quad F_2 \to G^*$

(2) 第 4 段：

$$\begin{aligned} f_4(E_1) &= \min_F \{v_4(E_1,F) + f_5(F)\} \\ &= \min\{v_4(E_1,F_1) + f_5(F_1), v_4(E_1,F_2) + f_5(f_2)\} \\ &= \min\{3+4, 5+3\} = 7 \end{aligned}$$

其最优点为 F_1，即 $E_1 \to F_1$

$$f_4(E_2^*) = \min_F \{v_4(E_2,F) + f_5(F)\}$$
$$= \min\{v_4(E_2,F_1) + f_5(F_1), v_4(E_2,F_2) + f_5(F_2)\}$$
$$= \min\{5 + 4, 2 + 3\} = 5$$

其最优点为 F_2，即 $E_2 \to F_2^*$

$$f_4(E_3) = \min_F \{v_4(E_3,F) + f_5(F)\}$$
$$= \min\{v_4(E_3,F_1) + f_5(F_1), v_4(E_3,F_2) + f_5(F_2)\}$$
$$= \min\{6 + 4, 6 + 3\} = 9$$

其最优点为 F_2，即 $E_3 \to F_2$

同理可推得：

(3) 第 3 段：$f_3(D_1^*) = \min_E \{v_3(D_1,E) + f_4(E)\} = 7$

最优点为 E_2^*。

$$f_3(D_2) = \min_E \{v_3(D_2,E) + f_4(E)\} = 6$$

最优点为 E_2。

$$f_3(D_3) = \min_E \{v_3(D_3,E) + f_4(E)\} = 8$$

最优点为 E_2。

(4) 第 2 段：$f_2(C_1) = \min_D \{v_2(C_1,D) + f_3(D)\} = 13$

最优点为 D_1。

$$f_2(C_2^*) = \min_D \{v_2(C_2,D) + f_3(D)\} = 10$$

最优点为 D_1^*。

$$f_2(C_3) = \min_D \{v_2(C_3,D) + f_3(D)\} = 9$$

最优点为 D_2。

$$f_2(C_4) = \min_D \{v_2(C_4,D) + f_3(D)\} = 12$$

最优点为 D_3。

(5) 第 1 段：$f_1(B_1^*) = \min_C \{v_1(B_1,C) + f_2(C)\} = 13$

最优点为 C_2^*。

$$f_1(B_2) = \min_C \{v_1(B_2,C) + f_2(C)\} = 16$$

最优点为 C_3。

(6) 第 0 段：$f_0(A) = \min\limits_{B}\{v_0(A,B) + f_1(B)\} = 18$

最优点为 B_1^*。

于是，由后反推回去（上述各段中带 * 号的式子）即得最优路径：$A \to B_1^* \to C_2^* \to D_1^* \to E_2^* \to F_2^* \to G^* = G$。

若引用状态变量 x_k 和决策变量 u_k 的记号，则上述各式可合并成下列方程组，这就是最短路线问题的动态规划数学模型：

$$
(\mathrm{DP})\begin{cases} f_k(x_k) = \min\limits_{u_k(x_k)\in D_k(x_k)} \{v_k(x_k, u_k(x_k)) + f_{k+1}(u_k(x_k))\} \\ \qquad k = 5,4,3,2,1,0 \\ \text{并规定：} f_6(x_6) = 0 \end{cases} \tag{1.2.1}
$$

这里，DP 是 Dynamic Programming 的缩写。$f_6(x_6)$ 表示第 6 段

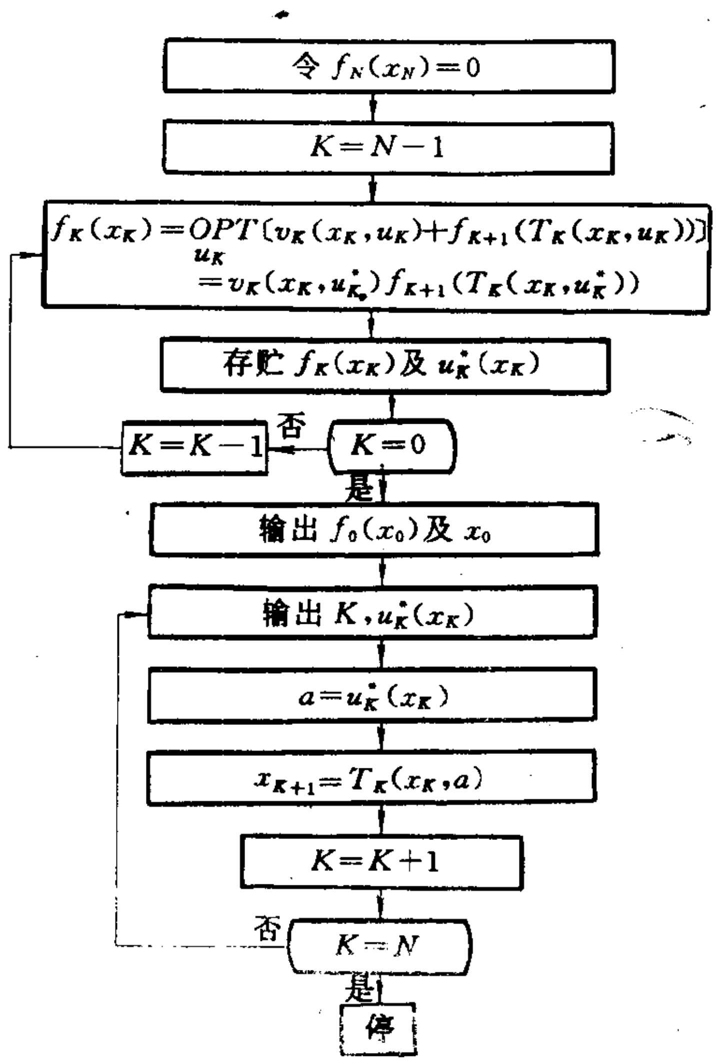

图 1.2.1

的 x_6 点(即 G 站)到过程的终点(也是 G 站)的最短距离,当然,它的值为 0。

一般地,多阶段确定性决策过程的动态规划数学模型为:

$$
(\mathrm{DP})\begin{cases} f_k(x_k) = \underset{u_k(x_k)\in D_k(x_k)}{\mathrm{opt}} \{v_k(x_k,u_k) + f_{k+1}(x_{k+1})\} \\ \qquad k = N-1, N-2, \cdots, 2, 1, 0 \\ \text{并规定:} f_N(x_N) = 0 \end{cases} \tag{1.2.2}
$$

注意 式(1.2.2)中目标函数是各段报酬函数和的形式。由于状态转移方程为 $x_{k+1} = T_k(x_k,u_k)$,所以式(1.2.2)又叫做该决策过程的逆序解法的基本方程。

方程组(1.2.2)的解算框图见图 1.2.1。

§1.3 最短行军路线问题的顺序解法

在最短路线问题中,因从 A 站出发到终点 G 站,故从 A 站到 G 站的顺序划分阶段,但解题时却是从 G 站开始逆推到 A 站,这种解法叫做逆序解法(或反向法)。但应指出,对于最短路线问题,它的两端都是固定的,且路线上的数字表示两点间的距离,那么从总体看,以 A 点为始点 G 为终点的最短路线,应与以 G 点为始点 A 为终点的最短路线相同。因而,我们也可以按 A 到 G 的顺推秩序来求解此问题。如用标号法求解,其顺推标号列于图 1.3.1。

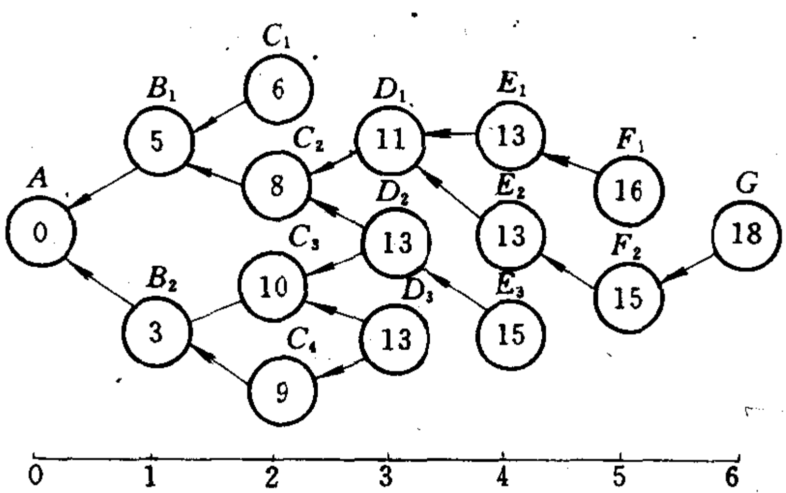

图 1.3.1