Some Results on the Analysis of Stochastic Processes with Uncertain Transition Probabilities and Robust Optimal Control

Keyong Li[+*] likeyong@ieee.org
Seong-Cheol Kang[*] jsckang@bu.edu
Ioannis Ch. Paschalidis[*] yannisp@bu.edu
Center for Information and Systems Engineering,
Boston University,
Brookline, MA 02446.

***Abstract*— This paper investigates stochastic processes that are modeled by a finite number of states but whose transition probabilities are uncertain and possibly time-varying. The treatment of uncertain transition probabilities is important because there appears to be a disconnection between the practice and theory of stochastic processes due to the difficulty of assigning exact probabilities to real-world events. Also, when the finite-state process comes as a reduced model of one that is more complicated in nature (possibly in a continuous state space), existing results do not facilitate rigorous analysis.**

Two approaches are introduced here. The first focuses on processes with one terminal state and the properties that affect their convergence rates. When a process is on a complicated graph, the bound of the convergence rate is not trivially related to that of the probabilities of individual transitions. Discovering the connection between the two led us to define two concepts which we call "progressivity" and "sortedness", and to a new comparison theorem for stochastic processes. An optimality criterion for robust optimal control also derives from this comparison theorem. In addition, this result is applied to the case of mission-oriented autonomous robot control to produce performance estimate within a control framework that we propose.

The second approach is in the MDP frame work. We will introduce our preliminary work on optimistic robust optimization, which aims at finding solutions that guarantee the upper bounds of the accumulative discounted cost with prescribed probabilities. The motivation here is to address the issue that the standard robust optimal solution tends to be overly conservative.

I. Introduction

The theory of stochastic processes has produced many elegant results. However, there seems to be a disconnection between theory and practice[1], which is largely due to the difficulty of assigning exact probabilities to real-world events. This is particularly true in the case of autonomous robot control, which will be discussed later in this paper.

+ Research partially supported by the Air Force under grant F49620-02-1-0388 and from the NSF under grant ECS-0329743.

* Research partially supported by the NSF under grants EFRI-0735974, DMI-0330171, CNS-0435312, ECS-0426453, and by the DOE under grant DOE DE-FG52-06NA27490.

[1]There are some notable exceptions such as the quantitative approach to analyzing the financial market

For the most part of this paper, we consider processes with one terminal (absorbing) state and concentrate on their convergence rate. This is largely motivated by our study of mission-oriented autonomous robot control, in which the robot usually need to execute several steps, some enabling the others, and finally accomplishing the mission. A very relevant body of research include those on symbolic methods of control and Motion Description Languages. See [1], [2], [3], [7] and the collection of articles in [4]. For such systems, accomplishment of the mission can be modeled as a terminal state, and one can compute the risk of mission not accomplished by some deadline from the convergence rate.

Here, we assume only bounds of the transition probabilities are known. When the process is on a complicated graph, the bound of the convergence rate is not trivially related to that of the probabilities of individual transitions. That is, whether the convergence rate would increase or decrease when a particular transition probability increases is not clear in general when the other transition probabilities are only known up to a range. For instance, consider the graph in Figure 1, in which both nodes 2 and 3 may reach the goal directly. The figure can actually be misleading in this case. For this chain, it is possible that Stage 2, even Stage 1, is actually "closer" to the terminal stage than Stage 3. Then, the seemingly forward transitions (suggested by how the nodes are ordered) from nodes 1 and 2 to 3 are actually backward ones.

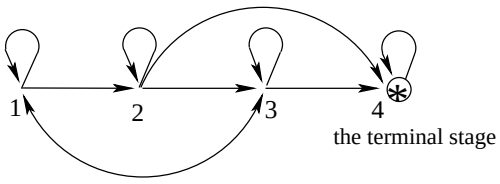

Fig. 1. A diagram of stages and transitions in a process in which the order of the stages may or may not be misleading.

To address this problem, we defined two new concepts which we call *progressivity* and *sortedness*. "More progressive" is a relation between probability transition matrices that is different from the commonly used elementwise "$\geq$" relation. "Sortedness" is on the other hand a property that certifies a transition matrix to serve as a reference for pro-

gressivity based comparison. The main theorem of this paper states that if a constant transition matrix $\hat{P}$ is sorted, then all processes whose transition matrices are more progressive than $\hat{P}$ for all time converges faster than the Markov process associated with $\hat{P}$. In addition, we will show that this result leads to an optimality criterion for robust optimal control.

Another topic discussed in this paper is optimistic robust optimization (ORO) of Markov Decision Processes (MDPs), again, with uncertain transition probabilities. Robust optimization is traditionally studied in the sense of optimizing the worst-case outcome, see [12], [13], [9], [6] for example. However, the optimal policy that concentrates on the worst cases can often be too conservative. With ORO, one may make use of available probabilistic information to come up with optimal solutions that focus on the probably scenarios rather than the worst-case scenarios. [10], [11] has studied the ORO of linear programming problems with uncertain constraints, in which the ORO optimal solution has a controlled level of probability of violating the constraints. The work reported here is an extension to the MDP case, although instead of having a chance of violating some constraints, the ORO solution here has a chance of giving an estimate of the accumulative cost that is lower than the actual outcome.

In what follows, the first topic mentioned above will be discussed in Section II, III, and IV. Among them, Section II introduces the concepts of "progressivity" and "sortedness", and proves the main theorem; Section III discusses the implication to robust optimization; and Section IV applies the results to mission-oriented autonomous robot control. The second topic of the paper (the MDP case) will be discussed in Section V.

II. A Comparison Theorem for the Time-to-Finish of Stochastic Processes with Uncertain and Time-Varying Transition Probabilities

Consider a stochastic system with $L+1$ possible states, where the state $L+1$ is absorbing. Let $\phi(k) = (\phi_1(k), \ldots, \phi_L(k))^T$, where $\phi_i(k),\ i = 1, \ldots, L$ is the probability of reaching state i at time k. Let $\psi(k)$ be the probability of reaching the absorbing state by time k. Let $P(k)$ be the transition probability matrix with the row and column corresponding to the absorbing state removed. This shall be understood when we mention probability transition matrices in what follows. Thus

$$\phi(k+1) = P(k)\phi(k), \quad \psi(k) = 1 - \mathbf{1}^T\phi(k). \qquad (1)$$

The matrix $P(k)$ is nonnegative and further sub-stochastic. Let's review some useful properties of such matrices first. For a sub-stochastic matrix P, the magnitude of its eigenvalues cannot exceed 1. Recall that a square nonnegative matrix is called *irreducible* if given any two indices $i, j \in \{1, \ldots, L\}$, there exists a sequence of indices $i_1, i_2, \ldots, i_l$ such that

$$p_{i_1 i} \cdot p_{i_2 i_1} \cdot p_{i_3 i_2} \cdots p_{i_l, i_{l-1}} \cdot p_{j, i_l} > 0.$$

Intuitively, this means that there is a possible chain of transitions connecting any pair of states. If P is further irreducible, then (see [8], Chapter 1, Theorems 4.1, 4.3 and 5.1)

(i) P has a unique real and positive eigenvalue r such that $r \geq |\lambda|$ for any other eigenvalue λ of P. This eigenvalue r is called the maximal eigenvalue of P. Note that a complex eigenvalue with the same magnitude, if it exists, is not referred to as a maximal eigenvalue.
(ii) The maximal eigenvalue of P has a positive eigenvector.
(iii) The maximal eigenvalue of P is greater than the maximal eigenvalue of any principal submatrix of P.

(Note: Properties (i) and (ii) are the main part of the well-known Perron-Frobenius Theorem. If P is reducible, then it still has a maximal eigenvalue, but it might have multiplicity greater than one and the associated eigenvectors are only guaranteed nonnegative.)

The transition matrix $P(k)$ is generally time-varying, and we do not assume that it is known exactly for any given k. Then, the question is: What do we need to know about $P(k)$ in order to make a statement about the time-to-finish distribution of the system. Previous works typically use elementwise bounds of $P(k)$. Although intuitive, that does not seem to be closely connected to the behavior of the process. Here, we define a different relation that is more relevant to the problem under study.

Definition 1 (Progressivity): Consider two $L \times L$ transition matrices P_1 and P_2. We say that P_2 is *no less progressive* than P_1, denoted by $P_2 \succeq P_1$, if

$$q^T(P_2 - P_1) \leq 0$$

for any vector $q = (q_1, q_2, \ldots, q_L)^T$ such that

$$q_1 > q_2 > \ldots > q_L > 0.$$

If $q^T(P_2 - P_1) < 0$ then P_2 is *more progressive* than P_1, denoted by $P_2 \succ P_1$.

The relation "no less progressive" is a partial order of the set of sub-stochastic matrices. Intuitively, being no less progressive means having an equal or greater tendency of transiting to higher-index states. Indeed, if from each state, the transition probability to some state is reduced and the same amount is added to that of a higher-index state, then the transition matrix P becomes more progressive.

Example 1: Consider

$$P_1 = \begin{pmatrix} 1-p_1 & 1-p_2 & 1-p_3 \\ p_1 & 0 & 0 \\ 0 & p_2 & 0 \end{pmatrix},$$

$$P_2 = \begin{pmatrix} 1-p_1-\epsilon_1 & 1-p_2 & 1-p_3-\epsilon_3 \\ p_1+\epsilon_1 & 0 & 0 \\ 0 & p_2-\epsilon_2 & \epsilon_3 \end{pmatrix},$$

and

$$P_3 = \begin{pmatrix} 1-p_1-\epsilon_1 & 1-p_2 & 1-p_3-\epsilon_3 \\ p_1+\epsilon_1 & 0 & 0 \\ 0 & p_2 & 2\epsilon_3 \end{pmatrix},$$

where $\epsilon_1, \epsilon_2, \epsilon_3$ are small positive numbers such that P_1, P_2, P_3 are still sub-stochastic. One may verify using the definition that $P_2 \succ P_1$ and P_3 cannot be compared with P_1 or P_2 in terms of progressivity.

The "no less progressive" relation is a linear inequality of the elements of the transition matrices, as shown by the following lemma.

Lemma 1: Let

$$Q = \begin{pmatrix} 1 & 1 & \dots & 1 & 1 \\ 1 & 1 & \dots & 1 & 0 \\ & & \dots & & \\ 1 & 0 & \dots & 0 & 0 \end{pmatrix}_{L\times L}.$$

(i) $P_2 \succeq P_1$ if and only if $Q(P_2 - P_1) \leq 0$.
(ii) If $P_2 \succ P_1$, then $Q(P_2-P_1) \leq 0$ and $Q(P_2-P_1) \neq 0$.
(iii) If $Q(P_2 - P_1) < 0$, then $P_2 \succ P_1$.

(The matrix inequalities are in the elementwise sense.)

Proof: For the "only if" part of item (i), note that each row of Q is a vector on the boundary of the set of vectors whose elements are positive and in descending order. Then, from the definition of "$\succeq$" and using the continuity of $q^T(P_2 - P_1)$ in q, $Q(P_2 - P_1) \leq 0$ must hold.

For the "if" part of item (i), one only needs to verify that any L dimensional vector whose elements are positive and in descending order can be expressed as a linear combination of the rows of Q with positive coefficients.

For item (ii), first note that Q is invertible. Thus, $Q(P_2 - P_1) \neq 0$, otherwise P_1 and P_2 would be identical. The rest of item (ii) follows directly from the "only if" part of (i), since "$\succ$" is strictly stronger than "$\succeq$". Item (iii) holds for the same reason as the "if" part of (i). ■

One may expect a system to proceed to the absorbing state no slower if its probability transition matrix becomes no less progressive. However, this is not necessarily the case.

Example 2: Consider again the graph depicted in Figure 1. A possible transition matrix of this system is

$$P = \begin{pmatrix} 0.1 & 0 & 0.1 \\ 0.9-\epsilon & 0.1 & 0 \\ \epsilon & 0.1 & 0.1 \end{pmatrix},$$

where $\epsilon > 0$ is a parameter. Clearly, P is no less progressive for greater ϵ. However, the maximal eigenvalue of P increases when ϵ increases. That is, the probability of reaching the absorbing state approaches 1 slower for greater ϵ.

A careful examination of the example illustrates a general problem. Namely, the state 2 is actually "closer" to the absorbing state than the state 3 is. For cases with a more complicated graph of transitions, inspection can become difficult. To establish a comparison, we need an additional condition.

Definition 2 (Sortedness): Let P be a $L \times L$ irreducible nonnegative matrix with maximal eigenvalue r and associated positive left eigenvector $q = (q_1, q_2, \dots, q_L)^T$. We say that P is *sorted* if $q_1 > q_2 > \dots > q_L$, or semi-sorted if $q_1 \geq q_2 \geq \dots \geq q_L$.

Remark: From the Perron-Frobenius Theorem (properties (i) and (ii) of irreducible nonnegative matrices discussed above), every irreducible nonnegative matrix can be rearranged into a sorted matrix by simultaneous row/column permutation.

The significance of *sortedness* will become apparent in light of the following theorem.

Theorem 1: Consider process (1). Let $\hat{P}$ be a sorted substochastic matrix of the same dimensions as $P(k)$, and let r be the maximal eigenvalue of $P(k)$.

(i) If $P(k) \succeq \hat{P}$ for all $k = 1, 2, \dots, \infty$, then

$$\exists \beta, \quad 1 - \psi(k) < \beta \cdot r^k, \quad \text{as} \quad k \to \infty.$$

(ii) Further, if $P(k) \succ \hat{P}$, then

$$(1 - \psi(k)) \cdot r^{-k} \to 0 \quad \text{as} \quad k \to \infty.$$

(iii) If $\hat{P} \succ P(k)$, then

$$(1 - \psi(k)) \cdot r^{-k} \to \infty \quad \text{as} \quad k \to \infty.$$

Proof: Let $0 < r \leq 1$ be the maximal eigenvalue of $\hat{P}$ and q be the associated positive left eigenvector. Let

$$V(k) = r^{-k} q^T \phi(k).$$

Then,

$$\begin{aligned} V(k+1) - V(k) &= q^T(r^{-k-1}\phi(k+1) - r^{-k}\phi(k)) \\ &= r^{-k-1} q^T (P(k)\phi(k) - r\phi(k)) \\ &= r^{-k-1} q^T (P(k) - \hat{P})\phi(k). \end{aligned}$$

By virtue of $\hat{P}$ being sorted, the elements of q satisfy $q_1 > q_2 > \dots > q_L > 0$.

For the first claim, if $P \succeq \hat{P}$, then $q^T(P - \hat{P}) \leq 0$. Because $\phi(k)$ is always nonnegative, we have $V(k+1) - V(k) \leq 0$. By definition, $V(k) \geq 0$, so $V(k)$ converges to a bounded value for $k \to \infty$.

Let $\sigma(k) = r^{-k} \mathbf{1}^T \phi(k)$, then

$$V(k)/q_1 \leq \sigma(k) \leq V(k)/q_L.$$

Thus $\sigma(k)$ also converges to a finite value (possibly zero) as $k \to \infty$, say σ_∞. Then, $\mathbf{1}^T \phi(k) \sim \sigma_\infty \cdot r^k$ for $k \to \infty$. That is, $1 - \psi(k) \sim \sigma_\infty \cdot r^k$.

For the second claim, if $P \succ \hat{P}$, then $q^T(P-\hat{P}) < 0$. Let $\alpha > 0$ be the smallest element of $-q^T(P-\hat{P})$, then

$$\begin{aligned} V(k+1) - V(k) &\leq r^{-k-1} \cdot (-\alpha)\mathbf{1}^T\phi(k) \\ &< (-\alpha)\mathbf{1}^T\phi(k) \\ &< -\tfrac{\alpha}{q_L}V(k). \end{aligned}$$

Thus, $V(k)$ decays to zero at least as fast as $(1-\alpha/q_L)^k$, and so does $\sigma(k)$. Then $\sigma_\infty = 0$.

Similarly for the third claim, if $\hat{P} \succ P$, then $V(k)$, $\sigma(k) \to \infty$. Consequently, $(1-\psi(k)) \cdot r^{-k} \to \infty$ as $k \to \infty$. ■

Remarks:

1) One may wonder whether we could have obtained the same result if we defined "more progressive" as: $P_2 \succ P_1$ if for any vector q whose elements are positive and in descending order, $q^T(P_2 - P1) \leq 0$ and $\neq 0$. The answer is negative. To give an example, consider

$$\hat{P} = \begin{pmatrix} 0 & 0.8 \\ 0.7 & 0 \end{pmatrix},$$

$$P(k) = \begin{cases} \begin{pmatrix} 0 & 0.6 \\ 0.7 & 0.2 \end{pmatrix} & : \ k \text{ is even}, \\ \begin{pmatrix} 0 & 0.8 \\ 0.9 & 0 \end{pmatrix} & : \ k \text{ is odd}, \end{cases}$$

and $\phi(0) = (1,\ 0)^T$. One may verify that $\phi(k)$ would be the same whether the probability transtion matrix is $P(k)$ or $\hat{P}$, while $P(k)$ would be more progressive than $\hat{P}$ under the alternative definition.

2) The connection between progressivity and sortedness plays a central role in the above proof. These two concepts seem to be fundamental. In particular, more implications of the sortedness property seem worthy of further investigation. As a first step of this investigation, we will explore the connection between sortedness and robust optimal policies in Section III.

III. Robust Optimal Policy

Consider the following decision problem: Given a system

$$x(k+1) = H(k, x(k), u(k), w(k)),\ x(0) = x_0, \quad (2)$$

where the state $x \in \mathbb{R}^n$; the control u is in a command set $U = \{\mu_1, \mu_2, \ldots, \mu_L\}$; and w is a disturbance process in a probability space $(\Omega, \mathcal{F})$ with probability measure $\Pr[\cdot] : \mathcal{F} \mapsto [0,1]$. Denote the control policy by

$$u = h(x) : \mathbb{R}^n \mapsto U. \quad (3)$$

Suppose the command μ_i is admissible only when $x \in \mathcal{X}_i \subseteq \mathbb{R}^n,\ i = 1, \ldots, L$. Consider the problem of maximizing the probability of driving the system state into a target domain in the state space, denoted by $\mathcal{X}_{L+1}$, as time passes.

let $s(k)$ denote the index of the command issued at time k. Then, given a control policy and the initial state, $s(k)$ is a random process over $\{1, \ldots, L+1\}$, with $L+1$ being an absorbing state. We will call $s(k),\ k = 1, 2, \ldots$ the *command process*. The probability transition of $s(k)$ can be written in the form of (1).

Here we do not assume the probability measure of the disturbance process is known exactly. Instead, we will only assume that $\Pr[\cdot]$ falls within a collection of probability measures $\mathcal{M}$, such that the probability transition matrix of the process $s(k)$ is no less progressive than some matrix. This assumption can be satisfied if under each command μ_i and for any domain $D \subset \mathbb{R}^n$, we can obtain two bounds $\check{p}$ and $\hat{p}$, and

$$\check{p} \leq \Pr[H(k, x, \mu_i, w) \in D | x] \leq \hat{p},$$

given any $x \in \mathcal{X}_i$ and for all $k \geq 0$.

Consider the problem of designing a robust optimal policy to maximize the minimum rate of convergence of the system towards the domain $\mathcal{X}_{k+1}$. Formally, the problem can be written as

$$\begin{aligned} \max_{h:\mathbb{R}^n \mapsto U} \quad & a \\ \text{s. t.} \quad & \exists \beta,\ \lim_{k\to\infty} \Pr[x(k) \notin \mathcal{X}_{L+1}] e^{ak} \leq \beta, \\ & \forall \Pr[\cdot] \in \mathcal{M} \text{ and } \forall x_0, \end{aligned} \quad (4)$$

where the process $x(k)$ is defined by (2) and (3).

One feasible policy is the following:

$$\begin{aligned} u = h^*(x) = \mu_i \\ \text{if } x \in S_i = \{x \in \mathcal{X}_i | x \notin \mathcal{X}_j,\ j = i+1, \ldots, L+1\}. \end{aligned} \quad (5)$$

One may verify that $S_1, S_2, \ldots, S_{k+1}$ is a partition of $\mathbb{R}^n$. This policy is thus well-defined. In plain words, this policy takes the ordering of the commands in the command set as an indication of relative priority, and issues a certain command if firstly, it is admissible; and secondly, no command with a higher priority (a higher index) is admissible. This policy makes intuitive sense, and it can indeed be the optimal solution of (4). We will prove a criterion for the optimality of policy (5) in the rest of this section.

Lemma 2: Suppose the probability measure of $\Pr[\cdot]$ is exactly given, and $\Pr[H(k, x, \mu_j, w(k)) \in S_i | x = \xi] = p_{i,j}$ uniformly for all $\xi \in \mathcal{X}_j$ and $k = 0, 1, \ldots$, $i, j = 1, \ldots, L$. If $P = \{p_{i,j}\}$ is sorted, then the policy (5) is the optimal solution of

$$\begin{aligned} \max_{h:\mathbb{R}^n \mapsto U} \quad & a \\ \text{s. t.} \quad & \exists \beta,\ \lim_{k\to\infty} \Pr[x(k) \notin \mathcal{X}_{L+1}] e^{ak} \leq \beta. \end{aligned} \quad (6)$$

Proof: Consider the command process. The hypothesis of the lemma says that under the policy (5), the transition probabilities of this process are given by P for all time. Consider an alternative strategy such that for $i = 1, \ldots, L$,

$u(k)$ takes the value of μ_i when $x(k) \in S'_i$. (By definition, $S'_{L+1} = S_{L+1}$.) For the control strategy to be well-defined, $S'_1, S'_2, \ldots, S'_{L+1}$ must also be a partition of $\mathbb{R}^n$. At the same time, because of the admissibility constraints, $S'_i \subseteq \mathcal{X}_i$ for each index i. The change from the original partition to the new partition can be decomposed into a series of changes that each involve the domains of two actions. Suppose one of the changes in this series is to remove a nonempty set ΔS from the domain of action i_0 and add it to the domain of action $i_1, i_1 \neq i_0$. Then, $\Delta S \subseteq S_{i_0} \cap S'_{i_1}$. Note that S_{i_0} does not intersect $\mathcal{X}_i$ for $i = i_0 + 1, \ldots, L+1$, and S'_{i_1} is in $\mathcal{X}_{i_1}$. Then, for $S_{i_0} \cap S'_{i_1}$ to be nonempty, $i_1 < i_0$ must hold. A consequence of the above change is to render the probability transition matrix of the process $s(k)$ less progressive in terms of the current ordering of the actions. Because the less progressive relation is transitive (as a part of being a partial order), a series of changes like this results in an alternative strategy whose transition matrix is also less progressive than P. This is true for all alternative strategies. In addition, since P is sorted, the optimality of policy (5) follows from Theorem 1. ■

Theorem 2: Let $P(k)$ be the probability transition matrix of the command process $s(k)$. Suppose for any initial state x_0, time $k \geq 0$, and any $\Pr[\cdot] \in \mathcal{M}$, $P(k)$ is no less progressive than a constant matrix P^* when policy (5) is applied, with $P(k) \equiv P^*$ being one possibility. If P^* is sorted, then policy (5) is an optimal solution of the problem (4).

Proof: Denote the maximal eigenvalue of P^* by r^*, and let $a^* = -\ln r^*$. Then, from Theorem 1, there exists a β such that $Pr[x(k) \notin \mathcal{X}_{L+1}] \leq \beta e^{-a^* k}$ for $k \to \infty$ and all $\Pr[\cdot] \in \mathcal{M}$. This shows that the pair $(h^*(\cdot), a^*)$ is a feasible solution of (4).

On the other hand, let $\Pr^*[\cdot]$ be a probability measure such that $P(k) \equiv P^*$. That is, $\Pr[H(k, x, \mu_j, w(k)) \in S_i | x \in \mathcal{X}_j] = p^*_{i,j}$ uniformly for all $x \in \mathcal{X}_j$ and $k = 0, 1, \ldots$, $i, j = 1, \ldots, L+1$. $\Pr^*[\cdot]$ is then a member of $\mathcal{M}$. Lemma 2 says that if $\Pr^*[\cdot]$ is indeed the probability measure of the disturbance process, then no other policy can produce a higher convergence rate than a^*. Note that $\Pr^*[\cdot]$ is the worst-case probability measure for $h^*(\cdot)$. Thus, $(h^*(\cdot), a^*)$ is optimal. ■

IV. Mission Oriented Autonomous Robot Control

In this section, we consider the problem of mission oriented autonomous robot control as another application of Theorem 1. The design solutions to such problems are often hierarchical. Let's consider a two-layer model, where the lower layer produces the so-called elementary actions (EAs) together with their preconditions, postconditions and assessments of their minimum probability for "success"; and the upper layer produces a mission-accomplishing policy through ordering the EAs.

Here, an EA is a 5-tuple $(f, \varphi, \psi, \hat{p}, \Theta)$. Let $x \in \mathbb{R}^n$ be the state of the robot's world. The function $f(x)$ is a lower-level control law that produces the signals directly fed to the actuators. Denote the state evolution under f by $x(t+\tau) = F(t, x(t), w, \tau)$, where w is a random process. (Here, we use t to denote time in stead of k for reasons that will become clear later.) The functions $\varphi : \mathbb{R}^n \mapsto \{\mathsf{true}, \mathsf{false}\}$ and $\psi : \mathbb{R}^n \mapsto \{\mathsf{true}, \mathsf{false}\}$ correspond to *precondition* and *postcondition*. The precondition $\varphi(x)$ equals true if this EA is admissible at x, and false otherwise. The postcondition $\psi(x)$ indicates the nominal result that the EA is designed to achieve, with $\psi(x) = \mathsf{true}$ when x is in the nominal range of the terminal state of this EA. The number $\hat{p}$ is a probability, and the number Θ is a finite duration in time. Taken together, the execution of the control law f with initial state in $\varphi^{-1}(\mathsf{true})$ satisfies[2]:

1) The system will not go to ruin under f, where "ruin" is defined as a domain $\mathcal{R}$, a subset of $\mathbb{R}^n$ that is disjoint from both $\varphi^{-1}(\mathsf{true})$ and $\psi^{-1}(\mathsf{true})$, and such that when x enters $\mathcal{R}$, it becomes infeasible for any control law to drive x out of $\mathcal{R}$.
2) The system state x will enter $\psi^{-1}(\mathsf{true})$ from $\varphi^{-1}(\mathsf{true})$ in less than Θ units of time with a probability greater than $\hat{p}$. More precisely, for all t,

$$\Pr\left(\begin{array}{l}\exists \delta \in (0, \Theta), \\ \varphi(F(t, x(t), w, \tau)) = \mathsf{true} \text{ for } 0 \leq \tau < \delta, \\ \text{ and } \psi(F(t, x(t), w, \delta)) = \mathsf{true}\end{array}\right) > \hat{p}. \tag{7}$$

In plain words, when admissible, an EA is safe and has a chance to succeed in some given time. We will call $\hat{p}$ the *minimum success probability*, Θ the *nominal duration*, and $\varphi^{-1}(\mathsf{true})$ the *admissible domain* of this EA.

The system dynamics treated at the upper layer can be expressed in the form of (2):

$$x(t+1) = H(t, x(t), u(t), w(t)), \ x(0) = x_0, \tag{8}$$

where the command set U becomes a set of EAs — $\mathrm{EA}_i = (f_i, \varphi_i, \psi_i, \hat{p}_i, \Theta_i), \ i = 1, \ldots, L+1$. Then,

$$H(t, x, \mathrm{EA}_i, w) \equiv F_i(t, x(t), w, 1). \tag{9}$$

The admissible domain of the ith command is $\mathcal{X}_i = \varphi_i^{-1}(\mathsf{true})$. Assume the mission of the robot is characterized by rendering $\varphi_{L+1} = \mathsf{true}$. Recalling that $\mathcal{R}$ is the domain of ruin, assume for all $x \notin \mathcal{R}$, the precondition of at least one EA is true. The control policy in the form of (5) can be applied:

$$\begin{array}{l} h(x) = f_i(x) \quad \text{if} \quad \varphi_i(x) = \mathsf{true}, \text{ and} \\ \varphi_{i+1}(x) = \ldots = \varphi_L(x) = \varphi_{L+1}(x) = \mathsf{false}. \end{array} \tag{10}$$

[2]Here $\varphi^{-1}(\mathsf{true})$ and $\varphi^{-1}(\mathsf{false})$ denote the preimage of true and false, respectively, although φ^{-1} as a function might not exist. The same is true for ψ^{-1}.

There is a difference however: the probabilistic information assumed for the EAs is not in terms of the transitions of each step. This formulation is natural for autonomous robot control, because each EA may have a chance to reach its goal only after running for some time. But we can still apply Theorem 1 by rescaling time.

Theorem 3: Consider the policy (10). If for each index $i = 1, \ldots, L$, there is a $i' > i$ such that $\psi_i(x) = \texttt{true}$ implies $\varphi_{i'}(x) = \texttt{true}$, then this mission is accomplished w.p.1 for all initial conditions not in the ruin set $\mathcal{R}$. Further, the probability of mission-not-accomplished by time t, as $t \to \infty$ is less than

$$r^{t/\Theta_M},$$

where $0 < r < 1$ is the maximal eigenvalue of the matrix

$$\hat{P} = \begin{pmatrix} 1-\hat{p}_1 & 1-\hat{p}_2 & 1-\hat{p}_3 & \cdots & 1-\hat{p}_k \\ \hat{p}_1 & 0 & 0 & \cdots & 0 \\ 0 & \hat{p}_2 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & & \cdots & \hat{p}_{k-1} & 0 \end{pmatrix},$$

and $\Theta_M = \max\{\Theta_1, \Theta_2, \ldots, \Theta_L\}$.

Proof: Given a policy, the system (8) generates a symbolic sequence $\sigma(k) : \mathbb{R}^+ \mapsto \{1, 2, \ldots, L, L+1\}$, which records the EAs actually taken to accomplish the mission. (This sequence is somewhat different from the command process, but related.) When the mission is started, $\sigma(0)$ is generated to record the first EA applied. Each next symbol is generated when the policy switches to a different EA, or the EA currently applied reaches its nominal duration. Mission accomplishment is recorded as $\sigma(k) = L + 1$, and is an absorbing state of the sequence. Let t_k be the time when the kth symbol is generated. Then,

$$\frac{t_k}{k} \leq \Theta_M. \tag{11}$$

The probability transition of $\sigma(k)$ can be written in the form of (1):

$$\phi(k+1) = P(k)\phi(k), \quad \psi(k) = 1 - \mathbf{1}^T\phi(k),$$

where $\phi(k)$ is a L-dimensional vector recording the probability distribution of $s(k)$ over 1 through L, and ψ is the probability of the mission being accomplished by time t_k. Using the definition of the minimum success probabilities and Lemma 1, one can show that $P(k)$ is more progressive than $\hat{P}$ for all k. The matrix $\hat{P}$ is also sorted. To see this, let r be the maximal eigenvalue of $\hat{P}$. From property (iii) of irreducible matrices in Section II, $r > 1 - p_1$ because $1-p_1$ is a principal submatrix of $\hat{P}$. One may verify that the eigenvector of $\hat{P}$ associated with r is $q = (q_1, q_2, \ldots, q_k)^T$, with

$$\begin{aligned} &q_1 = 1, \\ &q_i = 1 - (1-r)\frac{r^{i-2} + r^{i-3}p_1 + \ldots + p_1p_2\cdots p_{i-2}}{p_1p_2\cdots p_{i-1}}, \\ &i = 2, \ldots, k. \end{aligned}$$

Using $1 < r < 1 - p_1$ and with some tedious calculations, it can be shown that $q_1, q_2, \ldots, q_k$ is indeed a decreasing sequence. So, the process $s(k)$ converges to $L + 1$ in probability at a rate greater than that of r^k. Almost sure convergence to mission accomplishment follows based on the well-known Borel-Cantelli Lemma ([5]). ■

Note that the probabilistic information assumed here is almost minimum for estimating the time-to-mission-accomplishment of an autonomous robot. If richer information is available, then more interesting analysis may be carried out along the same line.

V. A Robust Approach to Markov Decision Problems with Uncertain Transition Probabilities

As a different but very related problem, we consider a discrete-time infinite horizon discounted cost MDP with finite state space S and control space U. For notational simplicity, we assume that U is state-independent. When the system is at state $i \in S$ and control $u \in U$ is taken, a cost $c(i, u)$ is incurred, which we assume nonnegative and bounded. The system then makes a transition to state j with probability $p_{ij}(u)$. The process then repeats from state j. Let $\mathbf{p}_i(u) = (p_{ij}(u))_{j \in S}$ be the transition probability vector associated with the state-control pair (i, u). Consider finding a stationary policy $\pi : S \mapsto U$, and let Π be the set of all admissible candidates. Let $0 < \alpha < 1$ be the discount factor.

In the framework of "standard" MDPs, it is assumed that the transition probability vectors $\mathbf{p}_i(u)$, $\forall\, i \in S, u \in U$, are precisely known. Let us denote by ω the collection of the transition probability vectors, i.e., $\omega \triangleq \{\mathbf{p}_i(u)\}_{i \in S, u \in U}$. Given the initial state $i_0 = i$, the cost of policy π is calculated as

$$V_\omega^\pi(i) = E_\omega\Big[\sum_{t=0}^{\infty} \alpha^t c(i_t, \pi(i_t)) \mid i_0 = i\Big],$$

where i_t is the state at epoch t and the expectation is taken with respect to ω. The objective is then to find an optimal policy π^* that minimizes $V_\omega^\pi(i)$ for all $i \in S$, i.e.,

$$V_\omega^*(i) = \min_{\pi \in \Pi} V_\omega^\pi(i), \quad \forall\, i \in S.$$

For the robust problem, the uncertainty in the transition probabilities can be modeled by assuming that $\mathbf{p}_i(u)$ belongs to some bounded set $\Omega_i(u)$. For instance, $\Omega_i(u)$ can be described as $\Omega_i(u) \triangleq \{\mathbf{p} \mid \underline{\mathbf{p}} \leq \mathbf{p} \leq \overline{\mathbf{p}}, \mathbf{p} \in \Delta_{|S|}\}$, where $\overline{\mathbf{p}} \geq \underline{\mathbf{p}} \geq \mathbf{0}$ and $\Delta_{|S|}$ is the probability simplex in $\mathbb{R}^{|S|}$. Let $\Omega \triangleq \times\Omega_i(u)$, i.e., the Cartesian product of $\Omega_i(u)$, $i \in S$. When the transition probabilities are uncertain, the worst-case approach considered in the literature seeks a policy that minimizes the worst possible cost. Such a policy, denoted by

π_F, is the solution of the standard robust MDP:

$$V_F^*(i) = \min_{\pi\in\Pi}\max_{\omega\in\Omega} E_\omega\Big[\sum_{t=0}^{\infty}\alpha^t c(i_t,\pi(i_t)) \mid i_0 = i\Big], \quad \forall\, i\in S.$$

We refer to π_F as the *fat policy*. It was shown by [9] that $V_F^*(i), \forall\, i\in S$, satisfies the following set of equations, which is the robust version of the well-known Bellman equations:

$$V_F^*(i) = \min_{u\in U}\Big\{c(i,u) + \alpha \max_{\mathbf{p}_i(u)\in\Omega_i(u)}\sum_{j\in S} p_{ij}(u)V_F^*(j)\Big\}, \ \ \forall\, i.$$

The fat policy is found by solving

$$\pi_F(i) = \operatorname*{argmin}_{u\in U}\Big\{c(i,u) + \alpha \max_{\mathbf{p}_i(u)\in\Omega_i(u)}\sum_{j\in S} p_{ij}(u)V_F^*(j)\Big\}, \forall\, i.$$

A value iteration algorithm for solving the robust Bellman equations was proposed by [9]. Also, [6] proved independently the robust Bellman equations and developed both value and policy iteration algorithms.)

Let $V_\omega^{\pi_F}(i)$ be the cost of the fat policy of the initial state $i_0 = i$ when state transitions occur according to a given ω, i.e.,

$$V_\omega^{\pi_F}(i) = E_\omega\Big[\sum_{t=0}^{\infty}\alpha^t c(i_t,\pi_F(i_t)) \mid i_0 = i\Big].$$

Clearly, for any $\omega\in\Omega$, $V_\omega^*(i) \le V_\omega^{\pi_F}(i) \le V_F^*(i)$ for all $i\in S$.

In order to obtain a less conservative policy, let's consider an *optimistic robust formulation*, where we restrict our attention to a subset of the full uncertainty set Ω. Of course, by doing this, we can only produce performance bounds that holds in a probablistic sense. Let $\mathscr{R} \triangleq \times\mathscr{R}_i(u)$, where $\mathscr{R}_i(u)\subseteq\Omega_i(u)$. The optimistic robust MDP problem is:

$$V_R^*(i) = \min_{\pi\in\Pi}\max_{\omega\in\mathscr{R}} E_\omega\Big[\sum_{t=0}^{\infty}\alpha^t c(i_t,\pi(i_t)) \mid i_0 = i\Big], \quad \forall\, i\in S.$$

Let π_R denote an optimal policy of the optimistic robust MDP. Then, $V_R^*(i)$, $\forall\, i\in S$, satisfies the robust Bellman equations:

$$V_R^*(i) = \min_{u\in U}\Big\{c(i,u) + \alpha \max_{\mathbf{p}_i(u)\in\mathscr{R}_i(u)}\sum_{j\in S} p_{ij}(u)V_R^*(j)\Big\}, \ \ \forall\, i,$$

with

$$\pi_R(i) = \operatorname*{argmin}_{u\in U}\Big\{c(i,u) + \alpha \max_{\mathbf{p}_i(u)\in\mathscr{R}_i(u)}\sum_{j\in S} p_{ij}(u)V_R^*(j)\Big\}, \forall\, i.$$

Again, one can use a value iteration algorithm or policy iteration algorithm to determine $V_R^*(i)$ and $\pi_R(i)$ for all $i\in S$.

Having defined the robust MDP, we characterize its performance by comparing its optimal cost with the optimal cost of a random instance of the MDP. Specifically, we are interested in $P\big[V_\omega^{\pi_R}(i) \le V_R^*(i)\big]$ for a randomly selected $\omega\in\Omega$. To that end, let us consider the cost of the robust policy for the MDP with ω. Let $V_\omega^{\pi_R}$ be the corresponding value function, i.e.,

$$V_\omega^{\pi_R}(i) = E_\omega\Big[\sum_{t=0}^{\infty}\alpha^t c(i_t,\pi_R(i_t)) \mid i_0 = i\Big].$$

Consider the probability of the complement of $V_\omega^{\pi_R}(i) > V_R^*(i)$:

$$\begin{aligned} P\big[V_\omega^{\pi_R}(i) > V_R^*(i)\big] &= P\big[V_\omega^{\pi_R}(i) > V_R^*(i) \mid \omega\in\mathscr{R}\big]P\big[\omega\in\mathscr{R}\big] \\ &\quad + P\big[V_\omega^{\pi_R}(i) > V_R^*(i) \mid \omega\notin\mathscr{R}\big]P\big[\omega\notin\mathscr{R}\big] \\ &= P\big[V_\omega^{\pi_R}(i) > V_R^*(i) \mid \omega\notin\mathscr{R}\big]P\big[\omega\notin\mathscr{R}\big]. \end{aligned} \tag{12}$$

Let $\mathbf{p}_i(\pi_R(i)) \in \omega$ be the transition probability vector for the state-control pair $(i,\pi_R(i))$. Since $V_\omega^{\pi_R}(i)$ and $\mathbf{p}_i(\pi_R(i))$ satisfy $V_\omega^{\pi_R}(i) = c(i,\pi_R(i)) + \alpha\sum_{j\in S} p_{ij}(\pi_R(i))V_\omega^{\pi_R}(j)$, we can write the first probability in (12) as

$$\begin{aligned} &P\big[V_\omega^{\pi_R}(i) > V_R^*(i) \mid \omega\notin\mathscr{R}\big] \\ &= P\Big[c(i,\pi_R(i)) + \alpha\sum_{j\in S} p_{ij}(\pi_R(i))V_\omega^{\pi_R}(j) > V_R^*(i) \mid \omega\notin\mathscr{R}\Big] \\ &= P\Big[\sum_{j\in S} p_{ij}(\pi_R(i))V_\omega^{\pi_R}(j) > C(i) \mid \omega\notin\mathscr{R}\Big], \end{aligned} \tag{13}$$

where $C(i) = \frac{1}{\alpha}\big\{V_R^*(i) - c(i,\pi_R(i))\big\}$.

The $V_\omega^{\pi_R}(i)$ in (13) cannot be determined until ω is realized. To find a bound of $V_\omega^{\pi_R}(i)$ that is independent of ω, we compute the worst cost of the policy π_R for all $\omega\notin\mathscr{R}$ as follows:

$$V^{\pi_R}(i) = \max_{\omega\notin\mathscr{R}} E_\omega\Big[\sum_{t=0}^{\infty}\alpha^t c(i_t,\pi_R(i_t)) \mid i_0 = i\Big], \quad \forall\, i\in S.$$

One can calculate $V^{\pi_R}(i)$ through the following set of equations: for all $i\in S$

$$V^{\pi_R}(i) = c(i,\pi_R(i)) + \max_{\omega\notin\mathscr{R}}\sum_{j\in S} p_{ij}(\pi_R(i))V^{\pi_R}(j), \ \forall i\in S. \tag{14}$$

It may not be easy to compute $V^{\pi_R}(i)$ because the requirement of $\omega\notin\mathscr{R}$ could make the maximization problem in (14) complicated. In that case, one may find a bound by replacing $V^{\pi_R}(i)$ with $\widehat{V}^{\pi_R}(i)$, which is the solution of[3]

$$\begin{aligned} \widehat{V}^{\pi_R}(i) &= c(i,\pi_R(i)) \\ &\quad + \alpha \max_{\mathbf{p}_i(\pi_R(i))\in\Omega_i(\pi_R(i))}\sum_{j\in S} p_{ij}(\pi_R(i))\widehat{V}^{\pi_R}(j), \quad \forall\, i\in S. \end{aligned}$$

Replacing $V_\omega^{\pi_R}(i)$ in (13) with $V^{\pi_R}(i)$, we obtain

$$\begin{aligned} &P\Big[\sum\nolimits_{j\in S} p_{ij}(\pi_R(i))V_\omega^{\pi_R}(j) > C(i) \mid \omega\notin\mathscr{R}\Big] \\ &\quad \le P\Big[\sum\nolimits_{j\in S} p_{ij}(\pi_R(i))V^{\pi_R}(j) > C(i) \mid \omega\notin\mathscr{R}\Big] \\ &\quad \le P\big[\mathbf{V}'\mathbf{p}_i(\pi_R(i)) \ge C(i) \mid \omega\notin\mathscr{R}\big], \end{aligned} \tag{15}$$

[3]The use of $\widehat{V}^{\pi_R}(i)$ could make the analysis weaker.

where $\mathbf{V}$ is the vector whose components are the $V^{\pi_R}(i)$.

Putting (12), (13), and (15) together and using Markov's inequality, we obtain for $\theta \geq 0$

$$\begin{aligned} &P\big[V_\omega^{\pi_R}(i) > V_R^*(i)\big] \\ &\quad \leq P\big[\mathbf{V}'\mathbf{p}_i(\pi_R(i)) \geq C(i) \mid \omega \notin \mathscr{R}\big] P\big[\omega \notin \mathscr{R}\big] \\ &\quad \leq e^{-\theta C(i)} E\big[e^{\theta \mathbf{V}'\mathbf{p}_i(\pi_R(i))} \mid \omega \notin \mathscr{R}\big] P\big[\omega \notin \mathscr{R}\big] \\ &\quad = \exp\big[-\theta C(i) + \Lambda_{\mathbf{p}_i(\pi_R(i)) \in \omega \notin \mathscr{R}}(\theta \mathbf{V})\big] P\big[\omega \notin \mathscr{R}\big], \end{aligned}$$

where $\Lambda_{\mathbf{p}_i(\pi_R(i)) \in \omega \notin \mathscr{R}}(\theta \mathbf{V}) \triangleq \log E\big[e^{\theta \mathbf{V}'\mathbf{p}_i(\pi_R(i))} \mid \omega \notin \mathscr{R}\big]$. Optimizing over θ, we arrive at the following proposition.

Proposition 1: It holds that

$$\begin{aligned} &P\big[V_\omega^{\pi_R}(i) > V_R^*(i)\big] \\ &\leq \exp\Big[\inf_{\theta \geq 0} \Big\{-\theta C(i) + \Lambda_{\mathbf{p}_i(\pi_R(i)) \in \omega \notin \mathscr{R}}(\theta \mathbf{V})\Big\}\Big] P\big[\omega \notin \mathscr{R}\big]. \end{aligned} \tag{16}$$

In general, computing the probability bound in (16) exactly would still pose computational challenges. Sometimes, however, ω is induced by a few parameters. In this case, the computational challenge could be mild.

VI. Conclusion

The theme of this paper is on the analysis and control of stochastic processes whose transition probabilities are uncertain. We first formulated a basis of comparison between finite-state stochastic processes with one absorbing state, and showed that the result has useful implications to robust optimal control. Second, we discussed our preliminary work on optimistic robust solution of Markov Decision Processes, in which the key issue is to estimate the probability that the performance in an actual operation would fall short of the performance bound suggested by the optimistic robust formulation.

VII. Acknowledgement

We would like to thank Professor Raffaello D'Andrea at Cornell University for supporting a part of this work.

VIII. References

[1] S Andersson and D Hristu-Varsakelis. Stochastic language-based motion control. In *Proceedings of the 42nd IEEE Conference on Decision and Control*, pages 3313–18, 2003.

[2] R W Brockett. On the computer control of movement. In *Proceedings of the 1988 IEEE Conference on Robotics and Automation*, pages 534–540, 1988.

[3] R W Brockett. Formal languages for motion description and map making. In *Robotics*, pages 181–93. American Mathematical Society, 1990.

[4] M Egerstedt, E Frazzoli, and G Pappas, editors. *Symbolic Methods for Complex Control Systems.*, volume 51 of *Special Issue*. IEEE Transactions on Automatic Control, 2006.

[5] W Feller. *An Introduction to Probability Theory and Its Application*. John Wiley & Sons, 1968.

[6] G. Iyengar. Robust dynamic programming. *Mathematics of Operations Research*, 30(2):257–280, 2005.

[7] V Manikonda, J Hendler, and P S Krishnaprasad. Languages, behaviors, hybrid architectures and motion control. In J B Baillieul and J C Willems, editors, *Mathematical Control Theory*, pages 199–226, 1998.

[8] H Minc. *Nonnegative Matrices*. Wiley-Interscience, 1988.

[9] A. Nilim and L. El Ghaoui. Robust control of Markov decision processes with uncertain transition matrices. *Operations Research*, 53(5):780–798, 2005.

[10] I. Ch. Paschalidis and S.-C. Kang. Robust linear optimization: On the benefits of distributional information and applications in inventory control. In *Proceedings of the 44th IEEE Conference on Decision and Control*, pages 4416–4421, Seville, Spain, 2005.

[11] I. Ch. Paschalidis and S.-C. Kang. On the benefits of distributional information in robust linear optimization. In *Proceedings of the 5th IFAC Symposium on Robust Control Design*, Toulouse, France, 2006.

[12] J. K. Satia and R. E. Lave. Markovian decision processes with uncertain transition probabilities. *Operations Research*, 21(3):728–740, 1973.

[13] C. C. White and H. K. Eldeib. Markov decision processes with imprecise transition probabilities. *Operations Research*, 42(4):739–749, 1994.