第九章

动态规划

第九章 动态规划

- 9.1 动态规划的基本思想
  - 9.1.1多阶决策问题
  - 9.1.2最优性原理
- 9.2 动态规划的基本方程
  - 9.2.1 动态规划的基本方程
  - 9.2.2 应用动态规划基本方程解最优控制问题
- 9.3 连续时间系统的动态规划

§ 9.1 动态规划的基本思想

- 动态规划是用来解最优控制问题的另一种方法。
- 在二十世纪50年代，贝尔曼在研究多阶段决策问题时提出了动态规划法。
- 离散系统的最优控制问题可以看做一个多阶段决策问题，因此可用动态规划求解。
- 动态规划的指导思想简单，可以方便地将一个复杂的多阶段决策问题化为一系列的一阶段决策问题，使问题得到简化，可以顺序求解,从而它已成为解多阶段决策问题的一种有效方法。
- 动态规划已被广泛应用于解很多技术领域的动态最优化问题，如生产管理问题，资源分配问题，设备更新问题，多级工艺设备的优化设计问题和工程控制问题等。

9.1.1　多阶段决策问题

- **1.最短路径问题**

  最**短路**径问题可以看做是多阶段决策过程的一个例子，通过它可以了解利用动态规划解多阶段决策问题的基本思想。

考虑如图9-1的最**短路**径问题。

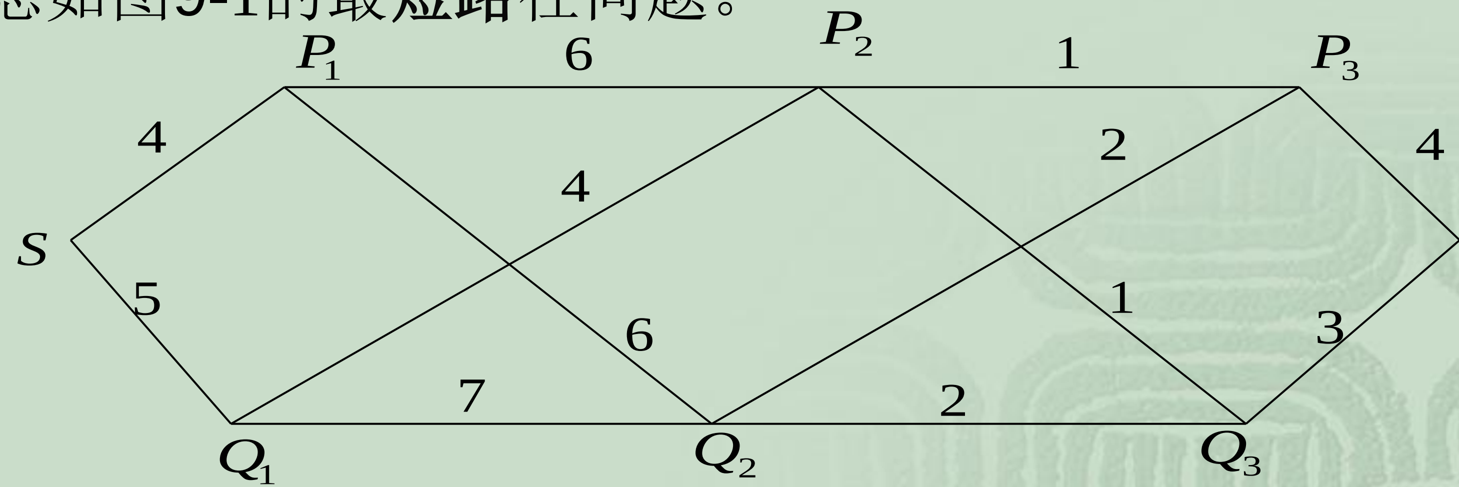

图9-1最短路径问题

在图9-1中，一汽车由 S 点出发到终点 F, P_i和Q_i 是一些可以通过的点。图中两点之间标出的数字是汽车走过这一段路程所需的距离（时间）。最**短路**径问题是确定一种走法使汽车由 S 到 F 所走距离（用时间）最短。

1.最优性原理——最短路径的任何最后一段还是最短路径。

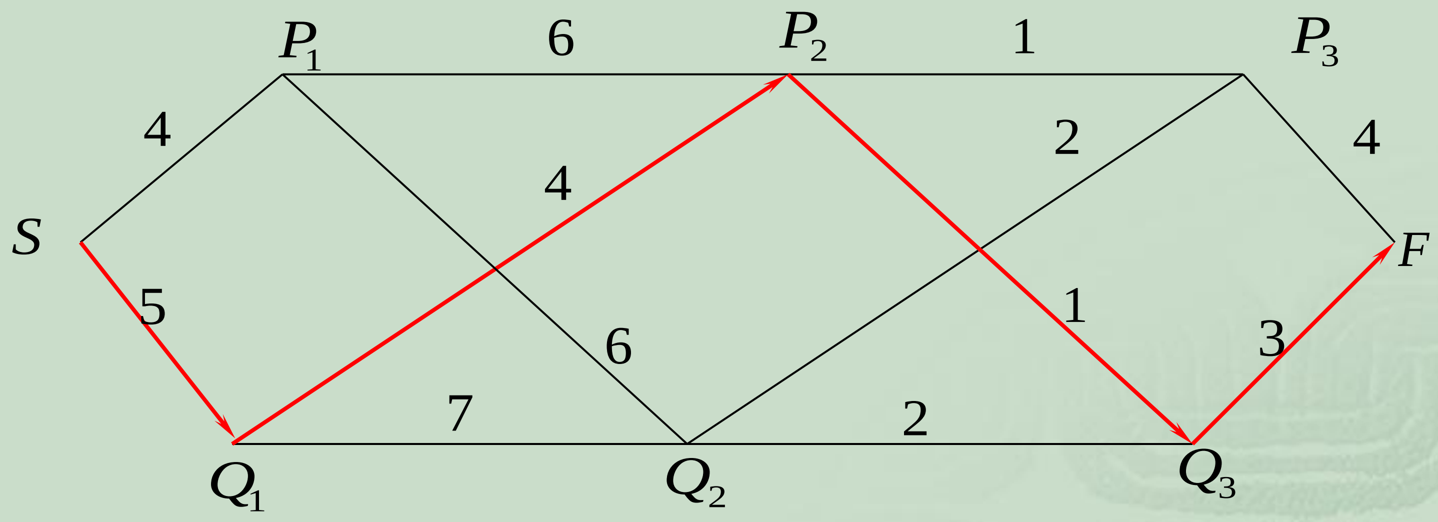

图9-2最短路径

在图9-2中，如果 $SQ_1P_2Q_3F$，是所求的最短路径，那么汽车从这一路径上任一点，例如 P_2，出发到 F 的最优途径必为 P_2Q_3F。根据这一原理我们可以由后向前递推求解最短路径问题。

2.计算步骤

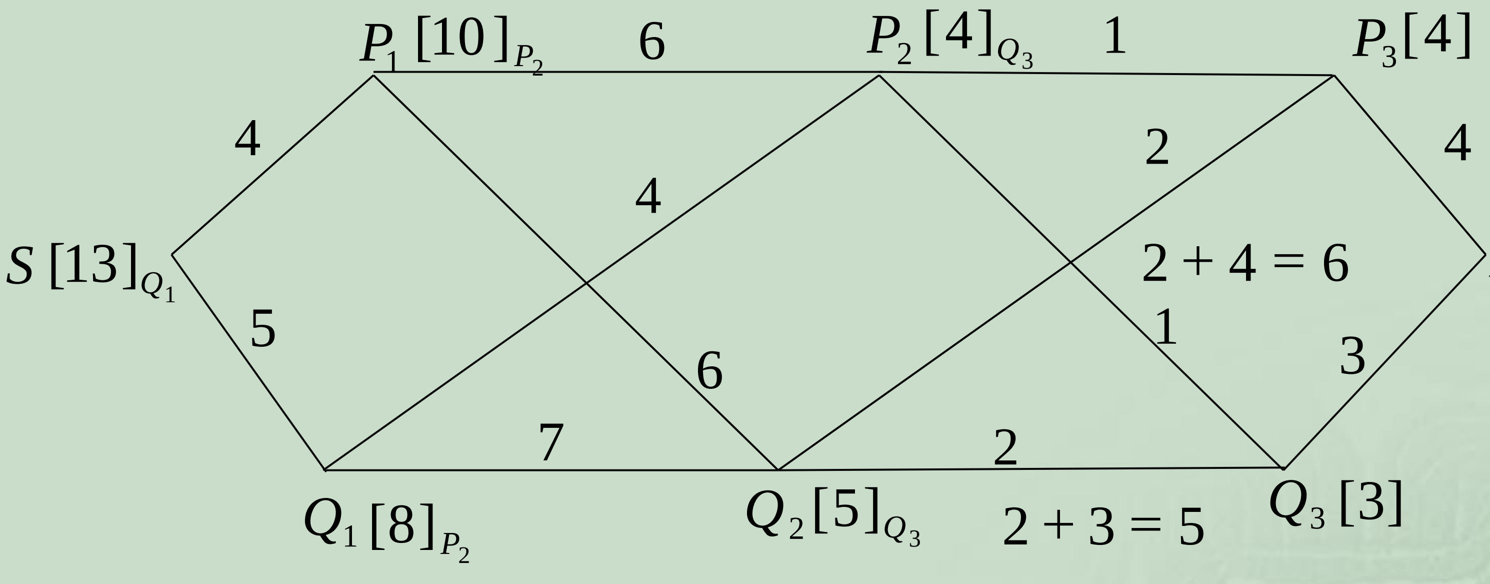

图9-3　最短路径的求解

结果：

由图9-3可见，汽车从 S 点出发到达 F 点最快需要13，最短的路径是 S Q_1 P_2 Q_3 F，最短时间 13。

按照前面的方法，由后向前递推的方法节省了大量运算，当汽车可经过的点增多时节省的运算次数会更加明显。10次加法运算。

如采用穷举法，需要将所有可能的途径一一列出，计算每种途径所需要的时间，然后再选出需要时间最短的途径，这样共需做24次的加法。

最短路径问题可以看做一个多阶段决策问题

由 S 到 P_1 或Q_1做为第一阶段，

由P_1，Q_1 到 P_2，Q_2 做为第二阶段，

由P_2，Q_2 到 P_3，Q_3 做为第三阶段，

由P_3，Q_3 到 F 做为第三阶段。

上面的求最**短路**径的方法，是把一个四阶段的最优决策问题，化成四个互相嵌套的子问题求解，从而使问题得到简化，这种方法叫动态规划法。

2.多阶段决策问题的一般提法

设系统的状态方程为

$$x(k+1) = f(x(k), u(k), k)$$

- 无后效性
- 目标函数的可分性

目标函数为

$$J_N = \sum_{k=0}^{N-1} L(x(k), u(k), k)$$

J_N 表示控制 N 步时的目标函数。

多阶段决策问题就是求最优控制策略序列 $u(0), u(1), \ldots, u(N-1)$ 使 J_N 最小（或最大）。

显然在这里多阶段决策问题也就是一个N 步最优控制问题。

假设状态方程中 f 和目标函数L 都不明显地依赖于时间变量k，x 和 u 都是标量，这时状态方程为

$$x(k+1)=f(x(k),u(k)) \quad \textbf{(9-1)}$$

$$J_N=\sum_{k=0}^{N-1}L(x(k),u(k)) \quad \textbf{(9-2)}$$

假设初始状态 $x(0)=x_0$ 是给定的，对目标函数(9-2)逐次应用式(9-1),可以得到

$$\begin{aligned}J_N &=L(x(0),u(0))+L(x(1),u(1))+\cdots+L(x(N-1),u(N-1))\\&=L(x(0),u(0))+L(f(x(0),u(0)),u(1))+\cdots\\&+L(f(f\cdots(f(x(0),u(0)),u(1)),\cdots u(N-2)),u(N-1))\end{aligned}$$

上式表明 J_N只依赖于 $x(0),u(0),u(1),\cdots u(N-1)$ 这样可记为

$$J_N = J_N(x(0),u(0),u(1),\cdots,u(N-1))$$

如果已用某种方法求出最优策略$u^*(0),u^*(1),\cdots,u^*(N-1)$, 那么$J_N$的最小值只依赖于初始值 $x(0)$, 把这个最小值记为$J_N^*(x(0))$, 于是

$$J_N^*(x(0)) = \min_{u(0),\cdots u(N-1)} J_N(x(0),u(0),\cdots,u(N-1))$$

初始状态 $x(0)$是可以变化的，用 $J_N^*(x)$ 表示初始条件为x，控制步数为 N 的目标函数的最小值。

9.1.2　最优性原理

- **定理9-1（最优性定理）**

  如果 $u^*(0),u^*(1),\cdots,u^*(N-1)$ 是最优策略序列，那么

  $u^*(k),u^*(k+1),\cdots,u^*(N-1)$ 也是一个最优控制策略序列，

  其初始状态是

$$x(k)=f(x^*(k-1),u^*(k-1))\qquad k\geq 1$$

证明　用反证法，设 $u^*(k),u^*(k+1),\cdots,u^*(N-1)$ 不是最优策略序列，而 $v^*(k),v^*(k+1),\cdots,v^*(N-1)$ 是最优策略序列，

那么

$$J_{N-k}(x(k),u^*(k),\cdots,u^*(N-1))>J_{N-k}(x(k),v^*(k),\cdots,v^*(N-1))$$

对 N 阶过程应用策略序列：$u^*(0),\cdots,u^*(k-1),v^*(k),\cdots,v^*(N-1)$

则有

$$
\begin{aligned}
&J_N(x(0),u^*(0),\cdots,u^*(k-1),v^*(k),\cdots,v^*(N-1)) \\
&= J_k(x(0),u^*(0),\cdots,u^*(k-1)) + J_{N-k}(x(k),v^*(k),\cdots,v^*(N-1)) \\
&< J_k(x(0),u^*(0),\cdots,u^*(k-1) + J_{N-k}(x(k),u^*(k),\cdots,u^*(N-1)) \\
&= J_N(x(0),u^*(0),\cdots,u^*(N-1))
\end{aligned}
$$

即

$$
\begin{aligned}
J_N(x(0),u^*(0),\cdots,u^*(k-1),\ v^*(k),\cdots,v^*(N-1)) \\
< J_N(x(0),u^*(0),\cdots,u^*(N-1))
\end{aligned}
$$

这与 $u^*(0),\cdots,u^*(N-1)$ 是最优策略序列矛盾，证毕。

§ 9.2 动态规划的基本方程

- 动态规划的基本方程

动态规划的基本方程给出 N 阶决策问题的目标函数的最优值与它的子问题（一个N-1 阶决策问题）的目标函数的最优值之间的递推关系。它是应用动态规划解多阶决策问题（N 步最优控制问题）的基础。

设 $u^*(0)$ 已求出，那么求 $u^*(1),\cdots,u^*(N-1)$ 的问题构成一个初始条件为

$$x(1)=f(x(0),u^*(0))$$

的 N-1 阶决策问题。

$J_N^*(x(0))$表示初值为 $x(0)$, 控制步数为N 时目标函数的最小值

$J_{N-1}^*(x(1))$ 表示初值为 $x(1)$,控制步数为N-1时目标函数的最小值

可以导出 $J_N^*(x(0))$ 与 $J_{N-1}^*(x(1))$ 之间的关系如下:

$$J_N^*(x(0)) = \min_{u(0),\cdots,u(N-1)} \{\sum_{k=0}^{N-1} L(x(k),u(k))\}$$

$$= \min_{u(0),\cdots,u(N-1)} \{L(x(0),u(0)) + \sum_{k=1}^{N-1} L(x(k),u(k))\}$$

上式第一项$L(x(0),u(0))$ 不依赖于 $u(1),\cdots,u(N-1)$, 因此上式可改写为

$$J_N^*(x(0)) = \min_{u(0)} \{L(x(0),u(0)) + \min_{u(1),\cdots,u(N-1)} \sum_{k=1}^{N-1} L(x(k),u(k))\}$$

$$= \min_{u(0)} \{L(x(0),u(0)) + J_{N-1}^*(x(1))\}$$

这样我们就得到了递推关系

$$J_N^*(x(0)) = \min_{u(0)}\{L(x(0),u(0)) + J_{N-1}^*(x(1))\} \qquad \textbf{(9-3)}$$

——动态规划的基本方程

式中 $x(1) = f(x(0),u(0))$

类似于上面的推导，可以得到

$$J_{N-i}^*(x(i)) = \min_{u(i)}\{L(x(i)),u(i)) + J_{N-i-1}^*(x(i+1))\} \qquad \textbf{(9-4)}$$

——动态规划基本方程的更一般的形式

式中

$$x(i+1) = f(x(i),u(i))$$

利用这一递推关系可以把一个多阶决策问题化为若干个子问题，在每个子问题中只对一个变量进行最优化，例如：

$$J_1^*(x(N-1)) = \min_{u(N-1)}\left\{L(x(N-1),u(N-1))\right\}$$

是一个对单变量 $u(N-1)$ 的最优化问题，当 $J_1^*(x(N-1))$ 求出后，由动态规划的基本方程(9-3)得到

$$J_2^*(x(N-2)) = \min_{u(N-2)}\left\{L(x(N-2),u(N-2)) + J_1^*(x(N-1))\right\}$$

式中 $x(N-1) = f(x(N-2),u(N-2))$

这又是对变量 $u(N-2)$ 的最优化问题。依次类推，通过解一系列的单变量最优化问题即可得到最优控制序列

$$u^*(0),u^*(1),\cdots,u^*(N-1)$$

前面的讨论假设了 $x(k),u(k)$ 是标量，所得结果对它们都是向量的情况也适用。这时式(9-1)化为

$$x(k+1)=f(x(k),u(k))$$

式(9-2)化为

$$J_N=\sum_{k=0}^{N-1}L(x(k),u(k))$$

动态规划的基本方程(9-4)化为

$$J_{N-1}^*(x(i))=\min_{u(i)}\{L(x(i),u(i))\}+J_{N-i-1}^*(x(i+1))\}$$

- 动态规划基本方程可用来解各类离散的最优控制问题

【例9-1】 设系统的状态方程为

$$x(k+1)=ax(k)+bu(k),\qquad a，b\ \text{是常数}$$

目标函数是

$$J_3=\sum_{k=0}^{2}[x^2(k)+qu^2(k)]\qquad\qquad q>0$$

求 $u(0),u(1),u(2)$ 使J_3最小。

（1）先考虑一步最优控制问题。此时

$$J_1 = x^2(2) + qu^2(2)$$

$$J_1^*(x(2)) = \min_{u(2)}\{x^2(2) + qu^2(2)\}$$

当 $u(2) = 0$时 $x^2(2) + qu^2(2)$ 达最小值，因此

$$u(2) = 0 \qquad\qquad J_1^*(x(2)) = x^2(2)$$

（2）由动态规划的基本方程

$$J_2^*(x(1)) = \min_{u(1)}\{x^2(1) + qu^2(1) + J_1^*(x(2))\}$$

$$= \min_{u(1)}\{x^2(1) + qu^2(1) + x^2(2)\}$$

将状态方程 $x(2)=ax(1)+bu(1)$ 代入上式，则得

$$J_2^*(x(1))=\min_{u(1)}\{x^2(1)+qu^2(1)+(ax(1)+bu(1))^2\}$$

令 $\dfrac{\partial J_2(x(1))}{\partial u(1)}=2qu(1)+2b[ax(1)+bu(1)]=0$，得到

$$u^*(1)=\frac{-ab}{q+b^2}x(1)$$

代入 $J_2^*(x(1))$ 中，得到

$$J_2^*(x(1))=x^2(1)+q\frac{a^2b^2}{(q+b^2)^2}x^2(1)+\left[ax(1)-\frac{ab^2}{q+b^2}x(1)\right]^2$$

$$=(1+\frac{ab^2}{q+b^2})x^2(1)$$

（3）再一次利用基本方程，有

$$J_3^*(x(0)) = \min_{u(0)}\{x^2(0) + qu^2(0) + J_2^*(x(1))\}$$

$$= \min_{u(0)}\{x^2(0) + qu^2(0) + (1 + \frac{qa^2}{q+b^2})[ax(0) + bu(0)]^2\}$$

令 $\dfrac{\partial J_3}{\partial u(0)} = 0$ 得到

$$u^*(0) = -\frac{ab(q+b^2+qa^2)}{(q+b^2)^2 + qa^2b^2}x(0)$$

这样通过三步计算得到最优控制序列为

$$u^*(0) = -\frac{ab(q+b^2+qa^2)}{(q+b^2)^2 + qa^2b^2}x(0) \qquad u^*(1) = \frac{-ab}{q+b^2}x(1) \qquad u^*(2) = 0$$

它们都是状态变量的反馈。

- 【例9-2】 系统的状态方程为

$$x(k+1)=0.5x(k)+0.3u(k)$$

目标函数为

$$J_4=\sum_{k=0}^{3}[3x(k)-u(k)]$$

求 $u^*(0),u^*(1),u^*(2),u^*(3)$使 J_4 最小，

并满足约束条件 $0\le u(k)\le x(k)$ 。

解 （1）由最后一步开始

$$J_1^*(x(3))=\min_{0\le u(3)\le x(3)}\{3x(3)-u(3)\}$$

显然 $u^*(3)=x(3)$ 时 $3x(3)-u(3)$ 最小，这时

$$J_1^*(x(3))=2x(3)$$

（2）由动态规划的基本方程

$$J_2^*(x(2)) = \min_{0\le u(2)\le x(2)}\{3x(2)-u(2)+J_1^*(x(3))\}$$

$$= \min_{0\le u(2)\le x(2)}\{3x(2)-u(2)+2[0.5x(2)+0.3u(2)]\} = \min_{0\le u(2)\le x(2)}\{4x(2)-0.4u(2)\}$$

显然 $u^*(2)=x(2)$ 使 $4x(2)-0.4u(2)$ 最小，这时 $J_2^*(x(2))=3.6x(2)$

（3）再一次应用动态规划的基本方程

$$J_3^*(x(1)) = \min_{0\le u(1)\le x(1)}\{3x(1)-u(1)+J_2^*(x(2))\}$$

$$= \min_{0\le u(1)\le x(1)}\{3x(1)-u(1)+3.6x(2)\}$$

$$= \min_{0\le u(1)\le x(1)}\{3x(1)-u(1)+3.6(0.5x(1)+0.3u(1))\} = \min_{0\le u(1)\le x(1)}\{4.8x(1)+0.08u(1)\}$$

显然 $u^*(1)=0$ 使 $4.8x(1)+0.08u(1)$ 最小，这时

$$J_3^*(x(1)) = 4.8x(1)$$

（4）

$$J_4^*(x(0)) = \min_{0\le u(0)\le x(0)}\{3x(0) - u(0) + J_3^*(x(1))\}$$

$$= \min_{0\le u(0)\le x(0)}\{5.4x(0) + 0.44u(0)\}$$

$$u^*(0) = 0 \qquad J_4^*(x(0)) = 5.4x(0)$$

求出的最优控制序列是 $u^*(0) = 0, u^*(1) = 0, u^*(2) = x(2), u^*(3) = x(3)$。

这个例子中应用动态规划方法求解了有约束 $0 \le u(k) \le x(k)$ 的最优控制问题，这说明该方法能方便地处理有约束的问题，应用面比较广。

【例9-3】 生产库存系统的状态方程为

$$x(k+1)=x(k)+u(k)-S(k)$$

式中 $x(k),u(k),S(k)$ 分别是k周期期初库存量、生产速度和销售速度。

假设生产费用等于 $0.005u^2(k)$，库存费用等于 $x(k)$, 那么四个季度的总费用是

$$J_4=\sum_{k=0}^{3}[0.005u^2(k)+x(k)]$$

现设初始库存量 $x(0)=0$, 四个季度的订货分别为

$S(0)=600$ 件 $S(1)=700$ 件 $S(2)=500$ 件 $S(3)=1200$ 件

生产库存系统的管理问题是：求最优生产速度

$$u^*(0),u^*(1),u^*(2),u^*(3)$$

使 $x(4)=0$（满足销售并到年底没有积压）并且使总费用最小。

解：

第1步 先从最后一个季度考虑起，

$$J_1 = 0.005u^2(3) + x(3)$$

由状态方程和终端条件 $x(4) = 0$，得到

$$x(4) = x(3) + u(3) - S(3) = x(3) + u(3) - 1200 = 0$$

从而得到

$$u^*(3) = 1200 - x(3)$$

将上式代入 J_1，得到

$$J_1^*(x(3)) = 0.005[1200 - x(3)]^2 + x(3) = 7200 - 11x(3) + 0.005x^2(3)$$

第2步 考虑三、四两个季度。这时动态规划的基本方程为

$$J_2^*(x(2)) = \min_{u(2)}\{0.005u^2(2) + x(2) + J_1^*(x(3))\}$$

$$= \min_{u(2)}\{0.005u^2(2) + x(2) + 7200 - 11x(3) + 0.005x^2(3)\}$$

式中 $x(3) = x(2) + u(2) - S(2) = x(2) + u(2) - 500$

把它代入上式得到

$$J_2^*(x(2)) = \min_{u(2)}\{0.005u^2(2) + x(2) + 7200 - 11[x(2) + u(2) - 500]$$

$$+ 0.005[x(2) + u(2) - 500]^2\}$$

$u(2)$ 应使上式右端花括号中的函数取最小值，令

$$\frac{\partial\{.\}}{\partial u(2)} = 0.02u(2) - 16 + 0.01x(2) = 0$$

可解出 $u^*(2) = 800 - 0.5x(2)$

为了保证 $u^*(2) \geq 0$ ，必须 $x(2) \leq 1600$ 。将 $u^*(2)$ 的表示式代入 $J_2^*(x(2))$ ，得

$$\begin{aligned} J_2^*(x(2)) = {} & 0.005[800 - 0.5x(2)]^2 + x(2) + 7200 \\ & -11[x(2) + 800 - 0.5x(2) - 500] \\ & +0.005[x(2) + 800 - 0.5x(2) - 500]^2 = 7550 - 7x(2) + 0.0025x^2(2) \end{aligned}$$

第3步 考虑第二至第四的三个季度。这时动态规划的基本方程为

$$J_3^*(x(1)) = \min_{u(1)}\{0.005u^2(1) + x(1) + J_2^*(x(2))\}$$

$$= \min_{u(1)}\{0.005u^2(1) + x(1) + 7550 - 7x(2) + 0.0025x^2(2)\}$$

式中 $x(2) = x(1) + u(1) - 700$, 将代入 $J_3^*(x(1))$ 得到

$$J_3^*(x(1)) = \min_{u(1)}\{0.005u^2(1) + x(1) + 7550 - 7[x(1) + u(1) - 700]$$

$$+ 0.0025[x(1) + u(1) - 700]^2\}$$

令

$$\frac{\partial\{.\}}{\partial u(1)} = 0.015u(1) - 7 + 0.005[x(1) - 700] = 0$$

可解出

$$u^*(1) = 700 - \frac{1}{3}x(1)$$

将上式代入 $J_3^*(x(1))$ 得到

$$J_3^*(x(1)) = 10000 - 6x(1) + \frac{0.005}{3}x^2(1)$$

第4步 对四个季度一起考虑，这时动态规划的基本方程为

$$J_4^*(x(0)) = \min_{u(0)}\{0.005u^2(0) + x(0) + J_3^*(x(1))\}$$

由于 $x(0)=0,\ x(1)=u(0)-600,$ 上式化为

$$J_4^*(x(0)) = \min_{u(0)}\{0.005u^2(0) + 10000 - 6(u(0)-600) + \frac{0.005}{3}(u(0)-600)^2\}$$

令 $$\frac{\partial\{.\}}{\partial u(0)} = 0$$

得到 $$0.01u(0) - 6 + \frac{0.01}{3}(u(0)-600) = 0$$

解得 $$u^*(0) = 600$$

把它代入 $J_4^*(x(0))$ 中，得到 $J_4^*(x(0)) = 11800.$

于是该生产库存系统的最优管理测略和相应的库存量是

$$x(0)=0 \quad u^*(0)=600 \quad x(1)=0 \quad u^*(1)=700$$

$$x(2)=0 \quad u^*(2)=800 \quad x(3)=300 \quad u^*(3)=900$$

$$x(4)=0$$

在这一管理策略下，总费用为 11800 元。如果每个季度都按订货量安排生产，即 $u(0)=600, u(1)=700, u(2)=500, u(3)=1200,$ 这时每个季度的库存量都是零，那么总费用 $J_4=12700$ 元。比最优策略要多用900元。

从上面三个例子可以看出，由于对多阶决策问题反复应用了动态规划的基本方程，使得一个多阶决策问题分解成一系列的相互嵌套的子问题，对每个子问题只需要解一个决策变量的最优化问题，这样使问题得到简化。

§ 9.3 连续时间系统的动态规划方法

利用动态规划的最优性原理,可以推导出性能泛函为极小应满足的条件—哈密顿-雅可比方程。它是动态规划的连续形式，解此方程可求得最优控制 $u^*(t)$ 。现在来推导这一方程。

设连续系统的状态方程为：$\dot{x} = f(x,u,t), x(t_0) = x_0$

目标函数为：$J(u) = \theta(x(t_f),t_f) + \int_{t_0}^{t_f} L(x,u,t)dt$

求 $u^*(t), u \in U$或任意

最优性原理

最优性原理，如果 $u^*(s),(t \le s \le t_f)$是由时刻 t状态$x(t)$开始到时刻t_f的最优控制，那么$u^*(s),(t+\Delta t \le s \le t_f)$必是由时刻$t+\Delta t$状态 $x(t+\Delta t)$开始到时刻t_f的最优控制

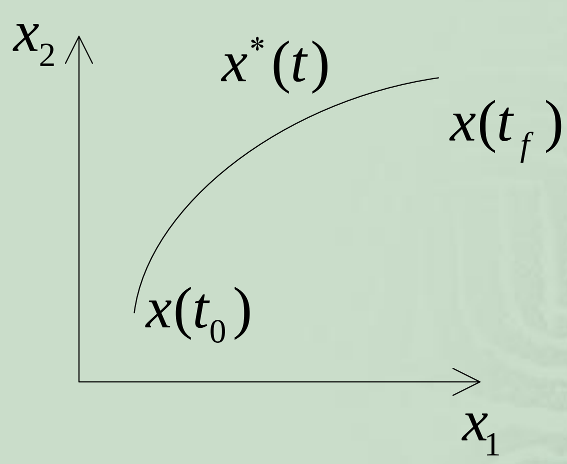

图9-3-1 连续系统的最优轨线

最优性原理

在图9-3-1中，如果 $x^*(t)$是以 $x(t_0)$为初始状态的最优轨迹，设 $t=t'$，状态为 $x(t')$，它将轨线分为前后两段。那么以 $x(t')$ 为初始状态的后半段也必然是最优轨线，而与系统先前是如何到达 $x(t')$无关。

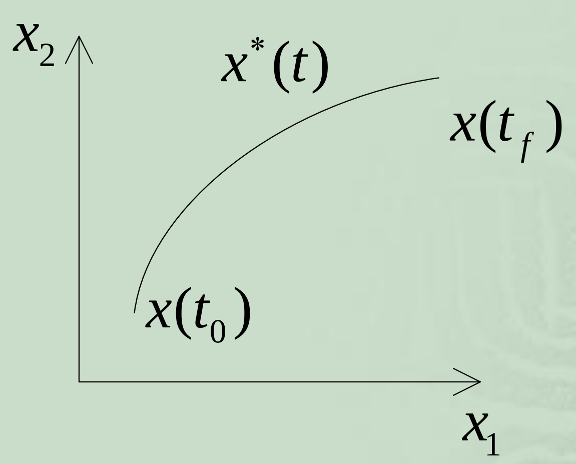

图9-3-1 连续系统的最优轨线

记系统从 t时刻的状态 $x(t)$ 到 t_f 时刻的目标函数的最小值为

$$J^*(x(t),t) = \min_{u\in U}\{\theta(x(t_f),t_f) + \int_t^{t_f} L(x,u,t)dt\}$$

$$= \min_{u\in U}\{\theta(x(t_f),t_f) + \int_t^{t+\Delta t} L(x,u,t)dt + \int_{t+\Delta t}^{t_f} L(x,u,t)dt\}$$

$$= \min_{u\in U}\{\int_t^{t+\Delta t} L(x,u,t)dt + J^*(x(t+\Delta t),t+\Delta t)\}$$

$$\int_t^{t+\Delta t} L(x,u,t)dt = L(x(t+\alpha\Delta t),u(t+\alpha\Delta t),t+\Delta t)\cdot\Delta t$$

$$\approx L(x(t),u(t),t)\cdot\Delta t$$

$$\int_t^{t+\Delta t} L(x,u,t)dt \approx L(x(t),u(t),t)\cdot\Delta t$$

$$x(t+\Delta t) \approx x(t) + \dot{x}(t)\Delta t$$

在$(x(t),t)$邻域做泰勒展开

$$J^*(x(t+\Delta t),t+\Delta t) \approx J^*(x,t) + [\frac{\partial J^*(x,t)}{\partial x}]^T \dot{x}\Delta t + \frac{\partial J^*(x,t)}{\partial t}\Delta t$$

$$J^*(x,t) \approx \min_{u\in U}\{L(x,u,t)\Delta t + J^*(x,t) + [\frac{\partial J^*(x,t)}{\partial x}]^T f(x,u,t)\Delta t + \frac{\partial J^*(x,t)}{\partial t}\Delta t\}$$

由于$J^*(x,t)$, $\frac{\partial J^*(x,t)}{\partial t}$ 与 u无关

$$= J^*(x,t) + \frac{\partial J^*(x,t)}{\partial t}\Delta t + \min_{u\in U}\{L(x,u,t)\Delta t + [\frac{\partial J^*(x,t)}{\partial x}]^T f(x,u,t)\Delta t\}$$

得到：

$$-\frac{\partial J^*(x,t)}{\partial t}=\min_{u\in U}\{L(x,u,t)+[\frac{\partial J^*(x,t)}{\partial x}]^T f(x,u,t)\} \qquad (9.3.1)$$

式（9.3.1）称为连续系统动态规划基本方程或HJB方程。它是一个关于$J^*(x,t)$的偏微分方程。解此方程可求得最优化控制使J达到极小，它的边界条件为

$$J^*(x(t_f),t_f)=\theta(x(t_f),t_f)$$

以下讨论与最大值原理的关系（殊途同归）

如果令哈密顿函数为$H(x,u,\lambda,t)=L(x,u,t)+[\frac{\partial J^*(x,t)}{\partial x}]^T f(x,u,t)$

$$=L(x,u,t)+\lambda^T f(x,u,t)$$

则（9.3.1）式可写成 $-\frac{\partial J^*(x,t)}{\partial t}=\min_{u\in U} H(x,u,\lambda,t) \qquad (9.3.2)$

（9.3.2）称为哈密顿-雅可比方程

下面由贝尔曼方程推导出协状态方程

$$\frac{\partial J^*(x,t)}{\partial t}+L(x,u,t)+[\frac{\partial J^*(x,t)}{\partial x}]^T f(x,u,t)=0$$

对 x 求偏导数，得

$$\frac{\partial^2 J^*(x,t)}{\partial x\partial t}+\frac{\partial L(x,u,t)}{\partial x}+[\frac{\partial J^*(x,t)}{\partial x}]^T \frac{\partial f(x,u,t)}{\partial x}+\frac{\partial^2 J^*(x,t)}{\partial x^2} f(x,u,t)=0$$

由于 $\frac{\partial J^*}{\partial x}$ 对 t 的全导数为 $\frac{d}{dt}[\frac{\partial J^*(x,t)}{\partial x}]=\frac{\partial^2 J^*(x,t)}{\partial x\partial t}+\frac{\partial^2 J^*(x,t)}{\partial x^2}\cdot\frac{dx}{dt}$

$$\frac{d}{dt}[\frac{\partial J^*(x,t)}{\partial x}]+\frac{\partial L(x,u,t)}{\partial x}+(\frac{\partial J^*(x,t)}{\partial x})^T \frac{\partial f(x,u,t)}{\partial x}=0$$

令 $\lambda(t)=\frac{\partial J^*(x,t)}{\partial x}$

$$\frac{d\lambda(t)}{dt}=-\{\frac{\partial L(x,u,t)}{\partial x}+\lambda^T(t)\frac{\partial f(x,u,t)}{\partial x}\}=-\frac{\partial H}{\partial x}$$

这与最大值原理的结果是一致的。

边界上 $J^*(x(t_f),t_f)=\theta(x(t_f),t_f)$

对 $x(t_f)$求偏导数，得

$$\frac{\partial J^*(x(t_f),t_f)}{\partial x(t_f)}=\frac{\partial \theta(x(t_f),t_f)}{\partial x(t_f)}$$

即 $\lambda(t_f)=\dfrac{\partial \theta}{\partial x(t_f)}$

上面的推导过程实际上用动态规划法间接证明了最大值原理。应该指出，与最大值原理比较，动态规划法需要解偏微分方程

$$-\frac{\partial J^*(x,t)}{\partial t}=\min_{u\in U}\{L(x,u,t)+[\frac{\partial J^*(x,t)}{\partial x}]^T f(x,u,t)\}$$

它要求 $J(x,t)$具有连续的偏导数，但在工程实际中，这一点常常不能满足，因此，限制了动态规划的使用范围。

例 设 $\dot{x}=u$ ，求最优控制 $u^*(t)$，使

$$\min_u J=\int_0^{t_f}(x^2+\frac{1}{2}x^4+u^2)dt$$ 达到极小值。

解：动态规划：构造哈密顿函数

$$H=L+\frac{\partial J}{\partial x}f=x^2+\frac{1}{2}x^4+u^2+\frac{\partial J}{\partial x}u$$

根据哈密顿-雅可比方程，有

$$-\frac{\partial J^*}{\partial t}=\min_u H$$

由于u不受限制，得 $\frac{\partial H}{\partial u}=2u+\frac{\partial J^*}{\partial x}=0, u^*=-\frac{1}{2}\frac{\partial J^*}{\partial x}$

$$-\frac{\partial J^*}{\partial t}=x^2+\frac{1}{2}x^4+\frac{1}{4}(\frac{\partial J^*}{\partial x})^2-\frac{1}{2}(\frac{\partial J^*}{\partial x})^2$$

$$=x^2+\frac{1}{2}x^4-\frac{1}{4}(\frac{\partial J^*}{\partial x})^2$$

边界条件 $J^*(x(t_f),t_f)=0$

最大值原理

$$H = L + \lambda^T f = x^2 + \frac{1}{2} x^4 + u^2 + \lambda u$$

$$\frac{\partial H}{\partial u} = 0, 2u + \lambda = 0, u^* = -\frac{1}{2} \lambda$$

正则方程：

$$\begin{cases} \dot{x} = u = -\frac{1}{2} \lambda \\ \dot{\lambda} = -\frac{\partial H}{\partial x} = -(2x + 2x^3) \end{cases}$$

边界条件：

$$\begin{cases} x(0) = x_0 \\ \lambda(t_f) = 0 \end{cases}$$

连续时间系统利用动态规划求解最优控制问题的步骤。

1.构造哈密顿函数 $H(x,u,t)=L(x,u,t)+[\frac{\partial J^*}{\partial x}]^T f(x,u,t)$

2.以 $H(x,u,t)$取极值得到 u^* ，即

$$\frac{\partial H(x,u,t)}{\partial u}=0$$ （当u取值无约束）

$$\min_{u\in U} H(x,u,t)$$ （当$u\in U$为容许控制时）

由上述条件得到的u^*是 $x,\frac{\partial J^*}{\partial x},t$ 的函数

3.将u^*代入哈密顿-雅可比方程，并根据边界条件。求 $J^*(x(t),t)$

4.将 $J^*(x(t),t)$ 代入$u^*(t)$

变分法，最大值原理，动态规划三者间的关系

- 三种方法的比较：
  - 1.古典变分法不能处理闭集性约束
  - 2.最大值原理扩展了古典变分法，把古典变分法概括为自己的特殊情况。古典变分法和最大值原理都得到了一组常微分方程表示的必要条件。
  - 3.动态规划法从另一方面发展了变分法。对于连续系统的最优化问题，动态规划法给出了一个偏微分方程条件。虽然，一般来说，解偏微分方程比解常微分方程更难，但结果是充分条件
- 这三种方法也互有联系，它们在哈密顿函数上统一了起来。