2018 European Control Conference (ECC)
June 12-15, 2018. Limassol, Cyprus

Performance of Model Predictive Control of POMDPs

Martin A. Sehr & Robert R. Bitmead

***Abstract*— We revisit closed-loop performance guarantees for Model Predictive Control in the deterministic and stochastic cases, which extend to novel performance results applicable to receding horizon control of Partially Observable Markov Decision Processes. The general intractability of stochastic optimal control relaxes for this particular instance of stochastic systems, provided reasonable problem dimensions are taken. This motivates extending available performance guarantees to this particular class of systems, which may also be used to approximate general nonlinear dynamics via gridding of state, observation, and control spaces. We demonstrate applicability of the novel closed-loop performance results on a particular example in healthcare decision making, which relies explicitly on the duality of the control decisions.**

I. Introduction

Model Predictive Control (MPC) is well applied and popular because of its capacity to handle constraints and its simple formulation as an open-loop finite-horizon optimization problem evaluated on the receding horizon [1], [2]. There are a few areas in which MPC is found wanting for more complete results, notably in the area of output feedback control and the associated requirement to manage the duality of the control signal in stochastic MPC (SMPC) problems, where exploration and exploitation have to be balanced by the ensuing control policies. When SMPC is developed as a logical extension of finite-horizon Stochastic Optimal Control, which demands computation of closed-loop policies, it inherits the computational intractability of this latter subject via the inclusion of the Bayesian filter, required to propagate the conditional state densities, and the stochastic dynamic programming equation.

Results exist relating the infinite-horizon performance of MPC to both the optimal performance and the performance computed as part of the finite-horizon optimization. These performance bounds are available in both the deterministic [3] and the stochastic [4] settings, were one ever able to solve the underlying finite-horizon stochastic problem computationally. While approximation of SMPC based on Stochastic Optimal Control via more tractable surrogate problems is possible, such as for instance in [5]–[8], one generally loses the associated closed-loop guarantees, in particular regarding infinite-horizon performance of the generated control laws.

In this paper, we derive new performance results for SMPC of systems described by Partially Observable Markov Decision Processes (POMDPs, see e.g. [9], [10]). POMDP system models of small to moderate dimensions admit tractable computation of finite-horizon stochastic optimal control laws while preserving the control signal duality, and so are attractive propositions with which to approach implementable SMPC [11], [12]. Here, *duality* refers to the joint but antagonistic roles of the feedback control signal in regulating the state and in probing the state to ensure observability [13], [14]. In deriving perfomance bounds for this specific class of problems, we examine their relation to the deterministic and stochastic continuous-state results, highlighting the role of value function monotonicity with horizon. All theorems discussed in this paper exhibit the same conceptual structure:

Assumption:
Terminal cost contraction

$\Downarrow$

Observation:
Value function monotonicity with horizon

$\Downarrow$

Result:
Infinite-horizon optimal cost
$\wedge|$
Achieved infinite-horizon MPC cost
$\wedge|$
Computed N-horizon optimal cost

While the capability of handling constraints is a raison-d'être for MPC, constraints complicate this analysis and add little to the discussion about closed-loop cost. Thus, as in most of [3], we omit their explicit consideration in this paper and point out that constraints may be reinserted subject to recursive feasibility assumptions.

The paper is organized as follows. We revisit a particular infinite-horizon performance result from [3] in Section II. We then proceed by reviewing a stochastic counterpart to this result, derived in [4], which we extend to receding horizon control of POMDPs in Section IV. A specific POMDP example from healthcare is studied in Section V to demonstrate numerically the satisfaction of assumptions, interpret control duality, and evaluate performance bounds on the infinite control horizon. The example, introduced in [12], displays in particular the dual nature of SMPC based on Stochastic Optimal Control. Our central aim in this paper is to demonstrate that the seemingly opaque and computationally intractable stochastic MPC formulation with attached performance guarantees is indeed feasibly evaluated for modestly sized: problem dimensions, MPC horizon, and gridding of the state and control spaces.

M.A. Sehr is with Siemens Corporate Technology, Berkeley, CA 94704, USA. R.R Bitmead is with the Department of Mechanical and Aerospace Engineering, UC San Diego, La Jolla, CA 92093, USA.
martin.sehr@siemens.com | rbitmead@ucsd.edu

II. Deterministic Model Predictive Control

This section revisits a performance result for deterministic MPC from [3], which we extend to SMPC for nonlinear systems (see also [4]) and POMDPs below. Consider the nonlinear dynamic system

$$x_{t+1} = f(x_t, u_t),$$

where $x_t \in X$ and $u_t \in U$ for $t \in \mathbb{N}_0 \triangleq \{0, 1, 2, \ldots\}$ and metric spaces X, U. Further define the space of control sequences $u : \mathbb{N}_0 \to U$ as $\mathcal{U}$. In principle, we aim to find a control policy $\mu : X \to U$ that minimizes the infinite-horizon cost functional with stage cost $c : X \times U \to \mathbb{R}_+$,

$$J_\infty(x_0, u) \triangleq \sum_{k=0}^{\infty} c(x_k, u_k). \tag{1}$$

Define the optimal value function for cost (1) as

$$J^\star_\infty(x_0) \triangleq \inf_{u \in \mathcal{U}} J_\infty(x_0, u)$$

Given that solution of this infinite-horizon optimal control problem, even in the deterministic case, is usually intractable, a popular approach is to replace (1) by a finite-horizon optimal control problem over horizon $N \in \mathbb{N}_0$, with cost functional

$$J_N(x_0, u) \triangleq \sum_{k=0}^{N-1} c(x_k, u_k) + c_N(x_N), \tag{2}$$

where $c_N : X \to \mathbb{R}_+$ denotes an optional terminal cost term. The optimal value function for (2) is defined by

$$J^\star_N(x_0) \triangleq \inf_{u \in \mathcal{U}} J_N(x_0, u). \tag{3}$$

Denote the sequence of optimal control policies in this finite-horizon problem by μ^N, with first control policy $\mu^N_0 : X \to U$, which is implemented repeatedly in MPC, by

$$\mu^N_{\text{MPC}} \triangleq \{\mu^N_0, \mu^N_0, \ldots\}.$$

We now provide computational estimates of the infinite-horizon achieved MPC cost $J_\infty(x_0, \mu^N_{\text{MPC}})$ in relation to the computed finite-horizon optimal cost $J_N(x, \mu^N)$. This goal can be achieved, for instance, by using the following assumption.

Assumption 1. For all $x \in X$, there exists $u \in U$ such that

$$c_N(f(x, u)) \leq c_N(x) - c(x, u).$$

This assumption on the terminal cost c_N in (2) then leads to the following performance guarantee.

Theorem 1 (Performance of deterministic MPC [3])**.** *Given Assumption 1, the inequality*

$$J^\star_\infty(x) \leq J_\infty(x, \mu^N_{MPC}) \leq J^\star_N(x)$$

holds for all $x \in X$.

This result, which is a special case of Theorem 6.2 in [3], provides bounds on the achieved infinite-horizon performance of the closed-loop system when choosing the terminal cost, c_N, as a Lyapunov function. This result is particularly useful because we compute the upper bound implicitly when generating our MPC control law, μ^N_{MPC}. Theorem 1 follows given that Assumption 1 implies that the underlying finite-horizon optimal value function $J^\star_N(x)$ is monotonically non-increasing with control horizon N. Notice this result provides infinite- and finite-horizon closed-loop performance guarantees. This follows simply by

$$J^M_\infty(x, \mu^N_{\text{MPC}}) \leq J_\infty(x, \mu^N_{\text{MPC}}),$$

for all $M \in \mathbb{N}_0$ and $x \in X$, where

$$J^M_\infty(x_0, u) \triangleq \sum_{k=0}^{M} c(x_k, u_k).$$

The stochastic extension of this observation is of interest in particular for applications such as the healthcare example provided in Section V below, where infinite-horizon performance may not be of particular interest given the inherent finite-horizon nature of the control problem. We next provide results of similar quality to Theorem 1 for SMPC and in particular SMPC applied to POMDPs in the following sections. We note that [3] also consider monotonic non-decreasing cost inequalities, but these fail to operate as Lyapunov functions and therefore provide bounds only asymptotically in N.

III. Stochastic Model Predictive Control

We next discuss closed-loop performance of SMPC as in [4]. Committing a slight abuse of notation, we shall recycle most of the symbols used previously in Section II above. Consider nonlinear stochastic systems of the form

$$x_{t+1} = f(x_t, u_t, w_t), \tag{4}$$
$$y_t = h(x_t, v_t), \tag{5}$$

where $x_t \in X$, $u_t \in U$, $y_t \in Y$ for $t \in \mathbb{N}_0$ and metric spaces X, U, Y, respectively. Starting from known initial state density $\pi_{0|-1} = \text{pdf}(x_0)$, we denote the data available at time t by

$$\zeta^t \triangleq \{y_0, u_0, y_1, u_1, \ldots, u_{t-1}, y_t\}, \qquad \zeta^0 \triangleq \{y_0\},$$

and impose the following standing assumption on the signals.

Assumption 2. The signals in (4-5) satisfy:

1. w_t and v_t are i.i.d. sequences with known densities.
2. x_0, w_t, v_l are mutually independent for all $t, l \in \mathbb{N}_0$.
3. The control input u_t at time instant $t \in \mathbb{N}_0$ is a function of the data ζ^t and given initial state density $\pi_{0|-1}$.

The *information state,* denoted π_t, is the conditional probability density function of state x_t given data ζ^t,

$$\pi_t \triangleq \text{pdf}\left(x_t \mid \zeta^t\right).$$

As a result of the Markovian dynamics (4-5), optimal control inputs must inherently be *separated* feedback policies (e.g. [13], [14]). That is, optimal control input u_t depends on the data ζ^t and initial density $\pi_{0|-1}$ solely through the current information state, π_t. Optimality thus requires propagating π_t and policies g_t, where

$$u_t = g_t(\pi_t).$$

Definition 1. $\mathbb{E}_t[\,\cdot\,]$ and $\mathbb{P}_t[\,\cdot\,]$ are expected value and probability with respect to state x_t – with conditional density π_t – and i.i.d. random variables $\{(w_k, v_{k+1}) : k \geq t\}$.

Notice that stochastic optimal control on the infinite horizon (see [13], [15]) typically requires a discount factor $\alpha < 1$, casting the stochastic version of (1) as

$$J_\infty(\pi_0, g) \triangleq \mathbb{E}_0\left[\sum_{k=0}^{\infty} \alpha^k c(x_k, g_k(\pi_k))\right], \tag{6}$$

with corresponding finite-horizon cost

$$J_N(\pi_0, g) \triangleq \mathbb{E}_0\left[\sum_{k=0}^{N-1} \alpha^k c(x_k, g_k(\pi_k)) + \alpha^N c_N(x_N)\right]. \tag{7}$$

Defining the optimal value function $J_N^\star(\pi_0)$ as in (3),

$$J_N^\star(\pi_0) \triangleq \inf_{g_k(\cdot)} J_N(\pi_0, g),$$

finite-horizon stochastic optimal feedback policies may be computed, in principle, by solving the stochastic dynamic programming equation,

$$J_{N-k}^\star(\pi_k) \triangleq \inf_{g_k(\cdot)} \mathbb{E}_k\left[c(x_k, g_k(\pi_k)) + \alpha J_{N-k-1}^\star(\pi_{k+1})\right], \tag{8}$$

for $k = 0, \ldots, N-1$. The equation is solved backwards in time, from its terminal value,

$$J_0^\star(\pi_N) \triangleq \mathbb{E}_N\left[c_N(x_N)\right]. \tag{9}$$

Similarly to Section II, we denote by: $J_\infty^\star(\pi)$ the infinite-horizon optimal value function; μ^N the sequence of optimal policies in (8-9); μ_0^N the first element of this sequence; $\mu_{\text{MPC}}^N \triangleq \{\mu_0^N, \mu_0^N, \ldots\}$ the receding horizon implementation of this sequence. We next impose the following stochastic counterpart to Assumption 1 to discuss the infinite horizon cost of the SMPC law μ_{MPC}^N.

Assumption 3. For $\alpha \in [0, 1)$, there exist $\eta \in \mathbb{R}_+$ and a policy $\tilde{g}(\cdot)$ such that

$$\mathbb{E}_\pi\left[\alpha\, c_N(f(x, \tilde{g}(\pi), w))\right] \leq \mathbb{E}_\pi\left[c_N(x) - c(x, \tilde{g}(\pi))\right] + \frac{\eta}{\alpha^{N-1}},$$

for all densities π of $x \in X$. The expectation $\mathbb{E}_\pi[\cdot]$ is with respect to state x – with conditional density π – and w.

This assumption then leads to the following extension of Theorem 1 to SMPC of system (4-5).

Theorem 2 (Performance of stochastic MPC [4])**.** *Given Assumption 3, SMPC with $\alpha \in [0, 1)$ yields*

$$J_\infty^\star(\pi) \leq J_\infty(\pi, \mu_{MPC}^N) \leq J_N^\star(\pi) + \frac{\alpha}{1-\alpha}\eta,$$

for all densities π of $x \in X$.

This result relates the following quantities in SMPC: *design cost*, $J_N^\star(\pi)$, which is evaluated as part of the SMPC computation; *optimal cost*, $J_\infty^\star(\pi)$, which is unknown (otherwise we would use the infinite-horizon optimal policy); unknown infinite-horizon SMPC *achieved cost* $J_\infty(\pi, \mu_{\text{MPC}}^N)$. The result, which must exhibit duality by satisfaction of the stochastic dynamic programming equation (8-9), is special; SMPC approaches relying on approximation of the finite horizon Stochastic Optimal Control problem, as commonly found in the literature, do not generally yield statements regarding performance of the implemented control laws on the infinite horizon. This fact is linked inherently to the loss of duality when avoiding solution of (8-9).

As in Section II and Theorem 1, the proof of Theorem 2 via Assumption 3 relies on verifying monotonicity of the underlying optimal value function $J_N^\star(\pi)$. We next proceed by extending this result and its proof to dual optimal receding horizon control of POMDPs.

IV. Stochastic MPC for POMDPs

POMDPs are typically characterized by probabilistic dynamics on a finite state space $X = \{1, \ldots, n\}$, finite action space $U = \{1, \ldots, m\}$, and finite observation space $Y = \{1, \ldots, o\}$. POMDP dynamics are defined by the conditional state transition and observation probabilities

$$\mathbb{P}\left(x_{t+1} = j \mid x_t = i, u_t = a\right) = p_{ij}^a, \tag{10}$$

$$\mathbb{P}\left(y_{t+1} = \theta \mid x_{t+1} = j, u_t = a\right) = r_{j\theta}^a, \tag{11}$$

where $t \in \mathbb{N}_0$, $i, j \in X$, $a \in U$, $\theta \in Y$. The state transition dynamics (10) correspond to a conventional Markov Decision Process (MDP, e.g. [16]). However, the control actions u_t are to be chosen based on the known initial state distribution $\pi_0 = \text{pdf}(x_0)$ and the sequences of observations, $\{y_1, \ldots, y_t\}$, and controls $\{u_0, \ldots, u_{t-1}\}$, respectively. That is, we are choosing our control actions in a Hidden Markov Model (HMM, e.g. [17]) setup.

Given control action $u_t = a$ and measured output $y_{t+1} = \theta$, the information state π_t in a POMDP is updated via

$$\pi_{t+1,j} = \frac{\sum_{i \in X} \pi_{t,j} p_{ij}^a r_{j\theta}^a}{\sum_{i,k \in X} \pi_{t,j} p_{ik}^a r_{k\theta}^a},$$

where $\pi_{t,j}$ denotes the j^{th} entry of the row vector π_t. To specify the cost functionals (6) and (7) in the POMDP setup, we write the stage cost as $c(x_t, u_t) = c_i^a$ if $x_t = i \in X$ and $u_t = a \in U$, summarized in the column vectors $c(a)$ of the same dimension as row vectors π_k. Similarly, the terminal cost terms are $c_N(x_t) = c_{i,N}$ if $x_N = i \in X$, summarized in the column vector c_N. The infinite horizon cost functional defined in Section III then follows as

$$J_\infty(\pi_0, g) = \mathbb{E}_0\left[\sum_{k=0}^{\infty} \alpha^k \pi_k c(g_k(\pi_k))\right],$$

with corresponding finite-horizon variant

$$J_N(\pi_0, g) = \mathbb{E}_0 \left[\sum_{k=0}^{N-1} \alpha^k \pi_k c(g_k(\pi_k)) + \alpha^N \pi_N c_N \right].$$

Extending (8-9), optimal control decisions are computed via

$$J^\star_{N-k}(\pi_k) = \min_{g_k(\cdot)} \Bigg\{ \pi_k c(g_k(\pi_k)) + \alpha \sum_{\theta \in Y} \mathbb{P}\left(y_{k+1} = \theta \mid \pi_k,\, g_k(\pi_k)\right) J^\star_{N-k-1}(\pi_{k+1}) \Bigg\}, \tag{12}$$

for $k = 0, \dots, N-1$, from terminal value function

$$J^\star_0(\pi_N) = \pi_N c_N. \tag{13}$$

Using the notation for optimal finite- and infinite-horizon value functions as well as MPC policies introduced in Section III, we next prove the following auxiliary result before extending the performance guarantees in Theorem 2 to SMPC on POMDPs.

Lemma 1. *If there exist $\gamma \in [0,1]$ and $\eta \in \mathbb{R}_+$ such that*

$$\mathbb{E}_0 \left[J^\star_N(\pi_1) - J^\star_{N-1}(\pi_1) \right] \;\le\; \gamma \mathbb{E}_0 \left[\pi_0 c(\mu^N_0(\pi_0)) \right] + \eta, \tag{14}$$

for all densities π_0 of $x_0 \in X$, then SMPC with discount factor $\alpha \in [0,1)$ yields

$$\begin{aligned}(1-\alpha\gamma)\, J^\star_\infty(\pi_0) &\le (1-\alpha\gamma)\, J_\infty(\pi_0, \mu^N_{MPC}) \\ &\le J^\star_N(\pi_0) + \frac{\alpha}{1-\alpha}\eta. \end{aligned} \tag{15}$$

Proof. Optimality of the initial policy $\mu^N_0(\cdot)$ implies

$$\begin{aligned} J^\star_N(\pi_0) = \mathbb{E}_0 &\left[\pi_0 c(\mu^N_0(\pi_0)) + \alpha J^\star_{N-1}(\pi_1) \right] \\ &+ \alpha \mathbb{E}_0 \left[J^\star_N(\pi_1) - J^\star_N(\pi_1) \right], \end{aligned}$$

which by (14) yields

$$\begin{aligned} (1-\alpha\gamma)\mathbb{E}_0 &\left[\pi_0 c(\mu^N_0(\pi_0)) \right] \le \\ & J^\star_N(\pi_0) - \alpha \mathbb{E}_0 \left[J^\star_N(\pi_1) \right] + \alpha\eta. \end{aligned} \tag{16}$$

Now denote by $J^M_\infty(\pi_0, \mu^N_{\text{MPC}})$ the first $M \in \mathbb{N}_1$ terms of the achieved infinite-horizon cost $J_\infty(\pi_0, \mu^N_{\text{MPC}})$ subject to the SMPC implementation of policy $\mu^N_0(\cdot)$. By (16), we have

$$\begin{aligned} (1-\alpha\gamma) & J^M_\infty(\pi_0, \mu^N_{\text{MPC}}) = \\ & (1-\alpha\gamma)\mathbb{E}_0 \left[\sum_{k=0}^{M-1} \alpha^k \pi_k c(\mu^N_0(\pi_k)) \right] \le \\ \mathbb{E}_0 & \left[J^\star_N(\pi_0) - \alpha J^\star_N(\pi_1) + \alpha\eta + \alpha J^\star_N(\pi_1) - \alpha^2 J^\star_N(\pi_2) + \right. \\ \alpha^2 & \left. \eta + \ldots + \alpha^{M-1} J^\star_N(\pi_{M-1}) - \alpha^M J^\star_N(\pi_M) + \alpha^M \eta \right], \end{aligned}$$

such that

$$\begin{aligned} (1-\alpha\gamma) J^M_\infty(\pi_0, \mu^N_{\text{MPC}}) &\le J^\star_N(\pi_0) \\ &- \alpha^M \mathbb{E}_0 \left[J^\star_N(\pi_M) \right] + \left(\alpha + \ldots + \alpha^M \right) \eta, \end{aligned}$$

which confirms the right inequality in (15) as $M \to \infty$. The left inequality follows from infinite-horizon optimality. ☐

This lemma then leads to the following assumption and subsequent performance result in the spirit of Theorems 1-2.

Assumption 4. For $\alpha \in [0,1)$, there exist $\eta \in \mathbb{R}_+$ and a policy $\tilde{g}(\cdot)$ such that

$$\mathbb{E}_0 \left[\alpha\, \pi_1 c_N \right] \le \mathbb{E}_0 \left[\pi_0 c_N - \pi_0 c(\tilde{g}(\pi_0)) \right] + \frac{\eta}{\alpha^{N-1}}, \tag{17}$$

for all densities π_0 of $x_0 \in X$.

Theorem 3. *[Performance of SMPC for POMDPs] Given Assumption 4, SMPC for POMDPs with $\alpha \in [0,1)$ yields*

$$J^\star_\infty(\pi) \le J_\infty(\pi, \mu^N_{MPC}) \le J^\star_N(\pi) + \frac{\alpha}{1-\alpha}\eta,$$

for all densities π of $x \in X$.

Proof. Use optimality and Assumption 4 to conclude

$$\begin{aligned} & J^\star_N(\pi_1) - J^\star_{N-1}(\pi_1) \\ &= \mathbb{E}_0 \Bigg[\left(\sum_{k=0}^{N-1} \alpha^k \pi_{k+1} c(\mu^N_k(\pi_{k+1})) + \alpha^N \pi_{N+1} c_N \right) \\ &\quad - \left(\sum_{k=0}^{N-2} \alpha^k \pi_{k+1} c(\mu^N_{k+1}(\pi_{k+1})) + \alpha^{N-1} \pi_N c_N \right) \Bigg] \\ &\le \mathbb{E}_0 [\alpha^{N-1} \pi_N c(\tilde{g}(\pi_N)) \\ &\quad + \alpha^N \pi_{N+1} c_N - \alpha^{N-1} \pi_N c_N] \\ &\le \eta, \end{aligned}$$

which implies (14) with $\gamma = 0$ and thus completes the proof by Lemma 1. ☐

V. Numerical Example in Healthcare

A. Problem Setup

The remainder of this paper discusses a particular numerical example of decisions on treatment and diagnosis in healthcare, displaying specifically the use of dual control in SMPC applied to a POMDP. Consider a patient treated for a specific disease which can be managed but not cured. For simplicity, we assume that the patient does not decease under treatment.

The example, introduced in [12] and based on a Hepatitis B model of [18], is set up as follows. The disease encompasses three stages with severity increasing from Stage 1 through Stage 2 to Stage 3, transitions between which are governed by a Markov chain with transition probability matrix

$$P = \begin{bmatrix} 0.8 & 0.2 & 0.0 \\ 0.0 & 0.9 & 0.1 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}.$$

Once our patient enters Stage 3, Stages 1 and 2 are inaccessible for all future times. However, Stage 3 can only be entered through Stage 2, a transition from which to Stage 1 is possible only under costly treatment. The same treatment inhibits transitions from Stage 2 to Stage 3. We have access

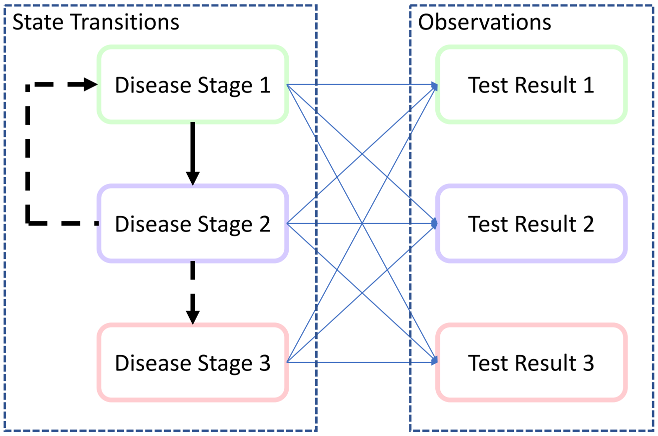

Fig. 1. Feasible state transitions and possible test results in healthcare example. Solid arrows for feasible state transitions and observations. Dashed arrows for transitions conditional on treatment and diagnosis decisions.

to the patient state only through tests, which will result in one of three possible values, each of which is representative of one of the three disease stages. However, these tests are imperfect, with non-zero probability of returning an incorrect disease stage. All possible state transitions and observations are illustrated in Figure 1.

At each point in time, the current information state π_t, but not the exact patient state x_t, is available to make one of four possible decisions:

1) Skip next appointment slot
2) Schedule new appointment
3) Order rapid diagnostic test
4) Apply available treatment

Skipping an appointment slot results in the patient progressing through the Markov chain without medical intervention, and without new information being available after the current decision epoch. Scheduling an appointment does not alter the patient transition probabilities but provides a low-quality assessment of the current disease stage, which is used to refine the next information state. The third option, ordering a rapid diagnostic test, allows for a high-quality assessment of the patient's state, leading to a more reliable refinement of the next information state than possible when choosing the previous decision option. The results from this diagnostic test are considered available sufficiently fast so that the patient state remains unchanged under this decision. The remaining option entails medical intervention, allowing transition from Stage 2 to Stage 1 while preventing transition from Stage 2 to Stage 3. Transition probabilities $P(a)$, observation probabilities $R(a)$, and stage cost vectors $c(a)$ for each decision are summarized in Table I. Additionally, we impose the terminal cost

$$c_N = \begin{bmatrix} 0 & 4 & 30 \end{bmatrix}^T.$$

B. Computational Results

The trade-off between these two principal decision categories is precisely what is encompassed by duality, which we can include in an optimal sense by solving (12-13) and applying the resulting initial policy in receding horizon fashion. This is demonstrated in Figure 2, which shows simulation results for SMPC with control horizon $N = 5$ and discount factor $\alpha = 0.98$. As anticipated, the stochastic optimal receding horizon policy shows a structure not drastically different from the decision structure motivated above. In particular, diagnostic tests are used effectively to decide on medical intervention.

In order to apply Theorem 3 to this particular example, we choose the policy $\tilde{g}(\cdot)$ in Assumption 4 always to apply medical intervention. Using the worst-case scenario for the expectations in (17), which entails transition from Stage 1 to Stage 2 under treatment, we can satisfy Assumption 4 with $\eta = 7.92$. The computed cost in our simulation is $J_N^\star(\pi_0) \approx 11.36$. Combined with the discount factor $\alpha = 0.98$, we thus have the upper bound

$$J_\infty(\pi_0, \mu_{\text{MPC}}^N) \leq J_N^\star(\pi_0) + \frac{\alpha}{1-\alpha}\eta \approx 400$$

via application of Theorem 3. Denoting by e_j the row-vector with entry 1 in element j and zeros elsewhere, the observed (finite-horizon) cost corresponding with Figure 2 is

$$J_\infty^{\text{obs}} = \sum_{k=0}^{29} e_{x_k} c(\mu_0^N(\pi_k)) \approx 38.53 < 400.$$

While this bound is not particularly tight, one may modify the discount factor α or the terminal cost c_N, such as by iterating to match the terminal cost with the infinite-horizon cost, to achieve a tighter estimate of the achieved MPC cost.

VI. Conclusions

We extended closed-loop achieved performance guarantees well-known in deterministic MPC to SMPC and in particular receding horizon control of POMDPs, which allow tractable solution of the underlying Stochastic Optimal Control problems and thus duality of the control inputs in an optimal sense. The basic formulations in this paper can be modified, for instance, by introducing state and input constraint sets and time-varying stage costs. We successfully demonstrated use of the novel results using a particular POMDP instance in healthcare decision making, demanding the use of probing control inputs in order to decide adequately upon the proper and cost-effective use of medical intervention.

References

[1] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. M. Scokaert, "Constrained model predictive control: Stability and optimality," *Automatica*, vol. 36, no. 6, pp. 789–814, 2000.

[2] D. Q. Mayne, "Model predictive control: Recent developments and future promise," *Automatica*, vol. 50, no. 12, pp. 2967–2986, 2014.

[3] L. Grüne and A. Rantzer, "On the infinite horizon performance of receding horizon controllers," *IEEE Transactions on Automatic Control*, vol. 53, no. 9, pp. 2100–2111, 2008.

[4] M. A. Sehr and R. R. Bitmead, "Stochastic model predictive control: Output-feedback, duality and guaranteed performance," *Automatica*, to appear.

[5] D. Sui, L. Feng, and M. Hovd, "Robust output feedback model predictive control for linear systems via moving horizon estimation," in *American Control Conference*, (Seattle, WA), pp. 453–458, 2008.

TABLE I

PROBLEM DATA FOR HEALTHCARE DECISION MAKING EXAMPLE.

Decision a	Transition Probabilities $P(a)$	Observation Probabilities $R(a)$	Cost $c(a)$
1: Skip next appointment slot	$\begin{bmatrix} 0.80 & 0.20 & 0.00 \\ 0.00 & 0.90 & 0.10 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}$	$\begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 5 \\ 5 \end{bmatrix}$
2: Schedule new appointment	$\begin{bmatrix} 0.80 & 0.20 & 0.00 \\ 0.00 & 0.90 & 0.10 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}$	$\begin{bmatrix} 0.40 & 0.30 & 0.30 \\ 0.30 & 0.40 & 0.30 \\ 0.30 & 0.30 & 0.40 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$
3: Order rapid diagnostic test	$\begin{bmatrix} 1.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}$	$\begin{bmatrix} 0.90 & 0.05 & 0.05 \\ 0.05 & 0.90 & 0.05 \\ 0.05 & 0.05 & 0.90 \end{bmatrix}$	$\begin{bmatrix} 4 \\ 3 \\ 4 \end{bmatrix}$
4: Apply available treatment	$\begin{bmatrix} 0.80 & 0.20 & 0.00 \\ 0.75 & 0.25 & 0.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}$	$\begin{bmatrix} 0.40 & 0.30 & 0.30 \\ 0.30 & 0.40 & 0.30 \\ 0.30 & 0.30 & 0.40 \end{bmatrix}$	$\begin{bmatrix} 4 \\ 2 \\ 4 \end{bmatrix}$

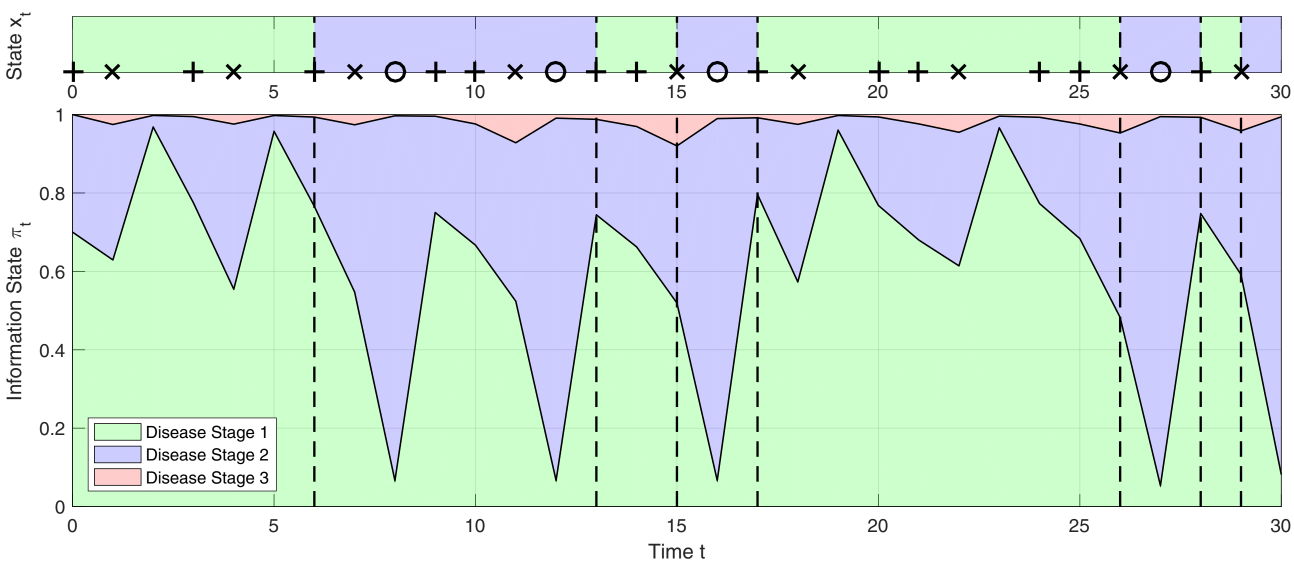

Fig. 2. Simulation results for SMPC with horizon $N = 5$ and discount factor $\alpha = 0.98$. Top plot: patient state with SMPC decisions based on current information state: appointment (pluses); diagnosis (crosses); treatment (circles). Bottom plot: information state evolution and instances of state transitions.

[6] D. Q. Mayne, S. V. Raković, R. Findeisen, and F. Allgöwer, "Robust output feedback model predictive control of constrained linear systems: Time varying case," *Automatica*, vol. 45, pp. 2082–2087, 2009.

[7] L. Blackmore, M. Ono, A. Bektassov, and B. C. Williams, "A probabilistic particle-control approximation of chance-constrained stochastic predictive control," *IEEE Transactions on Robotics*, vol. 26, no. 3, pp. 502–517, 2010.

[8] M. A. Sehr and R. R. Bitmead, "Particle model predictive control: Tractable stochastic nonlinear output-feedback MPC," in *IFAC World Congress*, (Toulouse, France), pp. 15361–15366, 2017.

[9] R. D. Smallwood and E. J. Sondik, "The optimal control of partially observable Markov processes over a finite horizon," *Operations Research*, vol. 21, no. 5, pp. 1071–1088, 1973.

[10] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial intelligence*, vol. 101, no. 1, pp. 99–134, 1998.

[11] Z. Sunberg, S. Chakravorty, and R. S. Erwin, "Information space receding horizon control," *IEEE Transactions on Cybernetics*, vol. 43, no. 6, pp. 2255–2260, 2013.

[12] M. A. Sehr and R. R. Bitmead, "Tractable dual optimal stochastic model predictive control: An example in healthcare," in *Conf. Control Technology and Applications*, (Kohala Coast, HI), IEEE, 2017.

[13] D. P. Bertsekas, *Dynamic programming and optimal control*. Belmont, MA: Athena Scientific, 1995.

[14] P. R. Kumar and P. Varaiya, *Stochastic Systems: Estimation, Identification, and Adaptive Control*. Englewood Cliffs, NJ: Prentice-Hall, 1986.

[15] D. P. Bertsekas and S. E. Shreve, *Stochastic optimal control: The discrete time case*, vol. 23. New York, NY: Academic Press, 1978.

[16] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[17] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov models: estimation and control*, vol. 29. Springer Science & Business Media, 2008.

[18] M. A. Sehr, K. D. Joshi, J. M. Fontanesi, R. J. Wong, R. R. Bitmead, and R. G. Gish, "Markov modeling in hepatitis B screening and linkage to care," *Theoretical Biology and Medical Modelling*, vol. 14, no. 1, p. 11, 2017.