Dynamics and Control 3, 175–185 (1993)

Differential Dynamic Programming Applied to Continuous Optimal Control Problems with State Variable Inequality Constraints

DAVID J.W. RUXTON
Department of Mathematics and Computing, University of Central Queensland, Australia

Received April 24, 1991; revised December 5, 1991 and June 4, 1992
Editor: R. Stonier

Abstract. The purpose of this article is to show that the differential dynamic programming (DDP) algorithm may be readily adapted to cater for state inequality constrained continuous optimal control problems. In particular, a new approach using a multiplier penalty function scheme incorporated with the DDP algorithm is shown to be effective. The DDP algorithm, implemented in conjunction with a multiplier penalty function scheme, is compared to an established DDP algorithm variant and the gradient-restoration method.

1. Introduction

The differential dynamic programming (DDP) algorithm developed by Jacobson and Mayne [1] does not cater directly for state variable inequality constraints (SVICs).

DDP algorithm variants, specifically designed to accommodate the continuous optimal control problem subject to constraints of SVIC form, have been developed by Martensson [2], Jarmark [3], and Tun and Dillon [4]. Although these different DDP algorithm variants can give encouraging results, they all involve increased and often considerable implementation and computational effort, particularly for highly nonlinear problems.

The optimal control problem with constraints of SVIC form has been extensively researched [5–9]. Although a number of numerical methods based on this work have emerged, including the DDP variant of Tun and Dillon [4], schemes using penalty functions remain a popular alternative [3, 4, 9–13]. Penalty function schemes are known to have stability and accuracy problems, but their flexibility and ease of implementation make them worthy of continued interest.

Apart from work by Jarmark [3], there appears to have been scant interest, or at least little reported, regarding the use of penalty function schemes with DDP applied to the SVIC problem. Chen and Chang et al. [14, 15] have recently reported on the effectiveness of incorporating a multiplier penalty function scheme, as first proposed by Bertsekas [16], with the DDP algorithm applied to *discrete* SVIC optimal control problems.

The intent in this article is to demonstrate the ease of implementation and effectiveness of penalty function schemes when combined with the DDP algorithm and applied to *continuous* SVIC optimal control problems. In particular, and following the lead of Chen and

Chang et al. [14, 15], a new DDP algorithm variant using a multiplier penalty function scheme is shown to compare favorably with an established DDP algorithm variant and the gradient-restoration algorithm.

2. Problem statement and notation

Solution of the continuous optimal control problem involves finding a control function $u(t)$ that minimizes a cost functional or performance index

$$V(x_0, t_0) = \int_{t_0}^{t_f} L(x, u; t)\, dt + F(x(t_f); t_f) \tag{1}$$

subject to dynamic system constraints described by a set of ordinary differential equations

$$\dot{x} = f(x, u; t); \qquad x(t_0) = x_0, \qquad x(t_f) = x_{t_f}, \tag{2}$$

where

$x(t)$ is an n-dimensional vector function of time that describes the state of the dynamic system at any time $t \in [t_0, t_f]$;

$u(t)$ is an m-dimensional vector function of time that describes the control available for adjustment at any time $t \in [t_0, t_f]$;

$f(x, u; t)$ is an n-dimensional vector function that describes the dynamical structure of the system. The notation adopted means that f is a function of $x(t)$ and $u(t)$ explicitly and possibly also of time explicitly;

L and F are scalar functions of their arguments; and

V is a scalar that describes the value of some performance index or cost functional associated with the dynamical system.

The object of the SVIC optimal control problem is to solve equation (2) for a control $u(t)$; $t \in [t_0, t_f]$, such that the cost functional given by equation (1) is minimized and the following SVIC is satisfied:

$$g(x; t) \leq 0 \ \forall \ t \in [t_0, t_f]. \tag{3}$$

3. DDP and penalty function schemes

A full description of the DDP algorithm can be found in Jacobson and Mayne [1]. Here only those sections of the algorithm that need significant alteration in order to accommodate penalty function schemes are discussed.

The essence of the DDP method lies in the backward integration of the following matrix differential equations:

$$-\dot{V}_x = H_x + V_{xx}(f - f(\bar{x}, \bar{u}; t)),$$
$$-\dot{V}_{xx} = H_{xx} + f_x^T V_{xx} + V_{xx} f_x - (H_{ux} + f_u^T V_{xx}) H_{uu}^{-1} (H_{ux} + f_u^T V_{xx}).$$

The Hamiltonian H is defined in the usual manner as

$$H = H(x, u, V_x; t) = L(x, u; t) + < V_x, f(x, u; t) >.$$

In addition to requiring partial derivatives of H with respect to state and control variables, the DDP algorithm also requires a complete minimization of H with respect to control at each time step in an iteration. This minimization of H is used to calculate improved controls between iterations.

It will be shown that exterior quadratic penalty function (EQPF) an multiplier penalty function (MPF) schemes require simple modifications to the form of the Hamiltonian and some minor modifications to the control of flow in the DDP algorithm. This is in marked contrast when compared with other DDP algorithm variants for the SVIC problem.

4. A DDP/EQPF scheme

The augmented integrand of the cost functional in equation (1) becomes

$$L^*(x, u, w; t) = L(x, u; t) + w \sum_{i=1}^{n} A_i g_i^2(x; t),$$

where $A_i = 1$ if the $g_i > 0$, $A_i = 0$ otherwise, and n is the number of constraints g.

This results in an augmented Hamiltonian:

$$H^*(x, u, V_x, w; t) = H(x, u, V_x; t) + w \sum_{i=1}^{n} A_i g_i^2(x; t).$$

Repeated iteration of the DDP algorithm, using H^* in place of H and using a high fixed value for, or successively increasing values of, the penalty weighting parameter w between iterations, usually results in satisfaction of the constraints of SVIC form to within some specified tolerance.

The only required modifications to the DDP algorithm are in the backward integration phase and involve simple changes to the partial derivatives of the augmented Hamiltonian with respect to the state variables. The incorporation of a weighting parameter updating scheme is a minor addition to the main flow of the algorithm. That is,

$$H^* = H + w \sum_{i=1}^{n} A_i g^2, \qquad H_x^* = H_x + 2w \sum_{i=1}^{n} A_i g g_x$$

and

$$H_{xx}^{*} = H_{xx} + 2w \sum_{i=1}^{n} A_i(gg_{xx} + g_x^2),$$

where $g = g_i(x; t)$ is used for notational convenience.

5. DDP and a MPF scheme—the new approach

Essentially, the MPF method merges the penalty function idea with a primal-dual Lagrange multiplier scheme, producing an algorithm that largely overcomes the ill conditioning and rate-of-convergence difficulties associated with EQPF schemes.

Bertsekas [16, 17], has analyzed the use of multiplier methods for general contrained minimization problems and concluded that they have significant advantages over traditional penalty function methods in terms of reliability and speed of convergence. In particular, he points out that convergence for the method of multipliers can usually be obtained without the need to increase the penalty weighting parameter to the point where ill conditioning becomes a problem. He notes that most researchers concur that multiplier penalty functions are the best method available for problems with nonlinear constraints in the absence of special structures that might otherwise be exploited. He suggests that MPF schemes are suitable for high-dimensional problems with multiple constraints, such as nontrivial constrained optimal control problems.

The convergence rate for MPF schemes applied to general constrained optimization problems can be shown to be linear or better for convex problems, but the same restriction exists as for penalty function techniques in the case of nonconvex problems: the solution of the primal problem may not be optimal unless the minimizaton of the dual is indeed global.

Chen and Chang et al. [14, 15] recently used ideas suggested by Bertsekas [17] to incorporate the MPF method with the DDP algorithm, giving encouraging results for discrete optimal control problems subject to constraints of SVIC form. Proceeding along similar lines, this article shows that the DDP algorithm is readily modified to incorporate the MPF method in order to effectively deal with continuous optimal control problems subject to constraints of the SVIC form.

5.1. Dual-problem formulation

The constrained optimal control problem specified by equations (1), (2), and (3) involves two sets of constraints: the system dynamics as specified by equation (2) and the explicit constraint on state given by equation (3). A dual problem may be constructed, relaxing all but the constraints on the system dynamics, by applying a MPF scheme along the lines proposed by Bertsekas [17] as follows.

Let

$$\alpha_j(x, u; t) = wg_j(x, u; t),$$

where w is a weighting parameter associated with a series of constriants $\{g_j\}$.

Note that x and u are functions of time with $g_j(x, u; t) = g_j(x)$ if the constraint does not involve u or t explicitly.

$P(\alpha, \mu; t)$ is introduced as a (yet to be fully defined) MPF associated with the constraints $g_j(x, u; t)$ as follows:

$$P(\alpha, \mu; t) = \sum_{j=1}^{n} P_j(\alpha_j, \mu_j; t),$$

where $\mu_j(t)$ is a multiplier and n is the total number of constraints.

The integrand of the cost functional given by equation (1) is then modified to accommodate this MPF, thus forming an augmented cost integrand

$$L^*(x, u, \alpha, \mu; t) = L(x, u; t) + 1/wP(\alpha, \mu; t).$$

The dual problem corresponding to equations (1), (2), and (3) is then

$$\max_u D(\alpha, \mu; t), \tag{4}$$

with

$$D(\alpha, \mu; t) = \min_u \int_{t_0}^{t_f} L^*(x, u, \alpha, \mu; t)\, dt + F(x(t_f), t_f) \tag{5}$$

subject to equation (2).

Solution of this dual problem, and thus also of the constrained primal problem, is simply a matter of using a numerical method, suitable for unconstrained optimal control problems, to minimize equation (5). This algorithm would start with nominal values for $\mu(t)$ and w and then iteratively generate improved values until both equations (4) and (5) are satisfied to the required accuracy. In other words, equations (4) and (5) are alternated in an iteration sequence until convergence.

The updated multipliers at each unconstrained problem iteration sequence are generated for each of the j constraints from

$$\mu^{i+1}(t) = \nabla_\alpha P(\alpha, \mu^i; t)$$
$$\alpha_j = w^i g_j[x(\mu^i, w^i; t), u(\mu^i, w^i; t); t],$$

where $\nabla_\alpha = \partial P/\partial \alpha_j$, i is an iteration index and $\{w^i\}$ is a positive nondecreasing sequence.

5.2. *Properties of a combined DDP/MPF algorithm*

Chen and Chang et al. [14, 15] point out that an algorithm combining DDP and MPF methods inherits many of the desirable properties of both, namely:

- Convergence of the DDP/MPF algorithm to at least a local minimum of equation (5) is quadratic (at least in the discrete case), provided certain smoothness and convexity assumptions apply;
- The convergence rate for the multipliers $\mu(t)$ is at least linear, provided both the primal and dual problems are convex [17];
- The selection of the MPF $P(\alpha, \mu; t)$ can drastically affect the rate of convergence and the general performance of the algorithm [17].

5.3. *Selection of a multiplier penalty function*

The following MPF form, which is both smooth and numerically stable, was suggested by Bertsekas [17], and then subsequently modified and used by Chen and Chang et al. [14, 15] in solving discrete SVIC problems:

$$\begin{aligned} P(\alpha, \mu; t) &= \mu\alpha + \mu\alpha^2, \qquad \text{for } \alpha \geq 0 \\ &= \frac{\mu\alpha}{1 - \alpha}, \qquad \text{for } \alpha < 0. \end{aligned} \tag{6}$$

5.4. *The MPF scheme and DDP*

In what follows it is assumed, for notational simplicity and clarity, that only one SVIC is active. However, the rationale extends to the case of multiple constraints.

The DDP algorithm is readily modified to accommodate the SVIC problem by incorporation the MPF given by equation (6). The augmented cost integrand becomes either:

1. **Constraint active**, that is, $g(x) \geq 0$

$$\begin{aligned} P(\alpha, \mu; t) &= \mu\alpha + \mu\alpha^2, \\ L^* &= L + \mu g + \mu w g^2 \end{aligned}$$

(note the similarity of the last term to the exterior quadratic penalty form); or

2. **Constraint inactive**, that is, $g(x) < 0$

$$\begin{aligned} P(\alpha, \mu; t) &= \mu\alpha(1 - \alpha)^{-1}, \\ L^* &= L + \mu g(1 - wg)^{-1}. \end{aligned}$$

These modifications result in an augmented Hamiltonian

$$H^* = H + 1/wP(\alpha, \mu; t)$$
$$= H + A\mu g(1 + wg) + B\mu g(1 - wg)^{-1},$$

where $g = g(x)$ and $A = 1$, $B = 0$ for $g \geq 0$ or $A = 0$, $B = 1$ for $g < 0$.

Successive iterations of the DDP algorithm using H^* in place of H will, under convexity assumptions already noted, converge to a solution of the relaxed problem given by equation (5) for current values of $\mu(t)$ and w.

As in the case for the EQPF method, the only modifications required in the backward integration phase of the DDP algorithm are changes to the partial derivatives of the Hamiltonian. These partial derivatives become

$$H_x^* = H_x + A\mu(g_x + 2wgg_x) + B\mu g_x(1 - wg)^{-2},$$
$$H_{xx}^* = H_{xx} + A\mu(g_{xx} + 2w(gg_{xx} + g_x^2)) + B\mu(g_{xx} - wgg_{xx} + 2wg_x^2))(1 - wg)^{-3}.$$

The DDP algorithm is readily modified to update multipliers and penalty function weights between iterations as follows:

$$\mu^{i+1}(t) = \frac{\partial P}{\partial \alpha}(\alpha, \mu^i; t) = \mu^i(t)(1 + 2w^i g), \; g \geq 0,$$
$$= \mu^i(t)(1 - w^i g)^{-2}, \; g < 0,$$

and $w^{i+1} = a.w^i$, $a > 0$ and $a = 1$ for a fixed weighting scheme.

6. Comparisons

The following "evergreen" problem, considered by many researchers including [2, 5, 9, 18–20], is chosen as a convenient basis for comparing select algorithms. The problem is to minimize

$$V(x_0, t_0) = \int_0^1 u^2 \, dt$$

subject to

$$\dot{x}_1 = x_2 \; ; \; x_1(0) = 0 \; ; \; x_1(1) = 0$$
$$\dot{x}_2 = u \; ; \; x_2(0) = 1 \; ; \; x_2(1) = -1$$
$$g(x_1) = x_1 - p \leq 0, \; p > 0$$

where in comparisons both $p = 0.15$ and $p = 1/9$ are considered.

6.1. Case 1: $p = 1/9$

Results of applying the DDP/EQPF, DDP/MPF and Martensson's DDP algorithm variants to the problem with $p = 1/9$ are presented in table 1.

When interpreting these results, the following points should be noted:

- The mnemonics pf, m, and mts refer to the EQPF, MPF, and Martenson's DDP algorithm variants, respectively;
- The optimal cost with $p = 1/9$ is known to be 8.0000 [19];
- In this case constraint boundary adhesion for $x_1(t)$ and $u(t)$ can only be estimated qualitatively and relative to the solution presented by Martensson, where he makes the point that the EQPF method yields relatively soft adhesion even for high values of the weighting parameter. Because Martensson does not publish full trajectory details, the measure of constraint boundary adhesion used is Δx_1, which gives the maximum deviation of the $x_1(t)$ trajectory from the optimal solution;
- Martensson's algorithm variant involves significantly more computational effort than the others because it involves transforming the constraint from SVIC to another form using a hyperplane conversion technique and then the application of a DDP algorithm variant with substantially extended internal structure.

6.2. Case 2: $p = 0.15$

Results of applying the DDP/EQPF, DDP/MPF algorithm variants and the gradient-restoration algorithm to the problem with $p = 0.15$ are presented in tables 2 and 3.

Table 1. Summary of results with $p = 1/9$.

Method	Cost	Its.	$max\ \Delta x_1$	w	a	μ	CB adhesion
pf	7.9932	24	0.0002	10^7	–	–	Soft
m	7.9996	22	0.00004	90	5	50	Hard
mts	8.0008	18	0.00003	–	–	–	Hard

Table 2. State and control trajectories with $p = 0.15$.

t	x_1^{gr}	$x_1^{pf\&m}$	u^{gr}	u^{pf}	u_{16}^{m}	u_{18}^{m}	x_1^{opt}	u^{opt}
0.4	0.1496	0.1498	-0.4730	-0.4954	-0.4982	-0.4961	0.1498	-0.4938
0.5	0.1499	0.1500	+0.0375	-0.0514	-0.0416	-0.0206	0.1500	0.0000
0.6	0.1497	0.1498	-0.4405	-0.4938	-0.4923	-0.4940	0.1498	-0.4938

Table 3. Summary of results with $p = 0.15$.

Method	Cost	Its.	$E(u)$	w	a	μ
gr	5.927	16	0.019	–	–	–
pf	5.925	13	0.007	10^4 to 10^6	–	–
m_{16}	5.926	16	0.006	100	5	50
m_{18}	5.926	18	0.004	100	5	50

When interpreting these results, the following points should be noted:

- Gradient-restoration (gr) results are as published by Miele [18];
- The optimal cost with $p = 0.15$ is known to be 5.926 [19];
- Results for the MPF scheme are given for both 16 and 18 iterations;
- Note especially that $u^{gr} > 0$ at $t = 0.5$, whereas u^{m} and $u^{opt} < 0 \; \forall \; t \in [0, 1]$;
- Constraint satisfaction by either state or control trajectories may be measured by

$$E(p) = \frac{\sum_{i=1}^{n} | (p_a - p_c) |}{n}$$

where p_a is the actual optimal value of the trajectory and p_c is the computed value of the trajectory, measured for the n discrete time steps used over the entire interval [0, 1]. Here $E(u)$ is considered, rather than the more natural $E(x_1)$, for convenience in matching and presenting results from the various methods.

- The gradient-restoration algorithm involves transformation of the original problem into one of higher dimension. Although the first-order gradient-restoration algorithm is computationally more efficient, being of $O(n^2)$, than the second-order DDP algorithm, at $O(n^3)$, it has more work to do for a given problem, since the state-space dimension n increases according to the number of inequality constraints present. Thus, at least for problems of low dimension, the two algorithms have similar work to do per iteration.

7. Conclusions

The results for the single problem considered here are indicative of those obtained by Ruxton [20, 21], where the DDP/MPF algorithm variant is tested on a number of nonlinear problems and a problem involving an SVIC that is nonlinear and time dependent.

It would seem that both the DDP/MPF and DDP/EQPF algorithm variants give more accurate results, and especially better constraint boundary adhesion, than the gradient-restoration algorithm, albeit for increased computational effort. Furthermore, the DDP/MPF algorithm appears to produce significantly better constraint boundary adhesion than the DDP/EQPF algorithm. This is to be expected given the modus operandi of the two schemes. For example, it is a simple matter to show that the Hamiltonian is continuous across the constraint boundary $g(x_1) = 0$ for the DDP/MPF formulation, whereas for the DDP/EQPF formulation this is not the case.

To date, algorithm variants based on Martensson's work have been the most effective means of dealing with SVIC problems when using the DDP method. The results presented here demonstrate that the DDP/MPF algorithm is easier to implement and is as accurate as Martensson's DDP algorithm.

The DDP method has useful and largely unrecognized potential when combined with penalty function schemes. Avenues for further research include

- Investigating the effectiveness of different forms of the MPF given by equation (6);
- Adapting the DDP/MPF variant for more demanding practical problems, especially those with highly nonlinear dynamics and possibly those involving singular arcs.

Acknowledgment

The author acknowledges the encouragement of Dr. R.J. Stonier of the Department of Mathematics and Computing, University of Central Queensland, Australia.

References

1. D.H. Jacobson, and D.Q. Mayne, *Differential Dynamic Programming*. American Elsevier: New York, 1970.
2. K. Martensson, "New approaches to the numerical soluton of optimal control problems," Studentlitteratur (Ph.D. thesis), Lund, Sweden, 1972.
3. B. Jarmark, "Calculation aspects on an optimization program," School of Electrical Engineering, Chalmers University of Technology, Goteborg, Sweden, Report R 82-02, March 1982.
4. T. Tun and T.S. Dillon, "Extensions of the differential dynamic programming method to include systems with state dependent control constraints and state variable inequality constraints," *J. Appl. Sci. Eng. A*, vol. 3, pp. 171–192, 1978.
5. A.E. Bryson, Jr., W.F. Denham and S.E. Dreyfus, "Optimal programming problems with inequality constraints I: necessary conditions for extremal solutions," *AIAA J.*, vol. 1, no. 11, pp. 2544–2550, November 1963.
6. D.H. Jacobson, M.M. Lele and J.L. Speyer, "New necessary conditions of optimality for control problems with state-variable inequality constraints," *J. Math. Anal. Appl.*, vol. 35, pp. 255–284, 1971.
7. J. McIntyre and B. Paiewonsky, "On optimal control with bounded state variables," in C.T. Leondes (ed.), *Advances in Control Systems*, vol. 5. Academic Press: New York, 1967.
8. J.L. Speyer and A.E. Bryson, Jr., "Optimal programming problems with a bounded state space," *AIAA J.*, vol. 6, no. 8, pp. 1488–1491, August 1968.
9. L.S. Lasdon, A.D. Warren and R.K. Rice, "An interior penalty method for inequality constrained optimal control problems," *IEEE Trans. Auto. Control*, vol. AC-12, no. 4, pp. 388–395, August 1967.
10. R.V. Mayorga, V.H. Quintana and V. Gerez, "A numerical solution for state constrained continuous optimal control problems using improved penalty functions," in *Proc. 22nd IEEE Conf. Decision Control*, San Antonio, TX, vol. 1, pp. 432–434, December 1983.
11. M.M. Lele and D.H. Jacobson, "A proof of the convergence of the Kelley-Bryson penalty function technique for state-constrained control problems," *J. Math. Anal. Appl.*, vol. 26, pp. 163–169, 1969.
12. A.Q. Xing and C.L. Wang, "Applications of the exterior penalty method in contrained optimal control problems," *Optimal Control Appl. Meth. UK*, vol. 10, no. 4, pp. 333–345, 1989.
13. E. Polak, "An historical survey of computational methods in optimal control," *SIAM Rev.*, vol. 15, part 2, pp. 553–584, April 1973.
14. C.H. Chen, S.C. Chang and I.K. Fong, "An effective differential dynamic programming algorithm for constrained optimal control problems," in *Proc. Am. Control Conf.*, Pittsburgh, PA, vol. 2, pp. 1763–1764, June 1989.
15. S.C. Chang, C.H. Chen, I.K. Fong and P.B. Luh, "Hydroelectric generation scheduling with an effective differential dynamic programming algorithm," *IEEE Trans. Power Syst.*, vol. 5, no. 3, pp. 737–743, August 1990.

16. D.P. Bertsekas, "Multiplier methods: a survey," *Automatica*, vol. 12, pp. 133–145, 1976.
17. D.P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press: London, New York, 1982.
18. A. Miele, "Gradient algorithms for the optimization of dynamic systems," *Control Dynam. Syst.*, vol. 16, pp. 3–52, 1980.
19. A.E. Bryson Jr. and Y.C. Ho, *Applied Optimal Control—Optimization, Estimation and Control*. Halsted Press: New York, 1975.
20. D.J.W. Ruxton, "Differential dynamic programming and optimal control of inequality constrained continuous dynamic systems," M.Sc. thesis, University of Central Queensland, Australia, 1991.
21. D.J.W. Ruxton, "Applying the differential dynamic programming algorithm to state inequality constrained continuous optimal control problems," to appear in *Proc. Fifth Workshop Control Mech.*, University of Southern California, January 1992.