AIAA Guidance, Navigation, and
Control Conference and Exhibit
14-17 August 2000 Denver, CO

A00-37199

AIAA-2000-4559

Worst Case Dynamic Programming and Its Application to Deterministic Systems

Suman Chakravorty* David C. Hyland†
Department of Aerospace Engineering
University of Michigan, Ann Arbor

Abstract

In this paper, we investigate the numerical solution of discrete time ,infinite horizon,stationary ,discounted Dynamic Programming(DP) problem. We introduce the worst case variant of the standard DP operator and show that it retains the nice properties of the former.We apply this worst case formulation to deterministic systems in continuous space and show that for a class of such systems, any arbitrary degree of accuracy can be attained in the numerical solution to the DP problem. Finally we illustrate our results through a simple orbital dynamics example.

1 Introduction

Dynamic programming(DP) provides a methodology for optimal decision making under uncertainity. In the typical DP problem, a system evolves in a continuous state space S(usually$S \subseteq \Re^n$) and the interest is to find the fixed point of the DP operator T,

$$TP(x) = \inf_{u \in A}[C(x,u) + \alpha \int_S P(y)p(y/x,u)dy] \quad \forall x \in S \qquad (1.1)$$

where A is the set of control actions. $C(x,u)$ is the incremental cost incurred by the sytem in taking control action u at state x, $p(y/x,u)$ defines a probability distribution which given the current state x and control action u specifies the probability that the system will be in state y at the next instant and $\alpha \in (0,1)$ is a discount factor. The fixed point P^* of the DP operator T is the pivotal construct of this methodology and is known as the optimal "cost-to-go/value/reward" function. Unfortunately the DP equation $TP^* = P^*$ isn't usually amenable to closed form solutions thereby forcing us to attempt the solution numerically.

One method of solving the DP equation numerically is to break the state and control spaces into a finite number of the same and obtaining the solution to the finite state DP problem that results from this discretisation. There are a number of ways of solving finite state DP problems like value iteration(successive approximation), policy iteration and linear programming. A comprehensive survey of the state of the art in this field is given in Bertsekas[2].

This paper focuses on the problem of discretisation of state and control spaces such that a desired level of accuracy can be attained in the cost-to-go estimates. This problem has been addressed before by Hernandez-Lerma[3],Chow[4]and Whitt[5]. They show that for a class of sytems, the errors in the cost to go estimates can be made arbitrarily small if the discretisation of the state space is sufficiently fine.The assumptions made and the bounds obtained for the approximation errors are similar in [3,4,5]. For an exhaustive survey and discussion of these results, see Rust[6].A common underlying assumption of all these results is the lipschitz continuity of the probability density function $p(y/x,u)$. As mentioned in chow[4] this condition need not be true in general and an important special case where the results are not valid is that of deterministic systems where p corresponds to a singular measure , as opposed to a density.

The main contribution of this paper is to introduce a "worst case" variant of the standard DP problem and show that for the case of deterministic systems, under certain assumptions, with sufficiently fine discretisation of the state space, any desired level of accuracy can be obtained in the cost-to-go estimates.In order to facilitate this , instead of a probability density function, a step function $g(y/x,u)$ is introduced where

*Graduate Student Research Assistant
†Professor, Fellow AIAA

$$g(y/x,u) = \begin{cases} 1 & x \rightarrow^u y, \\ 0 & o.w. \end{cases} \quad (1.2)$$

correspondingly the DP problem boils down to finding the fixed point of the " worst case or Min-Max" DP operator T defined by

$$TP(x) = \inf_{u \in A}[\sup_{y \in S}[c(x,u) + \alpha P(y) g(y/x,u)]] \quad \forall x \in S \quad (1.3)$$

This problem is solved by discretising the control and state spaces to obtain the finite state variant of the worst case DP equation which can in turn be solved by any of the methods in (2). Finally the convergence result is illustrated through a simple example.

2 The worst case dynamic programming operator

In this section we introduce the worst case DP operator and discuss a few of its properties.

Definition 2.1 *The continuous state,discrete time system is defined by the equation:*

$$x_k = f(x_{k-1}, u_{k-1}, w_k), \quad (2.1)$$

where x_k is the state of the system at instant k,u_k is the control action taken at instant k and w_k is a noise term whose statistics are known.

Let S denote the state space of the system. $S \subset \Re^n$ and is compact. Let A denote the control/action space (it is a subset of some function space X).The **incremental cost** incurred by the system in taking control u_k at state x_k is given by the non negative and bounded function $c(x_k, u_k)$.

Definition 2.2 A **control policy** *is defined as a sequence of control actions i.e. a policy $\pi = (\pi(0), \pi(1), ..)$ where $\pi(k)$ belongs to A and is the control action chosen by the system at instant k.*

Consider any $x \in S$ and any policy π. Note now that if policy π were applied to the sytem starting at x, due to the uncertainity in the system there would be more than one path from x corresponding to π. Let these paths be denoted by $\gamma = (\gamma(0), \gamma(1), ..)$ such that

$$\gamma_k = f(\gamma_{k-1}, u_{k-1}, w_k), \quad (2.2)$$

$$\gamma_0 = x, \quad (2.3)$$

where $(w(1), w(2), ..)$ is any probable sequence of the noise term.

Definition 2.3 *The cost to go w.r.t path γ for policy π from x is defined as*

$$P_\pi^\gamma(x) = \sum_{k=0}^{\infty} \beta^k c(\gamma_k, u_k), \quad (2.4)$$

Definition 2.4 *Now we define the cost to go from state x w.r.t policy π as*

$$P_\pi(x) = \sup_\gamma P_\pi^\gamma(x). \quad (2.5)$$

Definition 2.5 *Finally we define the optimal cost to go from state x as*

$$P^*(x) = \inf_\pi P_\pi(x). \quad (2.6)$$

Now we define a step function $g(y/x,u)$ such that

$$g(y/x,u) = \begin{cases} 1 & x \rightarrow^u y, \\ 0 & o.w. \end{cases} \quad (2.7)$$

Definition 2.6 *We denote the next state set of some x under control action u by $\Gamma(x,u)$ and define it as*

$$\Gamma(x,u) = (y \in S | g(y/x,u) = 1). \quad (2.8)$$

This leads us to the definition of the worst case DP operator T:

$$TP(x) = \inf_{u \in A}(\sup_{y \in S}(c(x,u) + \beta P(y) g(y/x,u))) = \inf_{u \in A}(\sup_{y \in \Gamma(x,u)}(c(x,u) + \beta P(y))), \quad \forall x \in S. \quad (2.9)$$

With the above definitions we state and prove the following proposition,

Proposition 2.1 *The worst case DP operator T is a contraction mapping w.r.t the sup norm i.e*

$$||TP - TP^{'}||_\infty \leq \beta||P - P^{'}||_\infty$$
$$\text{where} \quad ||P||_\infty = \sup_{x\in S}|P(x)|.$$

Proof : To prove this result we need to quote a lemma from Hernandez-Lerma(3) p123.

Lemma 2.1 *Let X be any arbitrary nonempty set and let u and v be functions from X to $\Re$ bounded from above(i.e supu and supv exist) then*

$$|\sup_{x\in X} u(x) - \sup_{x\in X} v(x)| \leq \sup_{x\in X}|u(x) - v(x)|. \quad (2.10)$$

Let

$$G(x,u,P) = c(x,u) + \sup_{y\in\Gamma(x,u)} \beta P(y). \quad (2.11)$$

where P is any bounded functional from S into $\Re$. Then

$$|G(x,u,P) - G(x,u,P^{'})| = \beta|\sup_{y\in\Gamma(x,u)} P(y) - \sup_{y\in\Gamma(x,u)} P^{'}(y)|, \quad \forall x \in S. \quad (2.12)$$

(2.12) and lemma 2.1 $\Rightarrow$

$$|G(x,u,P) - G(x,u,P^{'})| \leq \beta \sup_{y\in\Gamma(x,u)}|P(y) - P^{'}(y)|, \leq \beta\sup_{y\in S}|P(y) - P^{'}(y)| = \beta||P - P^{'}||_\infty \quad \forall x \in S. \quad (2.13)$$

Note now that :

$$TP(x) = \inf_{u\in A} G(x,u,P), \quad (2.14)$$

hence

$$|TP(x) - TP^{'}(x)| = |\inf_{u\in A} G(x,u,P) - \inf_{u\in A} G(x,u,P^{'})| \quad \forall x \in S. \quad (2.15)$$

Again note that $G(x,u,P)$ is bounded and non negative $\forall\ u \in A$. Hence from lemma1 and (2.15),

$$|TP(x) - TP^{'}(x)| \leq \sup_{u\in A}|G(x,u,P) - G(x,u,P^{'})| \quad \forall x \in S, \quad (2.16)$$

(2.13) and (2.16) $\Rightarrow$

$$|TP(x) - TP^{'}(x)| \leq \beta||P - P'||_\infty \forall\ x \in S,$$

$$\Rightarrow \sup_{x\in S}|TP(x) - TP^{'}(x)| = ||TP - TP^{'}||_\infty \leq \beta||P - P^{'}||_\infty, \quad (2.17)$$

q.e.d

Next we state and prove the following proposition

Proposition 2.2 *The optimal cost to go from any state $x \in S$ satisfies*

$$TP^*(x) = P^*(x). \quad (2.18)$$

Proof: By definition we have that,

$$P^*(x) = \inf_\pi P_\pi(x), \quad (2.19)$$
$$\Rightarrow$$
$$P^*(x) = \inf_\pi(\sup_\gamma \sum_{k=0}^{\infty}\beta^k c(\gamma_k,\pi_k)),$$
$$= \inf_\pi(c(x,\pi_0) + \beta\sup_\gamma \sum_{k=0}^{\infty}\beta^k c(\gamma_{k+1},\pi_{k+1})),$$
$$= \inf_u \inf_\pi(c(x,u) + \beta \sup_{y\in\Gamma(x,u)} P_\pi(y)),$$
$$= \inf_u(c(x,u) + \beta \sup_{y\in\Gamma(x,u)} P^*(y)), \quad (2.20)$$
$$\Rightarrow$$
$$P^*(x) = TP^*(x)\ \forall\, x \in S. \quad (2.21)$$

q.e.d

By Propositon1 and the contraction mapping theorem(8) , there exists a unique fixed point of T in S. Hence it follows from Proposition 2 that the fixed point is P^*, the optimal cost to go.

Definition 2.7 *We define a stationary policy as a control policy under which the control action taken at a particular state is the same regardless of the instant it is taken.*

Notice that the optimal control policy defined by P^* i.e.

$$u^*(x) = arg \inf_{u\in A}(\sup_{y\in S}(c(x,u) + \beta P(y)g(y/x,u))) \quad \forall x \in S. (2.22)$$

defines a stationary policy. Hence to search for the optimal policy it is sufficient to search in the space of stationary policies. We also state without proof another proposition,

Proposition 2.3 *The operator T^π defined by*

$$T^\pi P(x) = \sup_{y\in\Gamma(x,\pi(x))}(c(x,\pi(x)) + \beta P(y)), \forall\, x \in S, \quad (2.23)$$

where π is any stationary policy, is a contraction operator and P_π is its unique fixed point.

Note : P_π is a function from S into $\Re$ where $P_\pi(x)$ is as previously defined for all x in S.

3 Bounds on errors in the cost-to-go approximation

In this section we present results which state that for a deterministic system,under certain assumptions,any arbitrary degree of accuracy can be attained in the cost-to-go estimates if the degree of discretisation is fine enough. We start by defining the deterministic system and making assumptions on the dynamics of the system.

System The dynamics of the system is governed by

$$x(k) = f(x(k-1), u(k-1)), \quad (3.1)$$

where $x(k)$ is the state of the system at the kth instant and $u(k)$ is the control action chosen at the kth instant.

State Space

The state space S is a subset of $\Re^n$. It is further divided into a finite number of non-empty subsets S_j, $j = 1, 2..N$ $\ni$

$$\bigcup_{j=1}^{N} S_j = S, \quad (3.2)$$

$$S_j \bigcap S_k = \phi \text{ if } j \neq k. \quad (3.3)$$

Control/action space: The control space is assumed to consist of a finite number of control actions and denoted by A. The control actions are assumed to be elements of some function space X.(usually however they are elements of $\Re^p$ where p denotes the number of inputs to the system).

It is easy to see that the system defined by (3.1) is the sytem in (2.1) without the noise term. However in this special case of a deterministic system there are a few notable changes:
For any policy π that is applied to the system starting at some $x \in S$ there can be only one path corresponding to the policy ,from x, and hence

$$P_\pi(x) = \sum_{k=1}^{\infty} \beta^k c(x_k, \pi_k), \quad (3.4)$$

where

$$x_k = f(x_{k-1}, \pi_{k-1}), \quad (3.5)$$

$$x_0 = x. \quad (3.6)$$

Also the DP operator changes to

$$TP(x) = \inf_{u\in A}(c(x,u) + \beta P(f(x,u))). \quad (3.7)$$

All the other results presented in the previous section still remain valid for this system with the abovementioned changes at the appropriate places.

For every S_j, we fix an arbitrary element in that set and call it the **exemplar** e_j corresponding to that set.

This leads us to define the dynamics of the discrete system. If the continuous system makes a transition from a point in set S_j to some point in S_i we say that the discrete system has made a transition from e_j to e_i.Also by taking control action u at any point in set S_j , the incremental cost incurred by the system is defined to be $c(e_j, u)$. Consider now the state space $S_d = \{e_1, e_2, ..., e_N\}$. S_d is a subset of $\Re^n$ and its dynamics can be formulated as in (2.7). Hence all the results that

were presented in section 2 are applicable to this system.But since we have a finite state space for this system, the following notation is added:

$$T_u(i,j) = \max_{x \in S_i, y \in S_j} g(x/y,u); \tag{3.8}$$

Let π be a stationary policy in continuous space. Then the discretised version of this policy in the discretised domain mentioned above is denoted by

$$\bar{\pi}(x) = \pi(x) \text{ if } x \in S_d \tag{3.9}$$

$$\bar{\pi}(x) = \pi(e_j) \text{ if } x \in S_j \tag{3.10}$$

$\mathcal{A}$ 3.1 *Let π be a stationary policy that asymptotically stabilises the origin of the system (3.1) and $\bar{\pi}$ its discretised counterpart. The dynamics of the system $x_k = f(x(k-1), \pi(x(k-1)))$ is* **lipschitz** *with respect to the state i.e*

$$\forall x, y \in S \ \exists \ K_f^{\pi} < \infty \ni$$
$$||f(x,\pi(x)) - f(y,\pi(y))|| \leq K_f^{\pi} ||x - y|| \tag{3.11}$$

Similarly for the discretised version of the policy π, the lipschitz inequality holds with the same lipshitz constant K_f^{π}.(this is for notational convenience) Further we assume that these constants over the space of all such asymptotically stabilising stationary policies is uniformly bounded above by a constant K_f i.e

$$\sup_{\pi} K_f^{\pi} \leq K_f < \infty \tag{3.12}$$

$\mathcal{A}$ 3.2 *Let π and $\bar{\pi}$ be as described above. Then we assume that $\exists \ N_1 < \infty \ni \forall n \geq N_1$*

$$\sup_k |\bar{x}(k) - x(k)| \leq K_s \rho(n), \tag{3.13}$$

where $(x_0 = x, x_1, ..x_k..)$ is the trajectory of the system from x under policy π and $(\bar{x}_0 = x, \bar{x}_1 ...)$ is the trajectory followed by the system under policy $\bar{\pi}$. Note that N_1 is independent of the policy π.

$\mathcal{A}$ 3.3 *The* **incremental** *cost function is Lipschitz in the domain S w.r.t the state of the system under policies π as above i.e*

$$\forall \ x, y \in S \ \exists \ K_c^{\pi} < \infty \ \ni$$
$$||c(x,\pi(x)) - c(y,\pi(y))|| \leq K_c^{\pi} ||x - y||. \tag{3.14}$$

The lipshitz condition is assumed to hold for the discretised version of the policy π with the same lipshitz constant and these constants are uniformly bounded above over the space of all such a.s. stationary policies i.e.

$$\sup_{\pi} K_c^{\pi} \leq K_c < \infty \tag{3.15}$$

Consider the division of the state space as described previously into N sets and let the exemplars corresponding to them be $(e_1, e_2, .., e_N)$.

Definition 3.1 *we define the* **diameter** *of the discretisation as*

$$\rho(N) = \max_{i=1}^{N} [\sup_{x,y \in S_i} ||x - y||]. \tag{3.16}$$

We also have the convention that $\rho(N) \geq \rho(N+1)$.

Let the optimal policy obtained by running the DP algorithm on the discrete state space $S_d = (e_1, e_2, ..., e_n)$ be denoted by π_n.

$\mathcal{A}$ 3.4 *The optimal policy in continuous state space ,π^* renders the origin of the system (3.1) asymptotically stable. In addition $\exists \eta < \infty \ \ni \ \forall \ n > \eta$, the optimal policy corresponding to $S_d = (e_1, ...e_n)$, π_n renders the origin of the system asymptotically stable.*

$\mathcal{A}$ 3.5 *For all stationary policies π that render the origin asymptotically stable(a.s) there exists a set Ω_π containing the origin such that $\forall \ x, y \in \Omega_\pi$*

$$||f(x,\pi(x)) - f(y,\pi(y))|| \leq \hat{k}_f^{\pi} ||x - y||. \tag{3.17}$$

where $\hat{k}_f^{\pi} < 1$, i.e in the set Ω_π any two solution trajectories can only come closer to each other and that

$$\sup_{\pi} \hat{k}_f^{\pi} = \hat{k}_f < 1 \tag{3.18}$$

Definition 3.2 *given a particular stationary policy that renders the origin a.s we define the diameter of the corresponding set Ω_π as*

$$\epsilon(\Omega_\pi) = \sup_{x,y \in \Omega_\pi} ||x - y||. \tag{3.19}$$

We also assume that this diameter is uniformly bounded below i.e. $\exists \epsilon_\infty \ \ni$

$$\inf_{\pi} \epsilon(\Omega_\pi) \geq \epsilon_\infty > 0. \tag{3.20}$$

$\mathcal{A}$ 3.6 *Let the time taken by the trajectory from state $x \in S$ to reach within $\epsilon_\infty/2$ of the origin under policy π be denoted by $t_\infty^{\pi}(x)$.*
We assume that

$$\sup_{\pi} \sup_{x \in S} t_\infty^{\pi}(x) \leq t_\infty < \infty. \tag{3.21}$$

With the above definitions and assumptions we make the following proposition,

Proposition 3.1 *Given a stationary policy π that renders the origin of the system a.s.*
$\exists(K_\infty < \infty, \quad N_\infty < \infty)$ *independent of* $\pi \ni \forall\, n \geq N_\infty$

$$|\bar{P}_{\bar{\pi}}(e_j) - P_\pi(x)| \leq K_\infty \rho(n). \tag{3.22}$$

Proof : Let $(x_0 = x, x_1, x_2, ...)$ denote the trajectory followed by the system from $x_0 = x$ under policy π.
Let the path followed by the discretised version of the policy be $(\bar{x}_0 = x, \bar{x}_1, ...)$. Then by (A 3.2) ,we have that $\forall n \geq N_1$

$$\sup_k ||x_k - \bar{x}_k|| \leq K_s \rho(n). \tag{3.23}$$

Hence it follows from (A 3.3) that

$$|P_{\bar{\pi}}(x) - P_\pi(x)| \leq \frac{K_c K_s}{(1-\beta)} \rho(n) \tag{3.24}$$

Let $E(x)$ denote the exemplar corresponding to the point x. In the discretised system there exists a path corresponding to $(\bar{x}_0 = x, \bar{x}_1, ...)$given by $((E(\bar{x}_0), ...))$. Let this path be denoted by γ.Then

$$|\bar{P}_{\bar{\pi}}(e_j) - P_{\bar{\pi}}(x)| \leq |\bar{P}_{\bar{\pi}}(e_j) - \bar{P}^\gamma_{\bar{\pi}}(e_j)| + |\bar{P}^\gamma_{\bar{\pi}}(e_j) - P_{\bar{\pi}}(x)|. \tag{3.25}$$

It follows from (A 3.3) that

$$|\bar{P}^\gamma_{\bar{\pi}}(e_j) - P_{\bar{\pi}}(x)| \leq \frac{K_c}{(1-\beta)} \rho(n) \tag{3.26}$$

We have that $\bar{P}_{\bar{\pi}}(e_j) = \bar{P}^\mu_{\bar{\pi}}(e_j)$ for some path μ in the discrete system.
From (A3.1, A 3.3) we have

$$||\gamma(k) - \mu(k)|| \leq (K_f^k + k_f^{k-1} + ... + 1) 2\rho(n) \tag{3.27}$$

However γ enters within $\frac{\epsilon_\infty}{2}$ of the origin in t_∞ time steps(A 3.6). Hence it follows from (A 3.5) that if we choose $N_2 \ni$

$$\rho(N_2) \leq min(\frac{\epsilon_\infty}{2(K_f^{t_\infty} + k_f^{t_\infty - 1} + ... + 1)}, \frac{\epsilon_\infty(1-\hat{k}_f)}{(1+\hat{k}_f)}) \tag{3.28}$$

then $\forall\, n \geq N_2$,

$$\sup_k ||\gamma(k) - \mu(k)|| \leq 2(K_f^{t_\infty} + k_f^{t_\infty - 1} + ... + 1)\rho(n) \tag{3.29}$$

which means that

$$|\bar{P}_{\bar{\pi}}(e_j) - P^\gamma_{\bar{\pi}}(e_j)| \leq \frac{2(K_f^{t_\infty} + k_f^{t_\infty - 1} + ... + 1) K_c \rho(n)}{(1-\beta)} \tag{3.30}$$

choose

$$N_\infty = max(N_1, N_2) \tag{3.31}$$

It follows from (3.24,3.26,3.30) and the triangle inequality

$$|\bar{P}_{\bar{\pi}}(e_j) - P_\pi(x)| \leq K^* \rho(n)\ \forall\, n > N_\infty \tag{3.32}$$

where $K^* = \frac{(K_s + 2(K_f^{t_\infty} + k_f^{t_\infty - 1} + ... + 1) + 1) K_c}{(1-\beta)}$.
q.e.d

Finally we make the following proposition:

Proposition 3.2 $\forall$ e_j *and* $x \in S_j$,

$$|\overline{P}^*(e_j) - P^*(x)| \leq 3K_\infty \rho(n) \quad \forall n \geq max(\eta, N_\infty) \tag{3.33}$$

Proof: We have that

$$\inf_\pi P_\pi(x) = P^*(x), \tag{3.34}$$

By definition , it follows that $\forall\, \delta > 0, \exists\, \pi^\delta \ni$

$$P_{\pi^\delta}(x) \leq P^*(x) + \delta, \tag{3.35}$$

Hence it follows that

$$|\bar{P}_{\bar{\pi}^\delta}(e_j) - P^*(x)| \leq |\bar{P}_{\bar{\pi}^\delta}(e_j) - P_{\pi^\delta}(x)| + \delta. \tag{3.36}$$

Let

$$\lambda(n) = K_\infty \rho(n). \tag{3.37}$$

Let π_n denote the optimal policy generated by running the dynamic programming algorithm on the discrete system with $S_d = (e_1, ..e_n)$. Hence by (A3.4),3.36 and Proposition3.1 it follows that for $n \geq min(\eta, N_\infty), \forall x \in S_j$

$$|\overline{P}_{\bar{\pi}^\delta}(e_j) - P^*(x)| \leq \lambda(n) + \delta \tag{3.38}$$

and

$$|\overline{P}^*(e_j) - P_{\pi_n}(x)| \leq \lambda(n). \tag{3.39}$$

Also, by definition, we have

$$\overline{P}^*(e_j) \leq \overline{P}_{\bar{\pi}^\delta}(e_j), \tag{3.40}$$

$$P^*(x) \leq P_{\bar{\pi}^*}(x). \tag{3.41}$$

(3.38), (3.39) $\Rightarrow$

$$\overline{P}_{\bar{\pi}^\delta}(e_j) - \lambda(n) - \delta \leq P^*(x) \leq \overline{P}_{\bar{\pi}^\delta}(e_j) + \lambda(n) + \delta, \tag{3.42}$$

$$\overline{P}^*(e_j) - \lambda(n) \leq P_{\bar{\pi}^*}(x) \leq \overline{P}^*(e_j) + \lambda(n). \tag{3.43}$$

(3.40) - (3.43) $\Rightarrow$

$$|\overline{P}^*(e_j) - \overline{P}_{\bar{\pi}^\delta}(e_j)| \leq 2\lambda(n) + \delta. \tag{3.44}$$

But

$$\begin{aligned} |\overline{P}^*(e_j) - P^*(x)| &\leq |\overline{P}^*(e_j) - \overline{P}_{\bar{\pi}^\delta}(e_j)| \\ &\quad + |\overline{P}_{\bar{\pi}^\delta}(e_j) - P^*(x)| \\ &\overset{3.38}{\leq} 3\lambda(n) + 2\delta. \\ &\overset{3.44}{} \end{aligned} \tag{3.45}$$

Note that the inequality (3.45) holds $\forall\ \delta > 0$. Hence

$$|\overline{P}^*(e_j) - P^*(x)| \leq 3\lambda(n) = 3K_\infty \rho(n) \quad \forall n \geq max(\eta, N_\infty) \tag{3.46}$$

which completes the proof of the proposition.
q.e.d

4 An illustrative example

For the purposes of illustration of the results presented in the previous sections we chose a simple orbital dynamics problem. The objective of the exercise is to put a satellite into a prescribed orbit around a massive body.

The equations of motion of the satellite around the massive body is governed by the following differential equations:

$$\dot{r} = v_r, \tag{4.1}$$

$$\dot{v_r} = v_\theta^2/r - k/r^2 + u_r, \tag{4.2}$$

$$\dot{v_\theta} = -v_r v_\theta/r + u_\theta, \tag{4.3}$$

where
r refers to the radial distance of the satellite from the massive body,v_r refers to its radial velocity around the massive body, v_θ represents the tangential velocity of the satellite and u_r and u_θ represent the control input to the system.Please refer to Greenwood[7] for further clarifications about the dynamics of the system. The sytem is further normalised so that k=1.
As can be seen from the equations of motion, the state space of the system is 3-dimensional. We discretise the state space by taking equispaced points along each axis and forming the corresponding grid.The continuous time dynamics of the sytem are discretised using a sample time of $\tau = 0.05(2\pi)$.The system is given a choice of two control actions: to take no control action or to give an impulse such that the satellite goes into a circular orbit at the current radius(this is achieved by resetting the states of the satellite so that it goes into a circular orbit). This kind of control action is imparted through a radial and/or a tangential thrust. The magnitude of the control is defined to be the square root of the sum of the squares of the radial and tangential thrusts. The circular orbit with radius one is the goal state and is given a zero cost. The steady state values of the state variables in this orbit are $r = 1,\ v_r = 0,\ and\ v_\theta = 1$. The incremental cost of being at a particular state at a given instant is defined as the sum of the square of the distance of the current state from the goal state and the square of the magnitude of the control action.The DP problem is solved under these conditions.

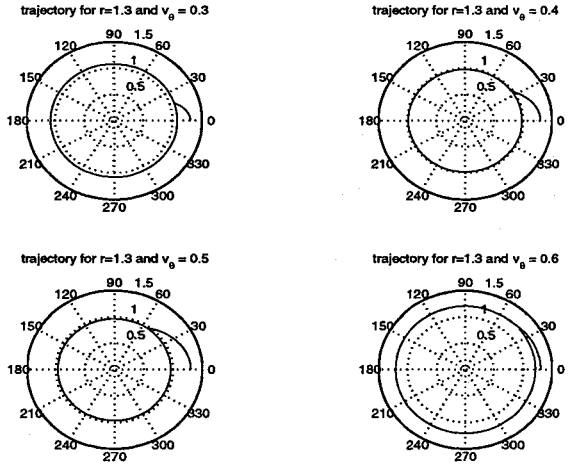

Figure 1: Optimal trajectories for the same initial v_θ but different initial radius

Let N denote the number of divisions along each axis i.e. we discretise r,v_r and v_θ into N discrete levels each.Hence note thatthe total number of states

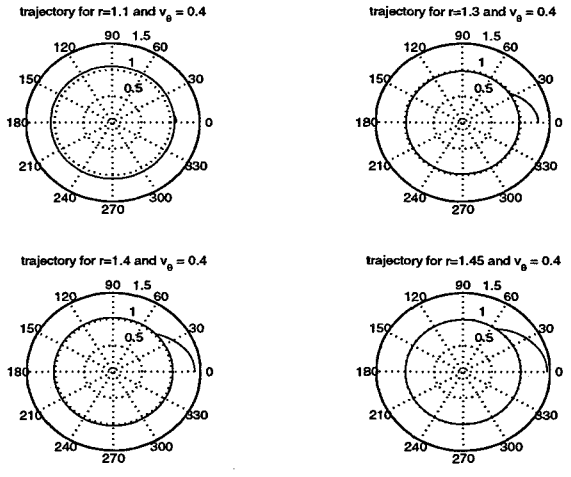

Figure 2: Optimal trajectories for same initial radius but different initial v_θ

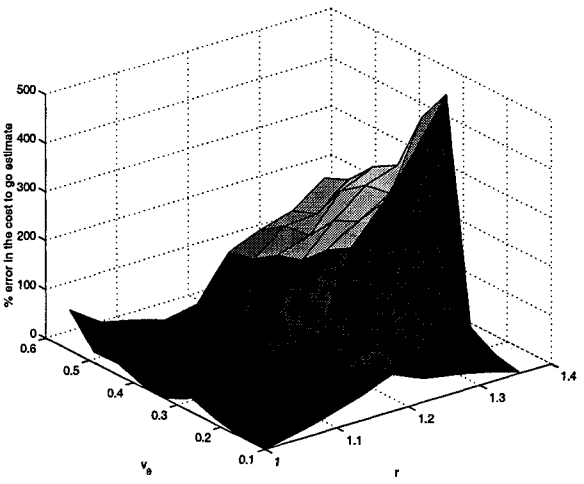

Figure 3: Surface plot of the estimation error in cost to go for 27 states

in the discrete system is going to be N^3. In fig.1 , the optimal trajectories learnt by the system are shown for a series of initial conditions with the same initial radius but different angular velocities. Fig.2 represents optimal trajectories for initial conditions with the same initial tangential velocity but different initial radius.Both these cases are for N= 5.As can be seen from the figures the optimal strategy is to take no control action till the unforced orbit crosses the target orbit when the system thrusts and resets the states of the sytem so that the satellite goes into orbit at the target radius. This tallies well with what is known about mimimum energy control in these cases. Also notice that the satellite doesn't go into orbit exactly at $r = 1$, this nebulousness of the orbit is due to the uncertainity that is introduced into the continuous system by the discretisation of the state space.

Figs. 3-6 represent the estimation error in the optimal cost to go for the case of N = 3 ,5 and 7 respectively. The optimal control policy for this example is to follow an unforced orbit till the target orbit is reached where the thrusters are fired to go into the target orbit. The actual cost- to-gos were calculated according to this policy. Fig.5 is a close up of fig.4 and helps to bring out the features of fig.4 better which are otherwise drowned by the scaling. As can be seen from the figures , there is a continuous improvement in the error plots from N = 3 to N = 7 as is to be expected. Please pay particular attention to the scales of the plots in order to appreciate the dramatic difference from a coarser grid to a finer one.Hence it is reasonable to expect that the errors become progressively lesser as the fineness of the grid increases, which is predicted by our results in section 3.

5 Conclusion

In this paper we have presented the worst case variant of the traditional DP operator and have discussed a few of its properties. We have also presented results which show that the optimal cost to go can be approximated arbitrartily well for deterministic systems by chosing a sufficiently fine degree of discretisation. However several issues reamain to be addressed , the logical next step being to extend these results to stochastic systems. Also, equispaced discretisation of the state space(as was attempted in the example) may not be enough to get over the curse of dimensionality and hence we should look towards finding an algorithm which would adaptively discretise the state space using some of the parameters in section3.We consider our work as a stepping stone in this direction.

References

[1] D.P.Bertsekas, Dynamic Programming: Deterministic and Stochastic models,Englewood cliffs,NJ:Prentice Hall,1987

[2] D P Bertsekas, J N Tsitsiklis, Neuro Dynamic Programming, Belmont,Mass:Athena Scientific

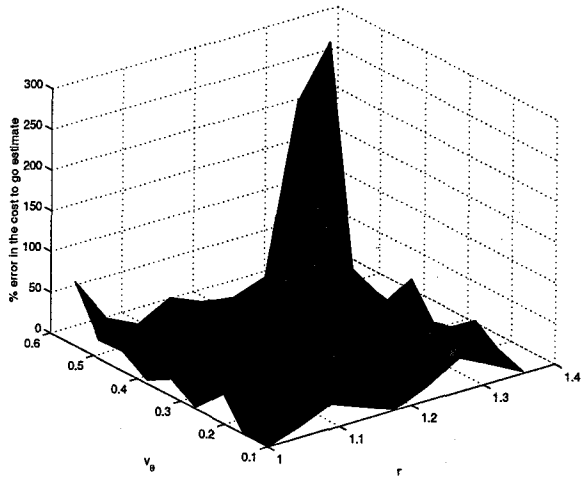

Figure 4: Surface plot of the estimation error in the cost to go for 125 states

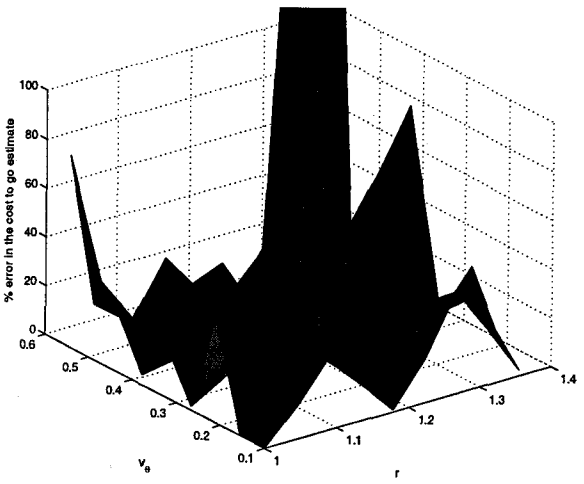

Figure 5: Close up of the estimation error for 125 states

[3] O. Hernandez-Lerma, Adaptive Markov Control Processes, New York:Springer Verlag,1989

[4] C S Chow , J N Tsitsiklis, "An optimal one-way multigrid algorithm for discrete-time stochastic control",IEEE transactions on Automatic Control,Vol36 no.8 Aug 1991,p 898-914

[5] W.Whitt,"Approximations of Dynamic Programs - 1", Matematics of Operations Research,vol.3,pp 231-243,1978

[6] J Rust, Handbook of computational Economics, vol.1,Elsevier,1996

[7] D T Greenwood, Principles of Dynamics,2nd ed., Englewood cliffs,NJ : Prentice Hall

[8] D G Luenberger, Optimisation by Vector Space Methods, John Wiley and sons

[9] H K Khalil, Nonlinear Systems, Upper Saddle River,NJ : Prentice Hall

[10] A M Bruckner, J B Bruckner, B S Thomson, Real Analysis, Upper Saddle River,NJ : Prentice Hall

[11] A N Shiryaev , Probability, Springer Verlag.

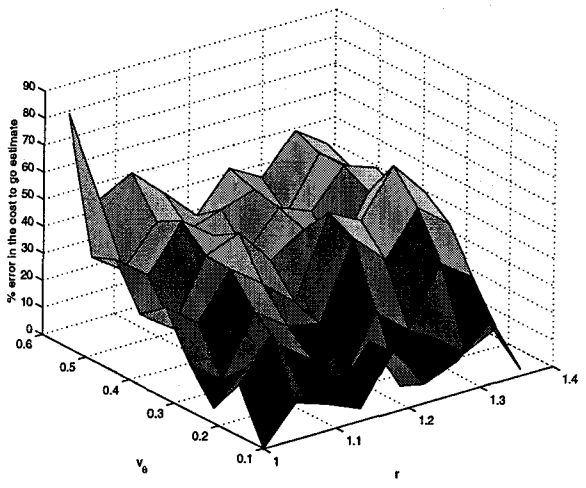

Figure 6: Surface plot of the estimation error in the cost to go for 343 states