The Optimal Search for a Markovian Target When the Search Path is Constrained: The Infinite-Horizon Case

Sumeetpal Singh and Vikram Krishnamurthy

***Abstract*—A target moves among a finite number of cells according to a discrete-time homogeneous Markov chain. The searcher is subject to constraints on the search path, i.e., the cells available for search in the current epoch is a function of the cell searched in the previous epoch. The aim is to identify a search policy that maximizes the infinite-horizon total expected reward earned. We show the following structural results under the assumption that the target's transition matrix is ergodic: 1) the optimal search policy is stationary, 2) there exists ϵ-optimal stationary policies which may be constructed by the standard value iteration algorithm in finite time. These results are obtained by showing that the dynamic programming operator associated with the search problem is a m-stage contraction mapping on a suitably defined space. An upper bound of m and the coefficient of contraction α is given in terms of the transition matrix and other variables pertaining to the search problem. These bounds on m and α may be used to derive bounds on suboptimal search polices constructed.**

***Index Terms*—Markovian target, optimal search, partially observed Markov decision process, stochastic shortest path problem.**

I. INTRODUCTION

A target moves among L cells according to a discrete-time homogeneous Markov chain. At discrete epochs of time $k \in \{1, 2, \ldots\}$, the searcher must choose an action from the set U. The set U may contain actions that search a particular cell or a group of cells simultaneously. Assuming action u is selected by the searcher at epoch k, it is executed with probability $1 - q(u)$. If the action cannot be executed, the searcher is said to be *blocked* for that epoch. This blocking situation models the scenario when the search sensors are a shared resource; the sensor could be shared between a number of searchers acting independently. If the searcher is not blocked and action u searches the cell that the target is in, the target is detected with probability $1 - \beta(u)$; failure to detect the target when it is in the cell searched is called an *overlook*. Additionally, the searcher is subject to constraints on the search path, i.e., if action $u_{k-1} = u$ was selected at epoch $k - 1$, then the action at epoch k must be selected from the set $U(u) \subset U$. The aim is to identify a search policy that minimizes the expected number of searches until detection or more generally, that maximizes the infinite-horizon total expected reward earned.

Reference [8] address the two-cell search problem for both the finite- and infinite-horizon case, but without blocking and search path constraints. Specifically, optimal search policies that: 1) maximize the probability of detecting the target in N attempts, and 2) minimize the expected number of searches for detection were constructed. For the simple two-cell scenario, the optimal search policy was solved for analytically for a number of special cases concerning the targets transition probabilities and the overlook probabilities. The L cell extension of [8] with search path constraints was addressed in [3] for criteria (1) and [6] for criteria (2). Let $J_\mu(\pi, j)$ denote the expected number of searches needed for detection when the target is initially distributed among the L according to π, the first action is to be selected from the set $U(j) \subset U$ and the search policy is μ. Let $J^*(\pi, j) = \inf_\mu J_\mu(\pi, j)$ denote the minimal expected number of searches for the pair (π, j) where the infimum is taken over the set of all possible search policies. It is shown in [1, Prop. 5.9] that J^* is the smallest nonnegative fixed point of Bellman's optimality equation (i.e., J^* may not be the only fixed point). Assuming U is finite, the *value iteration* (VI) algorithm converges pointwise to J^* for any initial iterate J_0 for VI provided $0 \le J_0 \le J^*$[1, Prop. 5.12 and Prop. 5.13]. With only pointwise convergence, one cannot derive meaningful bounds for sub-optimal policies constructed by iterating VI a finite number of times or more sophisticated procedures as in [5], nor can one assert the existence of ϵ-optimal stationary policies [1]. In [6], the author demonstrates the pointwise convergence of VI and how a careful choice of J_0 can lead to a closer approximation to J^* in less iterations.

In this note, we extend the search problem to the blocking scenario. Assuming that the transition probability matrix of the target is primitive, i.e., the matrix raised to some power l has all elements positive, we show that the dynamic programming (DP) operator for the search problem is a m-stage contraction mapping on a suitably defined space. We then give a (conservative) estimates of m and the coefficient of contraction α in terms of the transition matrix, the overlook and blocking probabilities. The m-stage contraction property enjoyed by the DP operator implies that ϵ-optimal stationary policies can be constructed in finite time using the VI algorithm, as is shown. Note that the search problem considered here uses the undiscounted infinite-horizon total expected reward performance criterion. For a discounted infinite-horizon performance criterion, uniform convergence of the VI algorithm is guaranteed irrespective of the structure of the Markov chain.

II. OPTIMAL CONSTRAINED SEARCH FOR A MARKOVIAN TARGET

Let $X = \{1, \ldots, L\} \cup \{T\}$, $Y = \{F(\text{found}), \bar{F}(\text{not found}), B(\text{blocked})\}$ and U be the search space, observation space and the action space, respectively. The search area of interest comprises of L cells. State T is a fictitious state that is added as a means of terminating the search upon detection. At the start of search epoch $k \in \{1, 2, \ldots\}$ the location of the target is $x_k \in X$ and an action $u_k \in U$ is adopted, for which an observation $y_k \in Y$ is received. The initial state of the target x_1 is distributed according to the probability distribution

$$\pi_1 \in \Pi(X) \triangleq \{\pi \in \mathbb{R}^{L+1} : \pi \ge 0, \sum_{k=1}^{L+1} \pi(k) = 1\}$$

with $\pi_1(T) = 0$. Let $\{U(u)\}_{u \in U}$ be an *arbitrary* family of subsets of U where each $U(u)$ is nonempty. The initial action u_1 is selected from an initial specified subset $U(j)$ where $j \in U$. For $k > 1$, u_k is selected from $U(u_{k-1})$. If the action u_k is blocked, then $y_k = B$. If the action u_k is not blocked and u_k searches the cell the target is in, then $y_k = \bar{F}$ with probability $\beta(u_k)$ and $y_k = F$ with probability $1 - \beta(u_k)$. The state of the target at the start of paging epoch $k + 1$, x_{k+1}, is characterized by the observation dependent transition probability matrices

$$\begin{aligned} P^F &= [\mathbf{0}_{(L+1)\times L}\ \mathbf{1}_{(L+1)\times 1}] \\ P^{\bar{F}} &= P^B = \begin{bmatrix} A & \mathbf{0}_{L\times 1} \\ \mathbf{0}_{L\times 1}^T & 1 \end{bmatrix} \end{aligned} \tag{1}$$

that is $\mathbf{P}(x_{k+1} = j | x_k = i, y_k = y) = P_{ij}^y$. ($\mathbf{1}$ and $\mathbf{0}$ represent matrices or vectors of 1's and 0's, respectively) The target moves among

Manuscript received April 3, 2002; revised September 26, 2002. Recommended by Associate Editor C. D. Charalambous.

S. Singh is Centre of Expertise in Networked Decision Systems (CENDS), Department of Electrical Engineering, University of Melbourne, Victoria 3010, Australia.

V. Krishnamurthy is with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: vikramk@ece.ubc.ca), and also with the University of Melbourne, Melbourne, Australia..

Digital Object Identifier 10.1109/TAC.2003.809165

the L cells according to sub-matrix A of (1) and transition to the absorbing state T occurs only when the target is detected. The law for the observation process is given by

$$\begin{aligned}\mathbf{P}(y_k = F|x_k = j, u_k = u) &= \begin{cases}(1-q(u))(1-\beta(u)) & \text{if } u \text{ searches cell } j\\ 0 & \text{otherwise}\end{cases}\\ \mathbf{P}(y_k = F|x_k = T, u_k = u) &= 1\\ \mathbf{P}(y_k = \bar{F}|x_k = j, u_k = u) &= \begin{cases}\beta(u)(1-q(u)) & \text{if } u \text{ searches cell } j\\ 1-q(u) & \text{otherwise}\end{cases}\\ \mathbf{P}(y_k = B|x_k = j, u_k = u) &= q(u) \end{aligned} \tag{2}$$

for all $u \in U$ and $j \in X\backslash\{T\}$. Note that observation F is always received when the target is in state T, regardless of the action taken.

In [8], [3], [6], the action space U is taken to be $X\backslash\{T\}$, i.e., there is an action corresponding to the search of each cell. We allow U to be arbitrary in the sense that $u \in U$ may correspond to the search of particular cell or a group of cells simultaneously. Without loss of generality, assume that

$$U = \{1, 2, \dots, K\}. \tag{3}$$

Following our comments on the arbitrary nature of actions in U, action $j \in U$ does not necessarily imply cell j is searched.

The following assumption, which holds throughout this note, effectively asserts that all actions in U are *useful*.

Assumption 2.1: For each $u \in U$, $q(u) < 1$, $\beta(u) < 1$, and $X_u \triangleq \{x \in X\backslash\{T\} : u \text{ searches cell } x\} \neq 0$.

If $q(u) = 1$ in (2), then action u is not useful as it is always blocked. Similarly, if $\beta(u) = 1$, then we will receive observation $\bar{F}$ even if the target is in the cell serached by u. Finally, $X_u = 0$ implies u does not search any of the L cells; for example, an action that corresponds to the suspension of search for one epoch.

Let $h(x_k, u_k)$ denote the instantaneous reward received for adopting action u_k, while the state of the target is x_k. Since the terminal state is fictitious, we restrict $h : X \times U \to \mathbb{R}$ so that

$$h(T, u) = 0 \qquad \forall u \in U. \tag{4}$$

Some examples of the choice of the instantaneous reward are as follows: 1) to minimize the expected number of searches for detection, set $h(x, u) = -1 \times I_{X\backslash\{T\}}(x)$; 2) to maximize the probability of detection, set $h(x, u) = \mathbf{P}(y_k = F|x_k = x, u_k = u) \times I_{X\backslash\{T\}}(x)$; and 3) to minimize the expected search cost for detection, set $h(x, u) = -c(u) \times I_{X\backslash\{T\}}(x)$ where $c : U \to (0, \infty)$ represents the cost associated with each action. Note that this formulation also captures the case when rewards depend on the current observation received: let $\bar{h} : X \times U \times Y \to \mathbb{R}$, then $h(x, u) \triangleq \sum_{y \in Y} \bar{h}(x, u, y)\mathbf{P}(y_k = y|x_k = x, u_k = u)$.

Let I_k denote the information (history) available at the start of paging epoch k and call it the *information vector*, i.e.,

$$I_k = \{u_1, y_1, \dots, u_{k-1}, y_{k-1}\} \qquad k > 1. \tag{5}$$

The *information state* at time k, which is denoted by π_k, is defined as the conditional probability distribution of the target state x_k given the available information I_k, $\pi_k(x) \triangleq \mathbf{P}(x_k = x|I_k)$ for $x \in X$. The information state can be computed recursively via Bayes' rule (also known as the hidden Markov model state predictor) as follows:

$$\pi_{k+1} = \tilde{\Phi}(\pi_k, u_k, y_k) \triangleq \frac{P^{y_k T}\tilde{Q}^{u_k}(y_k)\pi_k}{\mathbf{1}^T P^{y_k T}\tilde{Q}^{u_k}(y_k)\pi_k} \tag{6}$$

where $\tilde{Q}^u(y) \triangleq \text{diag}([\mathbf{P}(y_k = y|x_k = 1, u_k = u), \dots, \mathbf{P}(y_k = y|x_k = L, u_k = u), \mathbf{P}(y_k = y|x_k = T, u_k = u)])$. For convenience, we denote the denominator of $\tilde{\Phi}$ by $\tilde{\sigma}(y_k, \pi_k, u_k)$.

Let $\mathcal{U}$ denote the set of *admissible* search paging policies that are a function of the information state. A policy $\mu \in \mathcal{U}$ is a sequence $\mu = \{\mu_k\}_{k=1,2,\dots}$ where $\mu_k : \Pi(X) \times U \to U$ satisfies

$$\mu_k(\pi, u) \in U(u) \qquad \forall \pi \in \Pi(X), u \in U, k. \tag{7}$$

We denote by M the space of all such μ_k that satisfy (7). A policy $\mu \in \mathcal{U}$ is said to be *stationary* when $\mu_1 = \mu_2 = \dots$. Any given policy is executed during search as follows: the initial action is $u_1 = \mu_1(\pi_1, j)$. For $k > 1$, given the history I_k, compute π_k recursively using (6) and execute action $u_k = \mu_k(\pi_k, u_{k-1})$.

Let $(\Omega, \mathcal{F})$ be the underlying measurable space that is constructed in the usual manner, i.e., $\Omega \triangleq (X \times U \times Y)^\infty$ is the product space, which is endowed with the product topology and $\mathcal{F}$ is the corresponding product sigma-algebra [4]. For any $\pi_1 \in \Pi(X)$, initial subset $U(j)$ for u_1 and policy $\mu \in \mathcal{U}$, there exists a (unique) probability measure $\mathbf{P}^\mu_{\pi_1,j}$ on $(\Omega, \mathcal{F})$ satisfying certain *consistency conditions* concerning the initial state distribution, transition (1) and observation laws (2) as well as law generating the action process $\{u_k\}_{k=1,2,\dots}$ [cf. (7)]; see [4] for details. Let $\mathbf{E}^\mu_{\pi_1,j}$ denote the expectation with respect to the measure $\mathbf{P}^\mu_{\pi_1,j}$. For each initial distribution π_1, initial action constraint $U(j)$ and policy μ, the following infinite-horizon cost is associated:

$$J_\mu(\pi_1, j) \triangleq \lim_{N\to\infty} J^{(N)}_\mu(\pi_1, j) \tag{8}$$

where

$$J^{(N)}_\mu(\pi_1, j) \triangleq \mathbf{E}^\mu_{\pi_1,j}\left\{\sum_{k=1}^{N} h(x_k, u_k)\right\}. \tag{9}$$

The limits exists and is finite under Assumption 3.1 to follow. Let

$$J^*(\pi_1, j) \triangleq \sup_{\mu\in\mathcal{U}} J_\mu(\pi_1, j) \qquad \text{for } \pi_1 \in \Pi(X), j \in U. \tag{10}$$

Then, the aim is to determine a policy $\mu \in \mathcal{U}$ that satisfies

$$J_\mu(\pi_1, j) = J^*(\pi_1, j) \qquad \forall \pi_1 \in \Pi(X), j \in U. \tag{11}$$

Such a policy is called optimal.

We conclude this section by stating the DP recursion for calculating $J^{(N)}_\mu(\pi_1, j)$. Let $S \triangleq \Pi(X) \times U$ and let $\mathbf{B}(S)$ be the space of bounded functions $J : S \to \mathbb{R}$. $\mathbf{B}(S)$ is a Banach space under the supremum norm $\|J\|_\infty \triangleq \sup_{s\in S} |J(s)|$. Let $\mathbf{0}$ denote the constant function in $\mathbf{B}(S)$ with zero norm, i.e., evaluates to zero. Define the function $H : \Pi(X) \times U \times \mathbf{B}(S) \to \mathbb{R}$ as follows:

$$H(\pi, u, J) \triangleq \sum_{x\in X} h(x, u)\pi(x) + \sum_{y\in Y:\tilde{\sigma}(y,\pi,u)>0} \tilde{\sigma}(y, \pi, u) J(\tilde{\Phi}(\pi, u, y), u),\ \pi \in \Pi(X). \tag{12}$$

The DP recursion (Bellman's equation) [5] is

$$\begin{aligned} J^{(N)}_{\mu,N}(\pi, j) &\triangleq H(\pi, \mu_N(\pi, j), \mathbf{0})\\ J^{(N)}_{\mu,k}(\pi, j) &\triangleq H(\pi, \mu_k(\pi, j), J^{(N)}_{\mu,k+1}) \end{aligned} \tag{13}$$

for all $\pi \in \Pi(X)$, $j \in U$ and $k = 1, \dots, N-1$. Thus, $J^{(N)}_\mu(\pi_1, j) = J^{(N)}_{\mu,1}(\pi_1, j)$.

III. Constrained Search DP Operator—Contraction Property

In this section, we will establish that the constrained search problem is well defined in the sense that the limit in (8) exists and characterize the convergence properties of the VI algorithm. These results are established under Assumption 2.1 and the following assumption which will hold throughout this section.

Assumption 3.1: There exists some positive integer m *and* $\alpha \in (0,1)$ *such that* $\mathbf{P}^{\mu}_{\pi_1,j}\{x_{m+1} = T\} \geq \alpha$ *for all* $\pi_1 \in \Pi(X)$, $j \in U$ *and* $\mu \in \mathcal{U}$.

We have the following result that is due to the Markov structure in the dynamics of the search problem: for each $k \geq 1$, there exists some $\mu' \in \mathcal{U}$ such that

$$\mathbf{E}^{\mu}_{\pi_1,j}\{x_{(k+1)m+1} = T | I_{km+1}\} = \mathbf{E}^{\mu'}_{x_{km+1},u_{km}}\{x'_{m+1} = T\} \tag{14}$$

where $\mathbf{E}^{\mu'}_{\pi_{km+1},u_{km}}$ denotes expectation with respect to the probability measure on $(\Omega, \mathcal{F})$ that corresponds to the initial distribution for x_1, π_{km+1} [defined recursively from I_{km+1} using (6)], initial constraint set $U(u_{km})$ and policy μ'; see [4, p. 5] for details. μ' is essentially a *shifted version* of policy μ with shift proportional to k; see [4, p. 5] for details on construction of μ'. Assumption 3.1 and (14) give the result, shown in

$$\mathbf{P}^{\mu}_{\pi_1,j}\{x_{(k+1)m+1} \neq T\} \leq (1-\alpha)\mathbf{P}^{\mu}_{\pi_1,j}\{x_{km+1} \neq T\} \leq (1-\alpha)^{k+1} \tag{15}$$

that is exploited in the proof of Lemma 3.1. (The proof of (15), which is omitted, also uses the fact that a transition to state T follows once observation F is received (1), which implies $I_{\{x_{km+1} \neq T\}}$ is I_{km+1} measureable. For each $\mu \in M$, we define the mapping $T_\mu : \mathbf{B}(S) \to \mathbf{B}(S)$ by

$$T_\mu(J)(\pi, j) \triangleq H(\pi, \mu(\pi, j), J) \quad \forall \pi \in \Pi(X), j \in U. \tag{16}$$

We also define the (DP) map $T : \mathbf{B}(S) \to \mathbf{B}(S)$ by

$$T(J)(\pi, j) \triangleq \max_{u \in U(j)} H(\pi, u, J) \quad \forall \pi \in \Pi(X), j \in U. \tag{17}$$

For any $\mu = \{\mu_1, \mu_2, \ldots\} \in \mathcal{U}$, we denote by $(T_{\mu_1}T_{\mu_2}\ldots T_{\mu_k})$ the composition of the mappings $T_{\mu_1}, \ldots, T_{\mu_k}$, $k = 1, 2, \ldots$. It is obvious that $(T_{\mu_k}T_{\mu_{k+1}}\ldots T_{\mu_N})(\mathbf{0})(\pi, j) = J^{(N)}_{\mu,k}(\pi, j)$ for $k = 1, \ldots, N$.

Lemma 3.1: For every $\mu \in \mathcal{U}$, $\lim_{N\to\infty}(T_{\mu_1}T_{\mu_2}\ldots T_{\mu_N})(\mathbf{0})(\pi, j)$ exists and is a real number for each $\pi \in \Pi(X)$ and $j \in U$.

Proof: It is sufficient to show $\mathbf{E}^{\mu}_{\pi_1,j}\{\sum_{k=1}^{N} |h(x_k, u_k)|\} < d$, for some $d \in [0, \infty)$, for all $\pi \in \Pi(X)$, $j \in U$ and $N > 0$. Because state T is rewardless

$$\mathbf{E}^{\mu}_{\pi_1,j}\left\{\sum_{k=1}^{N} |h(x_k, u_k)|\right\} \leq \max_{x \in X, u \in U} |h(x, u)| \sum_{k=1}^{N} \mathbf{P}^{\mu}_{\pi_1,j}\{x_k \neq T\}.$$

Now, using (15), it follows that $\sum_{k=1}^{N} \mathbf{P}^{\mu}_{\pi_1,j}\{x_k \neq T\} \leq m/\alpha$. ■

The following lemma asserts that the mapping (16) is a contraction mapping on a suitably defined subset of $\mathbf{B}(S)$.

Lemma 3.2: Let $\bar{B} \triangleq \{J \in \mathbf{B}(S) : J([0, \ldots, 0, 1]^T, j) = 0$ for all $j \in U\}$. Then $\bar{B}$ is a closed subset of $\mathbf{B}(S)$. For all $J \in \bar{B}$, $\mu \in M$, the functions $T_\mu(J)$ and $T(J)$ belong to $\bar{B}$. Additionally, for every $\mu = \{\mu_1, \mu_2, \ldots\} \in \mathcal{U}$, $J, J' \in \bar{B}$

$$\|(T_{\mu_1}T_{\mu_2}\ldots T_{\mu_m})(J) - (T_{\mu_1}T_{\mu_2}\ldots T_{\mu_m})(J')\|_\infty \leq (1-\alpha)\|J - J'\|_\infty.$$

Proof: The fact that $\bar{B}$ is a closed set is established upon noting that any J that does not belong to $\bar{B}$ by definition cannot be an accumulation point of $\bar{B}$. The self-map property $T_\mu, T : \bar{B} \to \bar{B}$ follows since $H([0, \ldots, 0, 1]^T, u, J) = 0$ for all $u \in U$ and $J \in \bar{B}$; this in turn is true because of (4), $\tilde{\sigma}(F, [0, \ldots, 0, 1]^T, u) = 1$ and $\tilde{\Phi}([0, \ldots, 0, 1]^T, u, F) = [0, \ldots, 0, 1]^T$ for all $u \in U$. To show the contraction property, note that

$$(T_{\mu_1}T_{\mu_2}\ldots T_{\mu_m})(J)(\pi, j) = \mathbf{E}^{\mu}_{\pi,j}\left\{\sum_{k=1}^{m} h(x_k, u_k) + J(\pi_{m+1}, u_m)\right\}$$

where $\pi_2 = \tilde{\Phi}(\pi, u_1, y_1)$, $\pi_{k+1} = \tilde{\Phi}(\pi_k, u_k, y_k)$, $k = 2, \ldots, m$. So

$$\begin{aligned}
&|(T_{\mu_1}T_{\mu_2}\ldots T_{\mu_m})(J)(\pi, j) - (T_{\mu_1}T_{\mu_2}\ldots T_{\mu_m})(J')(\pi, j)| \\
&\leq \mathbf{E}^{\mu}_{\pi,j}\{|J(\pi_{m+1}, u_m) - J'(\pi_{m+1}, u_m)|\} \\
&= \mathbf{E}^{\mu}_{\pi,j}\{|J(\pi_{m+1}, u_m) - J'(\pi_{m+1}, u_m)| I_{\{x_{m+1} \neq T\}}\} \\
&\quad (\text{because } \pi_{m+1} = [0, \ldots, 0, 1]^T \\
&\quad \text{if } x_{m+1} = T) \leq (1-\alpha)\|J - J'\|_\infty.
\end{aligned}$$

■

Now we state the main result for the constrained search problem under Assumption 3.1. By proving Lemmas 3.1 and 3.2, we have shown that Assumption C of [2, Ch. 4] is satisfied by mappings T_μ and T for the constrained search problem. Thus, all the results proved in [2, Ch. 4] under Assumption C are applicable. Here, we quote the usual results of interest.

Proposition 3.1: The following results are true under Assumption 3.1.

i) For every $\mu \in \mathcal{U}$, $J_\mu = \lim_{N\to\infty}(T_{\mu_1}T_{\mu_2}\ldots T_{\mu_N})(J)$ for all $J \in \bar{B}$[2, Prop. 4.1 (a)].

ii) The mapping T *is an* m-stage contraction mapping with coefficient $1-\alpha$[2, Prop. 4.1 (c)]. The optimal reward function J^*(10) is the unique fixed point of T in $\bar{B}$, i.e., $\lim_{N\to\infty}\|T^N(J) - J^*\|_\infty = 0$ for all $J \in \bar{B}$[2, Prop. 4.2 (a,c)].

iii) A stationary policy $\mu^* = \{\mu, \mu, \ldots\} \in \mathcal{U}$ is optimal if and only if $T_\mu(J^*) = T(J^*)$. Furthermore, for every $\epsilon > 0$, there exists a stationary policy $\mu_\epsilon = \{\mu, \mu, \ldots\} \in \mathcal{U}$ such that $\|J^* - J_{\mu_\epsilon}\|_\infty \leq \epsilon$[2, Prop. 4.3 (a,c)].

(T^N denotes the composition of T with itself N times). ϵ-optimal stationary policies are constructed in the usual manner: let $J_\epsilon \in \bar{B}$ and $\mu \in M$ satisfy $\|J_\epsilon - J^*\|_\infty \leq \epsilon$ and $T_\mu(J_\epsilon) = T(J_\epsilon)$. Then, the policy $\mu_\epsilon = \{\mu, \mu, \ldots\}$ satisfies $\|J_{\mu_\epsilon} - J^*\|_\infty \leq 2\epsilon m/\alpha$. J_ϵ may be constructed using VI, i.e., $J_\epsilon = T^k(\mathbf{0})$ for large enough k where the value for k itself can be estimated using α and m.

We conclude this section by giving sufficient conditions for Assumption 3.1 to be satisfied. We define the following quantities first:

$$Q^u(y) \triangleq \text{diag}([\mathbf{P}(y_k = y | x_k = 1, u_k = u), \ldots \mathbf{P}(y_k = y | x_k = L, u_k = u)])$$

[cf. definition of $\tilde{Q}^u(y)$ in (6)]. Let $\min_{ij} A_{ij}$ denote the smallest element of the matrix A_{ij} in (1) and $\min_i (A^l)_{ij}$ denote the smallest element of column j of the matrix A^l. Let $\underline{Q}(F) \triangleq \min\{(Q^u(F))_{ii} > 0 : i \in X \setminus \{T\}, u \in U\}$ and $\underline{Q}(\bar{F}) \triangleq \min\{(Q^u(\bar{F}))_{ii} : i \in X \setminus \{T\}, u \in U\}$.

Proposition 3.2: Consider the constrained search problem under Assumption 2.1. The following conditions are sufficient for Assumption 3.1 to hold.

- If condition C1) is satisfied with $l = 1$, then $m = 2$ and $\alpha \geq \underline{Q}(F) \min_{ij} A_{ij}$.
- If condition C1) is satisfied with $l > 1$, then either C2) or C3) must hold and $m = l + 1$. If C2) is satisfied, then

$$\alpha \geq \left(\min_{u \in U} q(u)\right)^l \min_{u \in U} \mathbf{1}^T Q^u(F) \cdot [\min_i (A^l)_{i1}, \min_i (A^l)_{i2}, \ldots, \min_i (A^l)_{iL}]^T.$$

If C3) is satisfied, then

$$\alpha \geq \left(\underline{Q}(\bar{F})\right)^{2l} \min_{u\in U} \mathbf{1}^T Q^u(F) \cdot [\min_i (A^l)_{i1}, \min_i (A^l)_{i2}, \ldots, \min_i (A^l)_{iL}]^T.$$

C1) A in (1) is primitive, i.e., for some $l > 0$, $(A^l)_{ij} > 0$ for all i, j.
C2) In (2), $q(u) \in (0,1)$ for all $u \in U$.
C3) $\beta(u) \in (0,1)$ for all $u \in U$.

Proof: Case 1 C1) With $l = 1$:

$$\mathbf{P}^{\mu}_{\pi_1,j}(x_3 = T, y_2 = y', x_2 = j, y_1 = y, x_1 = i) = (P^{y'})_{jT}(\tilde{Q}^{u'}(y'))_{jj}(P^y)_{ij}(\tilde{Q}^u(y))_{ii}\pi_1(i)$$

where $u = \mu_1(\pi_1, j)$, $u' = \mu_2(\tilde{\Phi}(\pi_1, u, y), u)$. Let $u_2' = \mu_2(\tilde{\Phi}(\pi_1, u, \bar{F}), u)$, $u_2'' = \mu_2(\tilde{\Phi}(\pi_1, u, B), u)$ and $u_2''' = \mu_2(\tilde{\Phi}(\pi_1, u, F), u)$. Summing over the realizations $(i, y, j, y') \in XYXY$, we get

$$\begin{aligned}\mathbf{P}^{\mu}_{\pi_1,j}(x_3 = T) =& \mathbf{1}^T Q^{u_2'}(F) A^T Q^u(\bar{F})\tilde{\pi}_1 \\ &+ \mathbf{1}^T Q^{u_2''}(F) A^T Q^u(B)\tilde{\pi}_1 \\ &+ \mathbf{1}^T Q^u(F)\tilde{\pi}_1 + \pi_1(T)\end{aligned}$$

where $\tilde{\pi}_1 = [\pi_1(1), \ldots, \pi_1(L)]^T$. Note that $\mathbf{1}^T Q^u(F)\tilde{\pi}_1 = \mathbf{1}^T A^T Q^u(F)\tilde{\pi}_1 \geq \mathbf{1}^T Q^{u_2'''}(F) A^T Q^u(F)\tilde{\pi}_1$ since $Q^{u_2'''}(F)$ is a diagonal matrix with elements ≤ 1. Also, for any $u \in U$, by Assumption 2.1, there exists some i such that $(Q^u(F))_{ii} > 0$. So

$$\begin{aligned}\mathbf{1}^T Q^u(F) A^T \geq& (Q^u(F))_{ii}[A_{1i}, \ldots, A_{Li}] \\ \geq& (Q^u(F))_{ii} \min_{ij} A_{ij}\mathbf{1}^T \geq \underline{Q}(F) \min_{ij} A_{ij}\mathbf{1}^T.\end{aligned}$$

So

$$\begin{aligned}&\mathbf{P}^{\mu}_{\pi_1,j}(x_3 = T) \\ \geq& \underline{Q}(F) \min_{ij} A_{ij}\mathbf{1}^T(Q^u(\bar{F}) + Q^u(B) + Q^u(F))\tilde{\pi}_1 + \pi_1(T) \\ \geq& \underline{Q}(F) \min_{ij} A_{ij}.\end{aligned}$$

Case 2 C1) With $l > 1$ and C2): Only the sketch of the proof is given with details on algebraic manipulations omitted. Consider any $\pi_1 \in \Pi(X)$ with $\pi_1(T) \neq 1$ (otherwise $\mathbf{P}^{\mu}_{\pi_1,j}(x_{l+2} = T) = 1$)

$$\mathbf{P}^{\mu}_{\pi_1,j}(x_{l+2} = T) \geq \mathbf{P}^{\mu}_{\pi_1,j}(x_{l+2} = T | y_1 = B, \ldots, y_l = B) \cdot \mathbf{P}^{\mu}_{\pi_1,j}(y_1 = B, \ldots, y_l = B) + \pi_1(T).$$

$\mathbf{P}^{\mu}_{\pi_1,j}(y_1 = B, \ldots, y_l = B) \geq (\min_{u\in U} q(u))^l(1 - \pi_1(T))$. $\mathbf{P}^{\mu}_{\pi_1,j}(x_{l+2} = T | y_1 = B, \ldots, y_l = B) = \mathbf{P}^{\mu}_{\pi_1,j}(y_{l+1} = F | y_1 = B, \ldots, y_l = B)$ (because a transition to state T is only possible if $y_{l+1} = F) = \tilde{\sigma}(F, \pi_{l+1}, u_{l+1})$ where $\pi_{k+1} = \tilde{\Phi}(\pi_k, u_k, B)$ and $u_{k+1} = \mu_{k+1}(\pi_{k+1}, u_k)$ for $k = 1, \ldots, l$. Let $\tilde{\pi}_1 = [\pi_1(1), \ldots, \pi_1(L)]^T$. It is easy to show that $\pi_{l+1} = [\tilde{\pi}_1^T A^l / \mathbf{1}^T\tilde{\pi}_1 \; 0]^T$. So

$$\begin{aligned}\tilde{\sigma}(F, \pi_{l+1}, u_{l+1}) =& \mathbf{1}^T Q^{u_{l+1}}(F)(A^T)^l \frac{\tilde{\pi}_1}{\mathbf{1}^T\tilde{\pi}_1} \\ \geq& \mathbf{1}^T Q^{u_{l+1}}(F)[\min_i (A^l)_{i1}, \ldots, \min_i (A^l)_{iL}]^T \\ \geq& \min_{u\in U} \mathbf{1}^T Q^u(F)[\min_i (A^l)_{i1}, \ldots, \min_i (A^l)_{iL}]^T.\end{aligned}$$

Case 3 C1) With $l > 1$ and C3): Similar arguments to Case 2. Consider any $\pi_1 \in \Pi(X)$ with $\pi_1(T) \neq 1$

$$\mathbf{P}^{\mu}_{\pi_1,j}(x_{l+2} = T) \geq \mathbf{P}^{\mu}_{\pi_1,j}(x_{l+2} = T | y_1 = \bar{F}, \ldots, y_l = \bar{F}) \cdot (\mathbf{P}^{\mu}_{\pi_1,j}(y_1 = \bar{F}, \ldots, y_l = \bar{F}) + \pi_1(T)).$$

One may show $\mathbf{P}^{\mu}_{\pi_1,j}(y_1 = \bar{F}, \ldots, y_l = \bar{F}) + \pi_1(T) \geq (\underline{Q}(\bar{F}))^l$. Also, one may show

$$\mathbf{P}^{\mu}_{\pi_1,j}(x_{l+2} = T | y_1 = \bar{F}, \ldots, y_l = \bar{F}) \geq (\underline{Q}(\bar{F}))^l \cdot \min_{u\in U} \mathbf{1}^T Q^u(F)\,[\min_i (A^l)_{i1}, \ldots, \min_i (A^l)_{iL}]^T.$$

■

The basic idea in the proof of the case 2 (and similarly for case 3) is very simple. To establish that there is a minimum probability of termination, one shows that there is a positive probability that the first l actions will be blocked and this probability is independent of the initial distribution and the choice for the first l actions. After l transitions unobserved, the probability the target is in cell j is larger than $\min_i (A^l)_{ij}$. Thus, any action taken at epoch $l+1$ will yield a probability of termination larger than $\min_{u\in U} \mathbf{1}^T Q^u(F)[\min_i (A^l)_{i1}, \ldots, \min_i (A^l)_{iL}]^T$.

IV. Discussion

We have analyzed the constrained path search problem for general U and blocking by assuming that for any policy, the probability of not terminating diminishes at a rate of $1 - \alpha$ after m search epochs [cf. Assumption 3.1 and (15)]. We have shown that the mappings T_μ and T are m-stage contraction mappings on a suitably defined space with coefficient $1 - \alpha$; thus, value iteration converges to the optimal reward function in sup-norm. As shown in the proof of Lemma 3.1, the expected number of searches until termination is upper bounded by m/α. We then quoted in Proposition 3.1 results that are standard whenever the contraction mapping property holds. In Proposition 3.2, conservative estimates for m and α were given under an ergodicity assumption.

An equivalent formulation to ours is to express the constrained search problem as a *partially observed Markov decision process* (POMDP) (see [5] for a definition of a POMDP) by enlarging the state space to cope with the search path constraints; the POMDP will have the augmented process (u_{k-1}, y_{k-1}, x_k) as its state at time k. Once again, one may show that the DP operator for the POMDP is a contraction mapping on a suitable defined space. However, the enlargement of the state–space leads to less convenient expressions than presented here.

When Assumption 3.1 is not satisfied, one needs to impose an additional restriction on the reward structure. Assuming $h \leq 0$, then, using the techniques in [1, Ch. 5], it may be shown that J^* is a fixed point of T (not necessarily unique). Specifically, J^* is the largest nonpositive fixed point of T. [Note that the problem of minimizing the expected number of searches until detection is the same as maximizing J_μ when $h(x, u) = -1 \times I_{X\setminus\{T\}}(x)$]. Additionally, $T^k(J_0)$ converges only pointwise to J^* for any $J^* \leq J_0 \leq 0$. When $h \leq 0$, although the limit in (8) exists, it may be that $J_\mu(\pi, j)$ is equal to $-\infty$ for some policies μ and pairs (π, j). Thus, one may also have $J^*(\pi, j) = -\infty$. There is nothing much that can be done about the possibility of J_μ and J^* being extended real-valued functions in this general setting. One way to establish J^* is bounded below is to show that there exists at least one policy for which J_μ is bounded below, since by definition $J^* \geq J_\mu$.

A search problem can be cast into the framework of a *partially observed stochastic shortest path problem* (POSSP), and we refer to the reader to the recent work in [7] for the convergence properties of the VI algorithm for a POSSP. Note though that the work in [7] does not consider a setting with state-action (\search-path) constraints as we have here. In [7], the VI algorithms is shown to converge pointwise under a "weaker" assumption than Assumption 3.1; see [7, Ass. C]. We use "weaker" because while Assumption 3.1 need not be satisfied by all policies, any policy that fails to satisfy Assumption 3.1 must satisfy a divergence condition for the sequence of iterates generated by the DP recursion.

REFERENCES

[1] D. P. Bertsekas, *Dynamic Programming: Deterministic and Stochastic Models*. Upper Saddle River, NJ: Prentice-Hall, 1987.

[2] D. P. Bertsekas and S. E. Shreve, *Stochastic Optimal Control: The Discrete Time Case*. New York: Academic, 1978.

[3] J. N. Eagle, "The optimal search for a moving target when the search path is constrained," *Oper. Res.*, vol. 32, pp. 1107–1115, 1984.

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

[5] W. S. Lovejoy, "Computationally feasible bounds for partially observed Markov decision processes," *Oper. Res.*, vol. 39, no. 1, pp. 162–175, 1991.

[6] L. C. Thomas and J. N. Eagle, "Criteria and approximate methods for path-constrained moving-target search problems," *Naval Res. Log.*, vol. 42, pp. 27–38, 1995.

[7] S. D. Patek, "On partially observed stochastic shortest path problems," presented at the 40th IEEE Conf. Decision Control, Orlando, FL, 2001.

[8] S. M. Pollock, "A simple model of search for a moving target," *Oper. Res.*, vol. 18, pp. 883–903, 1970.

A Remark on Partial-State Feedback Stabilization of Cascade Systems Using Small Gain Theorem

Wei Lin and Qi Gong

***Abstract*—This note points out that input-to-state stability of zero dynamics having a continuously differentiable (instead of locally Lipschitz continuous) gain function suffices to guarantee the existence of globally stabilizing, smooth partial-state feedback control laws for cascade systems, *without imposing any extra condition*. This conclusion is proved via the small gain theorem and a novel variable separation technique combined with feedback domination design.**

***Index Terms*—Global stabilization, input-to-state stability (ISS), partial state feedback, small gain theorem.**

I. INTRODUCTION

In this note, we revisit the problem of global stabilization by *partial state* feedback for a class of cascade systems of the form

$$\begin{aligned}\dot{z} &= f_0(z, x_1)\\ \dot{x}_1 &= x_2 + f_1(z, x_1)\\ &\vdots\\ \dot{x}_{r-1} &= x_r + f_{r-1}(z, x_1, \ldots, x_{r-1})\\ \dot{x}_r &= u + f_r(z, x_1, \ldots, x_r)\end{aligned} \tag{1.1}$$

where $x = (x_1, \ldots, x_r) \in \mathbb{R}^r$ is the measurable state, $z \in \mathbb{R}^{n-r}$ the unmeasurable state and $u \in \mathbb{R}$ the control input, respectively. The functions $f_i : \mathbb{R}^{n-r+i} \rightarrow \mathbb{R}$, $i = 1, \ldots, r$, are C^1 with $f_i(0, \ldots, 0) = 0$, and $f_0 : \mathbb{R}^{n-r+1} \rightarrow \mathbb{R}^{n-r}$ is C^1 with $f_0(0, 0) = 0$. Throughout this note, it is assumed that (1.1) satisfies the following input-to-state-stable (ISS) condition.

Assumption 1.1: Suppose the system

$$\dot{z} = f_0(z, x_1)$$

with x_1 being the input and z the state, is ISS, i.e., there exists a K_∞ function $\gamma(\cdot)$ which is known such that the response $z(\cdot)$ to any bounded $x_1(\cdot)$ satisfies

$$\begin{aligned}\|z(t)\|_\infty &\leq \max\{\beta(\|z(0)\|, t), \gamma(\|x_1(\cdot)\|_\infty)\}\\ \lim_{t\rightarrow\infty} \sup \|z(t)\| &\leq \gamma(\lim_{t\rightarrow\infty} \sup \|x_1(t)\|)\end{aligned}$$

for some class KL function $\beta(\cdot, \cdot)$.

For the class of cascade systems (1.1) satisfying Assumption 1.1, global asymptotic stabilization by *partial state* feedback has received considerable attention. Indeed, the problem has been studied, for instance, in [1]–[4], under some *extra conditions* imposed on (1.1) such as growth hypotheses or gain-type matching conditions. These results were derived either by a Lyapunov-based design method combined with the idea of changing supply rate [7], or by using the small-gain theorem [3], [2] in a recursive manner. Note that both feedback design methods require certain ISS conditions on the z-subsystem of (1.1). Moreover, the Lyapunov-based design method needs to impose a sort of matching conditions between the driven system (i.e., z-subsystem) and the driving system (i.e., x-subsystem), while the small gain argument requires the crucial conditions iii)–iv) described in [2, Lemma 11.4.1] be fulfilled, as outlined in [2]. More specifically, it has been remarked in [2] that if at every step of the recursive design, the assumptions iii)–iv) of Lemma 11.4.1 are satisfied, a smooth virtual controller can be constructed in such a way that the resulted system is ISS and satisfies the small gain condition. To guarantee that the recursive design procedure can be carried out step by step, some extra conditions have been introduced. For example, in [3] it was assumed that the linearized system of the zero dynamics is asymptotically stable, i.e., the zero-dynamics must be *locally exponentially stable*. Recently, a consequence of [4] has indicated that a bit stronger ISS condition than Assumption 1.1 (basically, (GAS + LES)-type ISS condition) is enough for the solvability of the partial state feedback stabilization. More recently, the authors of [1] have proved that for a class of polynomial systems of the form (1.1), all the conditions of Lemma 11.4.1 can be rendered satisfied at each step by designing a suitable virtual controller. This, in turn, leads to the conclusion that global stabilization of the *polynomial* system (1.1) is solvable by partial state feedback under Assumption 1.1.

The purpose of this note is to point out that Assumption 1.1, together with the gain function $\gamma(\cdot)$ being C^1, suffices to guarantee the existence of a smooth partial-state feedback control law

$$u = u(x) = u(x_1, \ldots, x_r), \quad \text{with} \quad u(0, \ldots, 0) = 0 \tag{1.2}$$

such that the closed-loop system (1.1) and (1.2) is globally asymptotically stable at the equilibrium $(z, x) = (0, 0)$. That is, with Assumption 1.1, there is *no need to introduce any extra condition* such as those in [3], [4], and [1] for achieving global stabilization of the cascade system (1.1) via *partial state* feedback.

This conclusion will be proved in the next section, by effectively combining a small gain argument, the feedback domination design technique and a novel variable separation technique (Lemma 2.2). The proof that follows also offers a simplified design method which only needs to use explicitly the small-gain theorem once, rather than to use it repeatedly at every step of the recursive design.

Manuscript received April 19, 2002; revised November 8, 2002. Recommended by Associate Editor J. M. A. Scherpen. This work was supported in part by the National Science Foundation under Grants ECS-9875273 and DMS-0203387.

The authors are with the Department of Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, OH 44106 USA (e-mail: linwei@nonlinear.cwru.edu).

Digital Object Identifier 10.1109/TAC.2003.809166