最优控制计算中的微分动态规划(DDP)方法

孙茂相

（沈阳机电学院）

〔提要〕 本文介绍一种在最优控制计算中十分有效的算法：微分动态规划(DDP)方法。叙述了其原理，指出DDP在两方面所遇到的困难：收敛性的证明与处理纯状态变量约束问题，讨论了有关算法收敛性的研究成果，最后通过几个例子说明DDP在不同方面的应用。

一　问题提出

最优控制理论产生于50年代中期，以贝尔曼（Bellman）的动态规划与庞特里雅金(Понтрягцн）的最大原理为基础。在数学上，这一理论可表述为如下形式：

设系统动态方程为

$$\dot{x}=f(x,u;t)\quad x(t_0)=x_0\quad t\in[t_0,t_f] \tag{1.1}$$

x为动态系统的状态向量，维数为n；u为控制向量，维数为m；f为n维非线性向量函数。

系统的性能指标为

$$V(x_0;t_0)=\int_{t_0}^{t_f}L(x,u;t)dt+F(x(t_f);t_f) \tag{1.2}$$

L与F为纯量非线性函数，x与u可能还要满足下列约束条件

$$g(u;t)\leqslant 0 \tag{1.3}$$

$$\phi_1(x,u;t)\leqslant 0 \tag{1.4}$$

$$\phi_2(x;t)\leqslant 0 \tag{1.5}$$

$$\psi(x(t_f);t_f)=0 \tag{1.6}$$

其中g，ϕ_1，ϕ_2，ψ皆为非线性向量函数，f，L，F，g，ϕ_1，ϕ_2与ψ皆为三次可微。

控制问题的目的在于确定$u(t)$，$t\in[t_0,t_f]$，使V最小，并满足给定的约束条件。

上述问题出现在许多领域中，例如航天技术、军事、工业、经济系统、社会系统等等。由于最优控制问题的复杂性，一般不存在解析解，因此一些利用数值方法计算最优控制问题寻求最优解的最优控制计算方法便应运而生。

最优控制计算方法几乎与最优控制理论同时产生，它是将最优控制理论与最优化方法结合起来，以进行最优控制问题的数值计算。文献〔1〕，〔2〕，〔4〕是这一领域的代表著作，其中给出了许多实用算法，可以直接用于计算机计算，文献〔3〕回顾了最优控制计算方法的发展历史，有助于对这一领域的全面了解。

在最优控制的计算方法中，最著名，最引人注目者，当推微分动态规划（简称DDP）方法。DDP方法是由Jacobson与Mayne于1969年提出。由于这种方法在数值计算方面有种种优越性，因而很有吸引力。本文目的在于介绍DDP方法的基本原理，讨论其收敛性研究的结果，最后给出几个应用例子，以期引起读者的兴趣。

二　DDP 原理概述

DDP是一种逐次逼近方法，它以动态规划为基础，用于确定非线性系统的最优控制。为简单起见，我们仅考虑无约束情况，但其基本思路与有约束情况一致，见〔4〕，〔8〕。

我们知道，最优性能指标满足下列Bellman偏微分方程。

$$-(\partial V/\partial t)(x;t)=\min_{u}[L(x,u;t)+\langle V_x(x;t),f(x,u;t)\rangle] \tag{1.7}$$

设最优控制$u^\circ(t)$；$t\in[t_0,t_f]$是未知的，但标称控制$\overline{u}(t)$可以得到。根据方程(1.1)由$\overline{u}(t)$可得标称状态轨线$\overline{x}(t)$；$t\in[t_0,t_f]$，由(1.2)式可以计算标称性能指标。

方程（1.1），（1.2）与（1.7）可以通过标称轨线写出，这时令

$$x=\bar{x}+\delta x \tag{1.8}$$

$$u=\bar{u}+\delta u \tag{1.9}$$

在 $(\bar{x};t)$处，最优性能指标为 $\overline{V}(\bar{x};t)$

$$V(x;t)=\overline{V}(\bar{x};t)+a(\bar{x};t) \tag{1.10}$$

其中a定义为在 $(\bar{x};t)$处最优性能指标与标称性能指标之差。

引进 Hamilton函数

$$H(x,u,V_x;t)=L(x,u;t)+\langle V_x,f(x,u;t)\rangle \tag{1.11}$$

则有

$$-\frac{\partial\overline{V}}{\partial t}-\frac{\partial a}{\partial t}-\left\langle\frac{\partial V_x}{\partial t},\delta x\right\rangle-\frac{1}{2}\left(\delta x,\frac{\partial V_{xx}}{\partial t}\delta x\right)=\min_{\delta u}[H(\bar{x}+\delta x,\bar{u}+\delta u,V_x;t)+\langle V_{xx}\delta x+\tfrac{1}{2}V_{xxx}\delta x\delta_x,f(\bar{x}+\delta x,\bar{u}+\delta u;t)\rangle] \tag{1.12}$$

上式为在标称轨线邻域内的方程，是DDP中的基本公式，由此可以导出DDP算法中所需要的各种关系。经过简单推导，可得下列方程：

$$-\dot{a}=H-H(\bar{x},\bar{u},V_x;t) \tag{1.13}$$

$$-\dot{V}_x=H_x+\beta^T H_u^{-0}+V_{xx}(f-f(\bar{x},\bar{u};t)) \tag{1.14}$$

$$-\dot{V}_{xx}=H_{xx}+f_x^tV_{xx}+V_{xx}f_x-(H_{ux}+fu^TV_{xx})^T(H_{uu}^{-1}(H_{ux}+f_u^TV_{xx}) \tag{1.15}$$

在$t=t_f$时，有

$$V(\bar{x};t_f)=F(\bar{x}(t_f);t_f) \tag{1.16}$$

$$a(t_f)=0 \tag{1.17}$$

$$V_x(t_f)=F_x(\bar{x}(t_f);t_f) \tag{1.18}$$

$$V_{xx}(t_f)=F_{xx}(\bar{x}(t_f);t_f) \tag{1.19}$$

上式中，H_u^{-0}表示$H_u(\bar{x},u^*,V_x;t)=0$，其中u^*为使H最小时的u，并忽略V_{xxx}项。

有了这些关系式，我们便可以进行计算，计算大意如下：首先选择一标称控制$\bar{u}(t)$，$t\in[t_0,t_f]$（可猜试），由(1.1)式计算对应的标称轨线$\bar{x}(t)$，由(1.2)式计算对应的$\overline{V}(x_0,t_0)$。然后DDP进行反向与正向运算，在反向计算中，利用边界条件（1.17）到（1.19）从t_f到t_0对(1.13）到(1.15）反向积分，得到一个将$u(t)$表示为$x(t)$的线性函数的控制律。这个控制律是由一个$\bar{x}$与u的二次函数H对u求最小值而得到的，$u(t)=u^*(t)+\beta\delta x(t)$。在以后的正向积分中，应用这一控制律计算新的状态轨线x与对应的性能指标$V(x_0;t_0)$，如果结果得到改进，则新的控制将成为下一次DDP迭代中的标称控制。如果结果没有得到改善，则应用“步长调正法”。计算重复进行，直到产生一个充分逼近于最优控制函数的控制函数为止。

DDP可以处理（1.3），（1.4）与(1.6）型的约束，但是在处理含有型如（1.5）的纯状态变量约束时，十分困难，Mayne在[8]中曾有说明。我们将在另文中叙述一种能够处理纯状态变量约束的改进型DDP算法。

DDP方法最初提出时，其收敛性并没有得到证明。这是一个非常困难的课题，我们在下节介绍这方面取得的成果。

三 DDP算法收敛性研究

1975年，Polak与Mayne[5,6]研究了DDP型的一阶算法，用于求解具有控制不等式约束的连续最优控制问题。他们研究的问题可表示为：

动态系统方程为

$$\dot{x}(t)=f(x(t),u(t);t) \quad x(0)=x_0 \tag{3.1}$$

其中$x(t)\in R^n,u(t)\in R^m$，　$t\triangleq T\in[0,1]$

$$S=\{x,u;t\,|\,x\in R^n,u\in\Omega,t\in T\}$$

性能指标为

$$V(u)=\int_0^1 L[x^u(t),u(t);t]dt+F[x^u(1)] \tag{3.2}$$

其中x^u为由u与初始条件$(x_0,0)$所得到的(3.1)的解。

Polak与Mayne证明了其算法的局部收敛性。然而他们的算法是以最大原理为基础，与[4]中的一阶DDP算法也不完全相同。而且我

们知道，DDP算法通常是二阶。

1978年Ohno[7]提出一种新型DDP算法。他研究了如下问题：

动态系统方程为

$$x_{n+1}=f_n(x_n,u_n)\quad n=0,\cdots,N-1,\quad x_0=0 \tag{3.3}$$

性能指标为

$$V=\sum_{n=0}^{N-1}L_n(x_n,u_n)+L_N(x_N) \tag{3.4}$$

约束条件为

$$g_n(x_n,u_n)\leqslant 0\qquad h_n(x_n,u_n)=0 \tag{3.5}$$

Ohno证明了其算法的局部收敛性。然而这种算法是以Kuhn-Tucker条件为基础，与〔4〕中的DDP算法有很大差别。

1980年，Sakawa等[9]研究了下列问题：

动态系统方程为

$$\dot{x}(t)=f(x(t),u(t),t),x(t_0)=x_0 \tag{3.6}$$

性能指标为

$$V(u)=\int_{t_0}^{t_1}L[x(t),u(t);t]dt \tag{3.7}$$

约束条件为

$$u(t)\in U\qquad t\in[t_0,t_1]$$

他们提出了一种算法，该算法类似于一阶DDP但基本上以最大原理为基础。在算法中应用了Jarmark提出的收敛控制方法，并研究了算法的全局收敛条件。

四 DDP算法的应用

1. 火箭控制问题[4]

Dreyfus研究了下列问题，在固定时间将火箭发射到给定高度，最终垂直速度分量给定，使最终水平速度分量最大。设控制u为火箭推力对水平方向的倾角，现在要选择u，使最终水平速度分量在给定约束条件下最大。应用DDP解这一问题，6次迭代即得到所需结果，在IBM7090计算上用了一分钟时间。Dreyfus用一阶方法解这一问题，达到同样结果需29次迭代与三分钟计算时间。

2. 在非线性参数测辨中的应用[10]

解非线性参数测辨问题需要用迭代方法。固定区间测辨可以化为使一个适当的性能测量指标在动态约束下的最小化问题。这与最优控制问题等价，可以用DDP对这一问题求解。评价一种方法的优劣，关键在于比较其参数估计。常用的方法为非线性最小二乘法，即所谓输出误差法(OEM)，其中不考虑过程噪声。与OEM相比，DDP在几乎全部实验系统中，都得到重要的改进估计 此外，DDP在加入Jarmark的收敛控制原理之后，保证了算法的良好收敛性质，这对于有效的数值计算是十分必要的。

3. 多级水库联合运行问题[11]

用其他方法计算水库的联合运行，最多能做到四水库，而用DDP，可以轻而易举地处理十水库联合运行，DDP可以计算40维状态向量与控制向量的最优控制问题。

五 结论

DDP方法自1969年问世以来，已有10余年历史。由于在数值计算方面的种种优越性，如在LQP问题中，表现了一步收敛的性质，而在非LQP中，在最优轨线的邻域，收敛迅速，算法中包含步长调正方法，节省计算机的计算时间等等，而得到广泛应用。DDP方法既可用于连续系统，也可用于离散系统。关于算法收敛性的证明，至今还没有得到令人满意的结果。

参考文献

〔1〕Bryson, A.E.and Y.-C.Ho., Applied Optimal Control, John Wiley and Sons, New York, 1975.

〔2〕Dyer, P. and McReynolds, S.R., The Computation and Theory of Optimal Control, Academic Press, New York, 1970.

〔3〕Polak, E., An Historical Survey of Computational Methods in Optimal Control, SIAM Review, Vol.15, No.2, pp.553—584, 1973.

〔4〕Jacobson, D.H.and Mayne,D.Q., Differential Dynamic Programming, American Elsevier, New York, 1970.

〔5〕Mayne,D.Q.and Polak,E., First-Order Strong Variation Algorithms for Optimal Control, J.

Opt.Theory Appl.,Vol.16,pp.277—301,1975.
〔6〕Polak,E.and Mayne, D.Q., First-Order Strong Variation Algorithms for Optimal Control Problems with Terminal Inequality Constraints,J. Opt.Theory Appl.,Vol.16,pp.303—325,1975.
〔7〕Ohno, K., A New Approach to Differential Dynamic Programming for Discrete Time System, IEEE Trans.on Automatic Control,AC—23,pp. 37—47,1978.
〔8〕Mayne,D.Q.,Differential Dynamic Programming —A Unified Approach to the Optimization of Dynamic Systems, Control and Dynamic Systems,Vol.10,Academic Press,New York,1973.
〔9〕Sakawa,Y.and Shindo,Y.,On Global Convergence of an Algorithm for Optimal Control, IEEE Trans.on Automatic Control, AC—25, pp.1149—1153,1980.
〔10〕Jonsson,H-O., A Differential Dynamic Programming Approach to Nonlinear Parameter Identification, Internal Report.LITH-ISY-I-0606, Linkoping University,1983.
〔11〕Murray,D.M.and Yakowitz, S.J., Constrained Differential Dynamic Programming,Water Resources Research,Vol.15, pp.1017—1027,1979.
〔12〕Jarmark,B.,Convergence Control in Differential Dynamic Programming Applied to Air-to-Air Combat, AIAA J., 14,118—121, 1976.