Feedback control of state constrained optimal control problems

D.A. Redfern and C.J. Goh
Department of Mathematics, University of Western Australia
Nedlands, Western Australia 6907. e-mail: `goh@maths.uwa.edu.au`

Abstract

Solutions to nonlinear optimal control problems are usually computed in open-loop form, especially if the problem is subjected to state constraints. Open-loop controls, however, are not very useful from a control perspective due to their lack of robustness. We discuss a couple of useful techniques which reduce a state-constrained problem into an unconstrained problem, and subsequently apply a technique to synthesize an optimal state feedback controller for the reduced problem. The controller is able to recover the open-loop optimal state and control for arbitrary initial conditions arising from a nontrivial subset of the state space, while ensuring that the state constraint is satisfied at all time.

Keywords

Feedback control, state constraints, optimal control, neural networks.

1 INTRODUCTION

The linear-quadratic regulator is one of the most popular design methodologies in modern control theory. It is also one of the rare optimal control problems where the optimal control is furnished in the form of a feedback controller which is independent of the system's initial condition. For most nonlinear optimal control problems, one is usually concerned with the computation of an open-loop optimal control which is dependent on the system's initial condition. In general, optimal feedback control law for nonlinear optimal control problems with general cost functionals are extremely difficult to compute. For some simple problems such as the linear quadratic regulator, the optimal feedback controller can be obtained by solving the corresponding Hamilton-Jacobi-Bellman (HJB) dynamic programming equation. In practice, the curse of dimensionality associated with dynamic programming incurs serious computational difficulties for other problems. Sometimes local perturbation methods can be used for constructing the feedback controller around a nominally optimal trajectory, see for example Bryson and Ho (1975). However, these perturbation methods are unlikely to be effective in the event of a large shift in initial conditions.

In Edwards and Goh (1995), a direct training method is proposed for the synthesis of

an optimal feedback controller in the form of a feedforward neural network for continuous-time nonlinear dynamical systems. The controller is independent of the system's initial condition provided that it arises from some bounded domain of the state space, and is not constricted by any fixed model structure for the dynamical system. The underlying idea is motivated by a paper of Nguyen and Widrow (1990) where a trailer-truck is trained to back up to a loading dock from arbitrary initial conditions. Although it was not explicitly mentioned in Nguyen and Widrow (1990), the problem addressed therein was really a discrete-time optimal control problem. The controller in this case is also a neural network, but it is possible that other versatile function emulators such as radial basis functions can also be used for the controller.

It is the goal of this paper to extend the result in Edwards and Goh (1995) to nonlinear feedback optimal control problems subjected to further continuous state constraints. It is well-known that, even for computing open-loop optimal solutions, the state constrained optimal control problem is notoriously difficult. While there is no shortage of theoretical results in the form of necessary conditions for the constrained problem, practical methods for constructing the constrained solutions have been few. In this paper, we discuss two methods to transform a state-constrained problem into an unconstrained problem, and subsequently apply the method in Edwards and Goh (1995) to synthesize the optimal feedback controller for the resulting unconstrained problem.

2   FEEDBACK CONTROL OF UNCONSTRAINED OPTIMAL CONTROL PROBLEMS

We introduce the underlying notion of a feedback controller by formulating the unconstrained problem first. Consider the continuous-time dynamical system defined in the finite interval $[0, T]$:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, t), \tag{1}$$

$$\mathbf{x}(0) = \boldsymbol{\xi}, \tag{2}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the state; $\mathbf{u} \in \mathbb{R}^m$ is the control; and $\mathbf{f} = (f_1, f_2, \cdots, f_n)^\top : \mathbb{R}^n \times \mathbb{R}^m \times [0, T] \to \mathbb{R}^n$ is assumed known, smooth and Lipschitz. The initial state $\boldsymbol{\xi}$ is distributed randomly in some bounded subset Γ of the state space according to some density distribution $\rho(\boldsymbol{\xi})$ where ρ vanishes identically outside Γ. The control is subjected to the simple bound constraint:

$$\alpha_i \le u_i(t) \le \beta_i, \quad \forall i = 1, \cdots, m, \quad \text{and} \quad \forall t \in [0, T]. \tag{3}$$

A control $\mathbf{u}$ is said to be 'admissible' if it is a measurable function which satisfies the bound (3). Let $\mathcal{U}$ be the class of all such admissible controls. The (open-loop) optimal control problem can be stated as:

Problem P1: Given a fixed initial condition $\mathbf{x}(0) = \boldsymbol{\xi}$, find a $\mathbf{u}(\cdot; \boldsymbol{\xi})$ from $\mathcal{U}$ such that the cost functional:

$$J_1(\mathbf{u}(\cdot); \boldsymbol{\xi}) = \Phi(\mathbf{x}(T)) + \int_0^T L(\mathbf{x}(t), \mathbf{u}(t), t) dt \tag{4}$$

is minimized, where $L : \mathbb{R}^n \times \mathbb{R}^m \times [0, T] \to \mathbb{R}$ and $\Phi : \mathbb{R}^n \to \mathbb{R}$ are known smooth functions. Note that the (open-loop) optimal control $\mathbf{u}^*$ which solves P_1 is dependent on

the initial condition $\boldsymbol{\xi}$. Assuming that full state measurement is possible, the feedback problem is defined by:

Problem P2: Construct a state feedback controller $\mathbf{u} = \mathbf{g}(\mathbf{x}, t)$, such that for any given initial condition in Γ, the feedback control will give rise to the same minimum cost functional, optimal control and optimal state trajectory as that of the open-loop optimal solution to problem P1.

Such a feedback controller is then independent of any shift in initial condition (provided that it remains in Γ), and is almost surely more robust than the open-loop controller in the presence of noise and parameter variation. An exact analytical solution for $\mathbf{g}(\mathbf{x}, t)$ is only possible in trivial cases such as the linear quadratic regulator. We shall return in section 4 to discuss how we can synthesize an approximate optimal feedback controller.

3   TWO TECHNIQUES FOR REDUCING CONSTRAINED TO UNCONSTRAINED PROBLEMS

In some practical situations such as obstacle avoidance in robot control, the system is subjected to one or more of the following state constraint:

$$S(\mathbf{x}(t), t) \leq 0 \tag{5}$$

where $S : \mathbb{R}^n \times [0, T] \to \mathbb{R}$ is assumed to be smooth. We say that the constraint (5) is *of order K* if $dS/dt, d^2S/dt^2, \cdots, d^{K-1}S/dt^{K-1}$ do not contain $\mathbf{u}$ explicitly, and $d^K S/dt^K$ is the first term that contains $\mathbf{u}$ explicitly. We shall discuss two techniques to transform a state constrained problems into an unconstrained problem, and subsequently the direct training method of Edwards and Goh (1995) will be used in section 4 to construct an optimal feedback controller which, due to the prior transformation, will ensure that these state constraints are satisfied.

3.1   The penalty function approach

The first technique is based on the commonly used concept of exact penalty function. In particular, we use a special constraint transcription technique as presented in Teo et. al. (1992) which, apart from satisfying the usual constraint qualification, has several nice properties. While the penalty function method is entirely general, and may account for vector control with multiple state constraints, it is often slow in convergence, especially when one tries to train the feedback controller for the transformed problem. Furthermore, it is almost impossible to achieve perfect constraint satisfaction. We summarize the technique as follows and refer the readers to Teo et. al. (1992) for further details. As with all penalty function techniques, we append the constraint into the cost functional as follows:

$$J(\mathbf{u}(\cdot)) \;=\; \Phi(\mathbf{x}(T), T) + \int_0^T \left(L(\mathbf{x}(t), \mathbf{u}(t), t) + \mu L_\varepsilon(\mathbf{x}(t), t) \right) dt, \tag{6}$$

where $\;L_\varepsilon(\mathbf{x}(t), t) \;=\; \begin{cases} 0, & \text{if } S(\mathbf{x}(t), t) < -\varepsilon, \\ \frac{(S(\mathbf{x}(t), t) + \varepsilon)^2}{4\varepsilon}, & \text{if } -\varepsilon < S(\mathbf{x}(t), t) < \varepsilon, \\ S(\mathbf{x}(t), t), & \text{if } S(\mathbf{x}(t), t) > \varepsilon, \end{cases} \tag{7}$

and μ is some large penalty weight. To determine the solution to the constrained problem, ε is initially set to some moderately small positive value of, say 10^{-2}, and μ is set to some moderately large value, say 10^2. Note that if $\varepsilon > 0$ and if the penalty term is identically zero, then the constraint will be satisfied more than is necessary. In fact $S(t) \leq -\varepsilon < 0$. The optimal control problem is then solved and the feasibility of the constraint checked. If there is any constraint violation, increase μ, otherwise decrease ε, and the problem is solved again, until ε is less than a certain prescribed threshold. The rate of increasing μ and decreasing ε is based on some heuristic which comes about as the result of much trial and error. Multiple state constraints can be treated similarly by appending one penalty term for each constraint to the cost functional.

3.2   The Valentine transformation

The second technique is based on a result due to Valentine (1937) from the early calculus of variation literature. This result has subsequently been applied to compute the open-loop solution for state constrained optimal control problems in Jacobson and Lele (1969) and Miele et. al. (1979). We further extend the open-loop result to the synthesis of an optimal feedback controller. Unlike the penalty function approach, the Valentine transformation increases the order of the state space (by up uo the order of the constraint) using a new control. For some problems, this method works extremely well and the constraint is always satisfied exactly. The feedback controller can be trained in a fraction of the time taken by the penalty function approach. Nevertheless for the past two decades, very little attention has been paid to this seemingly elegant technique, but no reasons have been given to the best of our knowledge. We shall discuss in detail, the application of the Valentine transformation to the synthesis of feedback controllers, and point out the limitation which prevents this technique from becoming popular.

We consider firstly the case involving a single control and a single constraint. Generalization of the method to problems involving multiple constraints and multiple controls is slightly more tricky but not difficult. Firstly the inequality constraint is converted to an equality constraint using a slack variable:

$$S(\mathbf{x}(t), t) + \frac{1}{2}(\alpha(t))^2 = 0. \tag{8}$$

Repeated differentiation of equation (8) leads to the following set of equations, where $S^{(k)}$ and $\alpha^{(k)}$ are the kth derivative of S and α with respect to t, respectively:

$$S^{(1)}(\mathbf{x}(t), t) + \alpha(t)\alpha^{(1)}(t) = 0,$$
$$S^{(2)}(\mathbf{x}(t), t) + \left(\alpha^{(1)}(t)\right)^2 + \alpha(t)\alpha^{(2)}(t) = 0,$$
$$S^{(3)}(\mathbf{x}(t), t) + 3\alpha^{(1)}(t)\alpha^{(2)}(t) + \alpha\alpha^{(3)}(t) = 0, \tag{9}$$
$$\vdots$$
$$S^{(K)}(\mathbf{x}(t), u(t), t) + \{\text{terms involving } \alpha^{(1)}(t), ..., \alpha^{(K-1)}(t)\} + \alpha(t)\alpha^{(K)}(\mathrm{t}) = 0.$$

If the constraint (9) is of order K, then $\partial S^{(K)}/\partial u \neq 0$. The implicit function theorem asserts that the control u can be expressed as a function of the other variables:

$$u = \phi(\mathbf{x}, \alpha, \alpha^{(1)}, ..., \alpha^{(K-1)}, \alpha^{(K)}, t). \tag{10}$$

If this cannot be done analytically then an approximate implicit function can be emulated using a function emulator, typically either a neural network or a radial basis function. Treating $\alpha, \alpha^{(1)}, ..., \alpha^{(K-1)}$ as additional state variables and $v = \alpha^{(K)}$ as the new control

variable, and using (10), the following unconstrained problem is obtained.

$$\min \ J(v(\cdot)) \quad = \quad \Phi(\mathbf{x}(T),T) + \int_0^T L(\mathbf{x}(t), \phi(\mathbf{x},\alpha,\alpha^{(1)},...,\alpha^{(K-1)},v(t),t),t)dt \tag{11}$$

$$\begin{aligned} \dot{\mathbf{x}} &= f(\mathbf{x}(t), \phi(\mathbf{x},\alpha,\alpha^{(1)},...,\alpha^{(K-1)},v(t),t),t) \\ \dot{\alpha} &= \alpha^{(1)} \\ &\vdots \\ \dot{\alpha}^{(K-1)} &= v(t) \end{aligned} \tag{12}$$

with the initial conditions $\mathbf{x}(0) = \xi, \alpha(0), \alpha^{(1)}(0),...,\alpha^{(K-1)}(0)$; where $\alpha(0), \alpha^{(1)}(0),...,$ $\alpha^{(K-1)}(0)$ are obtained directly from equation (8) and the first $K-1$ equations from (9):

$$\begin{aligned} \alpha(0) &= \sqrt{-2S(\mathbf{x}(0),0)} \\ \alpha^{(1)}(0) &= -\frac{S^{(1)}(\mathbf{x}(0),0)}{\alpha(0)} \\ \alpha^{(2)}(0) &= -\frac{(\alpha^{(1)}(0))^2 + S^{(2)}(\mathbf{x}(0),0)}{\alpha(0)} \\ &\vdots \end{aligned} \tag{13}$$

Despite the guarantee of feasibility, this transformation technique does have a major drawback which severely limits the problems for which is appropriate. To be specific, for a certain class of problem this transformation results in a control which is unbounded at various points in the state space. As this singularity problem has not been reported elsewhere to the best of our knowledge, we shall illustrate it using a simple example, which is modified from a problem discussed in Bryson and Ho (1975).

Consider the following modified Zermelo problem. The problem represents the control of a ship traveling in a region of variable current strength. Instead of fixing the speed and controling the orientation of the ship as in the original Zermelo's problem, we fix the orientation (θ) and control the speed instead. The state constraint in this case is given by a parabolic boundary which the ship is to stay clear of.

$$\text{Min} \ \ J(u(\cdot)) \quad = \quad \int_0^1 u^2 dt$$

$$\begin{aligned} \text{with} \ \ \dot{x}_1 &= u\cos\theta + \beta x_2 \\ \dot{x}_2 &= u\sin\theta \end{aligned}$$

subject to the terminal constraints $x_1(1) = 0$, $x_2(1) = 0$; and the first order constraint:

$$x_2 \le a(x_1 - b)^2 + c. \tag{14}$$

Addition of an appropriate slack variable α leads to

$$x_2 + \frac{1}{2}(\alpha(t))^2 = a(x_1 - b)^2 + c. \tag{15}$$

Differentiating the above once and introducing the new control $v(t) = \alpha^{(1)}(t)$ to obtain:

$$u\sin\theta + \alpha v = 2a(x_1 - b)(u\cos\theta + \beta x_2), \tag{16}$$

and solving for $u(t)$ we get

$$u = \frac{2a\beta x_2(x_1 - b) - \alpha v}{\sin\theta - 2a(x_1 - b)\cos\theta}. \tag{17}$$

Note that u becomes unbounded whenever the denominator vanishes, i.e., when

$$x_1 = \frac{\tan\theta}{2a} + b. \tag{18}$$

For realistic values of the parameters a, b and θ this often causes problems. For example if we have $a = 5$, $b = -0.5$ and $\theta = \frac{\pi}{4}$ then the control u becomes unbounded whenever x_1 approaches -0.4. Thus the method will fail if we wish to find the optimal trajectory from an initial condition with $x_1(0)$ less than -0.4. In general, if $x_1(0)$ is less than $\tan(\theta)/2a +$ b, this singularity problem will surely arise. This difficulty was first observed when we attempt to control the movement of an inverted pendulum which results in consistent numerical singularity. The above example was then constructed to illustrate this singular effect. Despite this limitation, the Valentine transformation technique has been found to be generally effective for other nonsingular problems.

4   SYNTHESIS OF OPTIMAL FEEDBACK CONTROLLER

Once the appropriate transformation has been applied to reduce the constrained problem into an unconstrained problem, the technique of Edwards and Goh (1995) can be applied to construct a feedback controller that solves problem P2. The basic idea is to approximate the optimal controller by some parameterized model $\mathbf{u}(t) = \hat{\mathbf{g}}(\mathbf{x}(t), t; \mathcal{W})$, where $\mathcal{W}$ is the set of parameters for the controller, which is determined optimally by solving the following optimal parameter selection problem:

Problem P3: Find the optimal set of parameters $\mathcal{W}$ such that the cost functional

$$J_2(\mathcal{W}) = E(J_2'(\mathcal{W}; \boldsymbol{\xi})) = \int_{\boldsymbol{\xi} \in \Gamma} J_2'(\mathcal{W}; \boldsymbol{\xi}) \rho(\boldsymbol{\xi}) d\boldsymbol{\xi} \tag{19}$$

is minimized with respect to $\mathcal{W}$, where

$$J_2'(\mathcal{W}; \boldsymbol{\xi}) = \Phi(\mathbf{x}(T)) + \int_0^T L(\mathbf{x}, \hat{\mathbf{g}}(\mathbf{x}, t; \mathcal{W}), t) dt, \tag{20}$$

and $\mathbf{x}(\cdot)$ is uniquely determined by solving the (homogeneous) state differential equation:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \hat{\mathbf{g}}(\mathbf{x}(t), t; \mathcal{W}), t), \tag{21}$$
$$\mathbf{x}(0) = \boldsymbol{\xi}. \tag{22}$$

There are several possibilities for choosing the parameterized controller, the key requirement is that it must be flexible enough to emulate the true optimal controller. Theoretically, any function approximation schemes can be used, although in practice, our experience suggests that only radial basis functions or feedforward neural networks have any chance of success. In particular, when the controller has a large input dimension (greater than 5 say), as is usually the case after the application of Valentine transformation, neural networks appear to be the only feasible model without requiring an unwieldily large number of parameters. Furthermore, a neural network with a bounded sigmoidal output has the natural advantage of modeling the bounded control output as given by (3). It is also

versatile enough to emulate discontinuous control as is often encountered in bang-bang type control. In what follows, we shall assume that a neural network will be used for the controller, although the idea can also be used for any other parameterized model.

Problem P3 is a nonlinear programming problem in disguise, although the determination of the optimal set of weight parameters have to be computed in a somewhat round-about way. The direct training algorithm for the feedback controller in the form of a neural network is based on the gradient formulae for optimal parameter selection problems (see Teo et. al. (1992)), in conjunction with a steepest descent algorithm using instantaneous gradient similar to, although not quite the same, as the backpropagation training of neural networks. We shall summarize the important steps required in the training algorithm, and refer the readers to Edwards and Goh (1995) for details.

Modified Backpropagation Algorithm.

Initialization: generate a set of initial conditions $\Delta = \{\boldsymbol{\xi}_i, i = 1, \cdots, N\}$ from the distribution ρ which fills the set Γ in a sufficiently dense manner. Select an appropriate neural network and randomize all the weights $w \in \mathcal{W}$ to small values. While the average performance index can be reduced further, (more specifically, if the average performance index in subsequent iteration reduces by more than some pre-assigned threshold) do:

i. For each $\boldsymbol{\xi}_i$ solve the differential equation (21) and (22) forward in time from $t = 0$ to $t = T$.

ii. Solve the corresponding costate differential equation

$$\dot{\boldsymbol{\lambda}}^{\mathsf{T}} = -\frac{\partial L}{\partial \mathbf{x}} - \frac{\partial L}{\partial \mathbf{u}}\frac{\partial \hat{\mathbf{g}}}{\partial \mathbf{x}} - \boldsymbol{\lambda}^{\mathsf{T}}\left(\frac{\partial \mathbf{f}}{\partial \mathbf{x}} + \frac{\partial \mathbf{f}}{\partial \mathbf{u}}\frac{\partial \hat{\mathbf{g}}}{\partial \mathbf{x}}\right) \tag{23}$$

$$\boldsymbol{\lambda}^{\mathsf{T}}(T) = \frac{\partial \Phi}{\partial \mathbf{x}(T)}, \tag{24}$$

backward in time from $t = T$ to $t = 0$, where $\partial \hat{\mathbf{g}}/\partial \mathbf{x}$ is the Jacobian of the controller to be computed by some special formulae derived in much the same way as the backpropagation formulae (see equations (13)-(15) of Edwards and Goh (1995)).

iii. Compute, for each weight parameter $w \in \mathcal{W}$

$$\frac{\partial J_2'}{\partial w} = \int_0^T \left(\frac{\partial L}{\partial \mathbf{u}}\frac{\partial \hat{\mathbf{g}}}{\partial w} + \boldsymbol{\lambda}^{\mathsf{T}}\frac{\partial \mathbf{f}}{\partial \mathbf{u}}\frac{\partial \hat{\mathbf{g}}}{\partial w}\right)dt \tag{25}$$

where $\partial \hat{\mathbf{g}}/\partial w$ is computed by the usual backpropagation formulae.

iv. Update the weight $w, \forall w \in \mathcal{W}$, by $w \leftarrow w - \eta \partial J_2'/\partial w$, where η is some small learning rate.

The algorithm described above is only conceptual. For successful implementation of the algorithm, there are a number of fine details that warrant critical attention. Firstly, the choice of network's structure and parameterization (number of layers and nodes) for implementing the modified backpropagation algorithm is problem dependent. At this stage of research, there are no clever ways of choosing the network apart from systematic trial and error. Unfortunately, this has remained a major criticism of the application of neural

networks, despite the fact that they have been used successfully to solve many difficult problems. Secondly, when the system is inherently unstable, difficulty will arise during the integration of the state and costate equations in the initial training stage before the weights are appropriately initialized. An ad hoc way to overcome this is to start integrating with a small terminal time T, so that the system does not have sufficient time to escape to infinity, and gradually increase T until the desired T is reached. Nevertheless this has not been an entirely satisfactory way for highly unstable systems. Other issues such as the choice of training rate and the sequence in which the training initial conditions are used are also yet to be completely resolved.

5   CONCLUDING REMARKS

The proposed method has been tested on several non-trivial test problems, using both the penalty function method and the Valentine's transformation, with varying degree of success. In a test problem considered by Jacobson and Lele (1969), the trained feedback controller is able to achieve within 1% of the performance of the open-loop controller for all initial conditions arising from a fairly large subset. Furthermore, the feedback controller is also capable of generating the optimal solution for initial conditions not included in the training set, and even sometime for initial conditions *outside* the training set. Due to space constraint, these simulation results are not included but are readily available from the second author upon request.

REFERENCES

Bryson, A.E. and Ho, Y.C. (1975) *Applied Optimal Control.* Hemisphere publishing, Washington D.C..

Edwards, N.J. and Goh, C.J. (1995) A direct training method for continuous-time nonlinear optimal feedback controller. *Journal of Optimization Theory and Applications* **84(3)**.

Jacobson, D.H. and Lele, M.M. (1969) A transformation technique for optimal control problems with a state variable inequality constraint. *IEEE Trans. on Automatic Control,* **AC-14(5)**, 457-64.

Miele, A., Wu, A.K. and Liu, C.T. (1979) A transformation technique for optimal control problems with partially linear state inequality constraints. *Journal of Optimization Theory and Applications,* **28(2)**, 185-212.

Nguyen, D.H. and Widrow, B. (1990) Neural networks for self learning control systems. *IEEE Control Systems Magazine,* April, 18-23.

Teo, K.L., Goh, C.J. and Wong, K.H. (1991) *A Unified Computational Approach to Optimal Control Problems.* Pitman Series in Pure and Applied Mathematics, Longman.

Valentine, F.A. (1937) The problem of lagrange with differential inequalities as added side conditions. *Contribution to the Calculus of Variation,* University of Chicago press.