Welfare Maximization Algorithm for Solving Budget-Constrained Multi-Component POMDPs

Manav Vora, Pranay Thangeda, Michael N. Grussing, and Melkior Ornik, *Senior Member, IEEE*

Abstract—**Partially Observable Markov Decision Processes (POMDPs) provide an efficient way to model real-world sequential decision making processes. Motivated by the problem of maintenance and inspection of a group of infrastructure components with independent dynamics, this letter presents an algorithm to find the optimal policy for a multi-component budget-constrained POMDP. We first introduce a budgeted-POMDP model (*b-POMDP*) which enables us to find the optimal policy for a POMDP while adhering to budget constraints. Next, we prove that the value function or maximal collected reward for a special class of b-POMDPs is a concave function of the budget for the finite horizon case. Our second contribution is an algorithm to calculate the optimal policy for a multi-component budget-constrained POMDP by finding the optimal budget split among the individual component POMDPs. The optimal budget split is posed as a welfare maximization problem and the solution is computed by exploiting the concavity of the value function. We illustrate the effectiveness of the proposed algorithm by proposing a maintenance and inspection policy for a group of real-world infrastructure components with different deterioration dynamics, inspection and maintenance costs. We show that the proposed algorithm vastly outperforms the policies currently used in practice.**

Index Terms—**Optimization algorithms, Markov processes, distributed control.**

I. INTRODUCTION

SEQUENTIAL decision-making is an integral component of many real world problems like machine maintenance, structural inspection and autonomous robotics [1]. Markov Decision Processes (MDPs) have provided an efficient framework to model and solve such problems while accounting for

Manav Vora, Pranay Thangeda, and Melkior Ornik are with the Department of Aerospace Engineering and the Coordinated Science Laboratory, University of Illinois Urbana-Champaign, Urbana, IL 61801 USA (e-mail: mkvora2@illinois.edu; pranayt2@illinois.edu; mornik@illinois.edu).

Michael N. Grussing is with the Engineer Research and Development Center, U.S. Army Corps of Engineers, Champaign, IL 61822 USA (e-mail: michael.n.grussing@erdc.dren.mil).

the corresponding uncertainty [2]. POMDPs are a generalized version of MDPs, allowing for more uncertainty to be accounted for in the form of partial observability of the system state [3]. However, finding optimal policies for POMDPs is much more computationally intensive as compared to MDPs and is PSPACE-complete [4]. Synthesis of optimal policies for POMDPs is a classical problem and many algorithms have been proposed for the same [5], [6], [7].

This letter considers optimal planning for a class of structured budget-constrained POMDPs. This setting is motivated by infrastructure maintenance planning – a widely studied problem [8], [9] that involves finding the optimal policy for maintenance and inspection, of an infrastructure component or a group of components within a certain budget [10], [11]. For simplicity of planning, the stochastic dynamics of multi-component systems are modelled using a POMDP [12]. Also, the dynamics of individual components are assumed to be independent of each other [13]. We will thus model such a setting by multi-component budget-constrained POMDPs where the transition probabilities of the individual component POMDPs are decoupled.

An algorithm for solving cost-constrained POMDPs has been proposed in [14]. However, this algorithm becomes computationally infeasible for multi-component POMDPs with very large state spaces. A POMDP-based solution for optimal maintenance and inspection of structures using Dynamic Bayesian Networks is presented in [15]. However, in addition to computational infeasibility, this algorithm does not account for budget constraints. Optimal allocation for MDPs has been studied in [16]. This letter models the statistical ranking and selection problem as an MDP and derives an approximately optimal allocation policy using value function approximation. In our work, we study optimal budget allocation for multi-component POMDPs, for infrastructure management. A method for solving budget-constrained MDPs has been presented in [17]. This letter introduces a budgeted-MDP model which includes the budget as an implicit part of the state. This is because this letter uses a cost function, similar to Constrained MDPs [18], to keep a track of the cost incurred by the policy. Hence, this algorithm cannot be directly extended to POMDPs because the partial observability of the state would cause a violation of the budget constraint in some cases.

In this letter, we propose a computationally efficient algorithm for optimal policy synthesis of a multi-component budget-constrained POMDP. Our contributions here are:

- we introduce a b-POMDP model to facilitate strict adherence to budget constraints in POMDPs,
- we obtain an approximately optimal policy for multi-component POMDPs by finding the optimal budget split among the individual component POMDPs.

The b-POMDP model includes the total cost incurred up to a time instant k explicitly as a part of the state vector. We show that the value function for a particular class of b-POMDPs is a concave function of the budget. Next, we compute the optimal policy for the individual component POMDPs by modeling them as b-POMDPs and using an online solver like POMCP [7]. Doing so gives us the approximate maximal total reward collected by the policy in terms of the budget allocated to the b-POMDP. We use these rewards to calculate the optimum distribution of the total budget among the individual component b-POMDPs and find the approximately optimal policy of the multi-component POMDP. The budget splitting is posed as a welfare maximization problem constrained by the total budget of the multi-component POMDP. The concave nature of the value function renders this as a convex optimization problem, thus guaranteeing a global optimum. We demonstrate the utility of the proposed algorithm by finding optimal maintenance and inspection policies for multiple components of a realistic general administrative building, subject to a budget. Based on this real data, we show that our algorithm vastly outperforms the policy currently used in practice.

II. Preliminaries and Background

In this section, we provide background on Partially Observable Markov Decision Processes for sequential decision-making with stochastic dynamics. We start by defining the notation used in this letter.

Given a finite set $\mathcal{A}$, $|\mathcal{A}|$ denotes its cardinality and $\Delta(\mathcal{A})$ denotes the set of all probability distributions over the set $\mathcal{A}$. Notation $\mathbb{N}_0$ denotes the set of natural numbers including 0, i.e., $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$. The symbols $\lfloor . \rfloor$ and $\lceil . \rceil$ denote the floor and ceiling functions respectively.

A. Partially Observable Markov Decision Process

A discrete-time finite-horizon POMDP [19], [20] M is specified by the 8-tuple $(\mathcal{S}, A, \pi, T, \Omega, O, R, H)$, where $\mathcal{S}$ denotes a finite set of states, A denotes a finite set of actions, $\pi : \mathcal{S} \to A$ denotes the policy which specifies the action to take in a given state s and $T : \mathcal{S} \times A \to \Delta(\mathcal{S})$ denotes the transition probability function, where $\Delta(\mathcal{S})$ is the space of probability distributions over $\mathcal{S}$. Furthermore, Ω denotes a finite set of observations and $O : \Omega \times \mathcal{S} \times A \to \Delta(\Omega)$ denotes the observation probability function where $\Delta(\Omega)$ is analogous to $\Delta(\mathcal{S})$. Finally, $R : \mathcal{S} \times A \to [0, R_{max}]$ denotes the reward function and $H \in \mathbb{N}_0$ denotes the finite planning horizon.

For the above POMDP, at each time step, the environment is in some state $s \in \mathcal{S}$ and the agent interacts with the environment by taking an action $a \in A$. Doing so results in the environment transitioning to a new state $\bar{s} \in \mathcal{S}$ in the next time step with probability $T(s, a, \bar{s})$. Simultaneously, the agent receives an observation $o \in \Omega$ regarding the state of the environment with probability $O(\bar{s}, a, o)$ which depends on the new state of the environment and the action taken by the agent. In

a POMDP the agent doesn't have access to the true state of the environment. However, the agent can update it's belief about the true state of the environment using this observation. The agent also receives a reward $R(s, a)$.

The problem of optimal policy synthesis for a finite-horizon POMDP is that of choosing a sequence of actions which maximizes the expected total reward.

III. Problem Formulation

We consider a multi-component POMDP, which is a collection of n component POMDPs. The component POMDPs are weakly-coupled in the sense that they have independent transition probabilities and are connected only by the shared total budget. In this letter, we consider optimal policy synthesis for POMDPs with budget, i.e., each action incurs a cost and the total cost incurred by the optimal policy is limited by the budget. We first formally define the multi-component POMDP with a budget and then define the problem of finding the optimal policy for such a POMDP.

A. Multi-Component Decoupled POMDP With Shared Budget

For a multi-component POMDP, the state space $\mathcal{S} \subseteq \mathbb{N}_0^n$ is given by $\mathcal{S} = \mathcal{S}^1 \times \mathcal{S}^2 \times \ldots \times \mathcal{S}^n$. The state space $\mathcal{S}^i \in \mathbb{N}_0$, for component i, is given by $\mathcal{S}^i = \{0, 1, 2, \ldots, s_{max}\}$, where $s_{max} \in \mathbb{N}_0$. The state $s_k \in \mathcal{S}$, at time step k is given by $s_k = \{s_k^1, s_k^2, \ldots, s_k^n\}$ where $s_k^i \in \mathcal{S}^i$ represents the state of component i at time step k. The action space is given by

$$A = \prod_{i=1}^{n} A^i,$$

where the action space A^i for component i is given by $A^i = \{d^i, q^i, m^i\}$. Action d^i lets the component move to a new state according to the transition probabilities. The action q^i provides an observation which is equal to the next state $\bar{s}^i$ of the component and action m^i drives the component state to s_{max}. The transition probability function for the multi-component POMDP for $s, \bar{s} \in \mathcal{S}$ and $a \in A$ is given by

$$T(s, a, \bar{s}) = \prod_{i=1}^{n} T^i(s^i, a^i, \bar{s}^i).$$

T^i denotes the transition probability function for component i and is defined as

$$T^i(s, a, \bar{s}^i) = \begin{cases} 1, & \text{if } \bar{s}^i = s_{max} \text{ and } a^i = m^i, \\ p^i(s^i, a^i, \bar{s}^i), & \text{if } \bar{s}^i \leq s^i \text{ and } a^i \in \{d^i, q^i\}, \\ 1, & \text{if } \bar{s}^i = 0 = s^i \text{ and } a^i \in A^i, \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

The probability $p^i(s^i, a^i, \bar{s}^i)$ is chosen according to a probability distribution specific to component i. From the above equation, it can be observed that 0 is an absorbing state.

The observation space is given by $\Omega = \mathcal{S} \cup \{e\}$, where $e \in \mathbb{N}_0$ is an observation that does not provide any information regarding the true state of the system, i.e., $e \notin \mathcal{S}^i$ for all

$i \in \{1, 2, \ldots, n\}$. The observation function for the multi-component POMDP is given by

$$O(\bar{s}, a, o) = \prod_{i=0}^{n} O^i(\bar{s}^i, a^i, o^i).$$

Here, O^i is the observation probability function for component i and is defined as

$$O^i(\bar{s}^i, a^i, o^i) = \begin{cases} 1, & \text{if } o^i = \bar{s}^i \text{ and } a^i \in \{q^i, m^i\} \\ 1, & \text{if } o^i = e \text{ and } a^i = d^i \\ 0, & \text{otherwise.} \end{cases}$$

For each component i, each action d^i, q^i and m^i incurs a cost c_d^i, c_q^i and c_m^i respectively, against a total budget B.

B. Problem Statement

For a multi-component POMDP, given by the formulation in the previous section, we consider the problem of finding an optimal policy π^* which maximizes the time before reaching the absorbing state. Mathematically, π^* maximizes t such that $s_t > 0$. Furthermore, for a horizon of length H, π^* should be such that the total cost incurred for the multi-component POMDP does not exceed the total budget. We propose an approximately optimal solution for the above problem through a two-step approach. In the first step, we will solve a single-component POMDP for any given budget. In the second step, we will partition the total budget B into budgets for each individual component.

IV. SOLUTION APPROACH

In this section, we detail our methodology for solving the problem of optimal policy synthesis of a multi-component POMDP. First, we introduce the b-POMDP model and discuss how to solve a single b-POMDP. Next, we discuss why the value function for such a POMDP is a concave function of the budget. Finally, we present our proposed method for finding the optimal policy for an n-component POMDP by computing the optimal split of the total budget, among the individual component POMDPs.

A. Budgeted-POMDP Model (b-POMDP)

Our main goal is to find an optimal policy for a POMDP while adhering to a total budget for actions. The budgeted-MDP model in [17] tracks the incurred cost using a cost function similar to Constrained MDPs [18]. This model can't be extended directly to a POMDP because the partial observability of the state would lead to budget violation in some cases. Hence, we introduce a new b-POMDP model. In a b-POMDP, the budget constraint is incorporated by augmenting the total cost incurred upto time instant k, to each state of the state space. Thus, if we consider a single component POMDP with a total budget B, the modified state at time instant k according to the b-POMDP formulation is given by (s_k, c_k) where s_k is as defined in the previous section, for $i = 1$. We assume that unlike s_k, c_k is completely observable at all time instants. The transition function for the cost component of the state is given by:

$$T_c(c'|c, a) = \begin{cases} 1, & \text{if } c' = c + c_m \text{ and } a = m \\ 1, & \text{if } c' = c + c_q \text{ and } a = q \\ 1, & \text{if } c' = c + c_d \text{ and } a = d \\ 0, & \text{otherwise.} \end{cases}$$

The transition function for the overall b-POMDP is:

$$T'((s, c), a, (\bar{s}, c')) = T(s, a, \bar{s}) T_c(c, a, c'), \quad (2)$$

where $T(s, a, \bar{s})$ is defined in Section III-A. The new formulation prevents the policy from violating the budget at any time instant k. This is done by making $c_k > min\{B - c_m, B - c_i\}$ an absorbing state, similar to $s = 0$. The reward function for the b-POMDP is given by:

$$R'((s, c), a) = \begin{cases} r_1 > 0, & \text{if } s > 0, \\ r_2 = 0, & \text{if } s = 0. \end{cases}$$

To find an optimal policy for a b-POMDP, we use the method of Monte-Carlo Planning in POMDPs (POMCP [7]). POMCP is an online planning algorithm for large POMDPs, which combines a Monte-Carlo update of the agent's belief with a Monte-Carlo tree search for the best action from the current belief state. For a b-POMDP, the maximal collected reward (value function) obtained using POMCP will be a function of the budget B associated with it. We will now prove that this value function is concave in the budget for a special subclass of our overall problem.

B. Sketch of Proof for Concavity of Value Function

Consider an MDP with state space $\mathcal{S}_{MDP} = \{0, 1, 2 \ldots s_{max}\}$, where $s_{max} \in \mathbb{N}_0$ and action space $A_{MDP} = \{m, d\}$. The state of the system decreases by $d_0 \in \mathbb{N}_0$ unless we perform action m. The transition probability function is defined as:

$$T(s'|s, a) = \begin{cases} 1, & \text{if } \bar{s} = s_{max} \text{ and } a = m \\ 1, & \text{if } \bar{s} = s - d_0 \text{ and } a = d \\ 1, & \text{if } \bar{s} = 0 = s \text{ and } a \in A_{MDP} \\ 0, & \text{otherwise,} \end{cases}$$

From the above transition function, we can clearly see that state 0 is an absorbing state. The cost for the d action is $c_d = 0$ and the cost for the m action is $c_m = 1$. Let the available budget be denoted by $b \in \mathbb{N}_0$. This means that we can perform the action m at most b times. The reward function is the same as that for a b-POMDP.

Let V_H be the value function for a given horizon length with H steps to go. Then we prove that for all values of H, the following relations hold:

1) For a given initial state s_0, the value function $V_H(s_0, b)$ is an increasing function of b.
2) For $s_0 > d_0$, $V_H(s_0, b)$ is constant with respect to the budget for all $b \geq \lfloor H/2 \rfloor$,
3) For $s_0 \leq d_0$, $V_H(s_0, b)$ is constant with respect to the budget for all $b \geq \lceil H/2 \rceil$,
4) For budget values $b'' = b' + 1 = b + 2$ with $b \geq 0$ the following relation holds

$$V_H(s_0, b'') - V_H(s_0, b') \leq V_H(s_0, b') - V_H(s_0, b).$$

The first claim is trivial to prove by comparing the value functions, corresponding to the same optimal policy and initial state s_0 for any two budget values. We prove the other three claims using mathematical induction. First, we show that these properties hold true for a horizon of length 0. Next, we assume that these properties are true for a horizon of length H and use this to show that they hold true for a horizon of length $H + 1$. We do this by manipulating the Bellman equation given by

$$V_{H+1}(s_0, b) = r(s_0, a) + V_H(\bar{s}, b'),$$

where $s_0, \bar{s} \in \mathcal{S}$, $a \in A$ and $\bar{s}$ and b' are decided based on s_0 and a. In summary, the above claims imply that for $b' = \alpha b + (1 - \alpha) b''$ for some $\alpha \in [0, 1]$,

$$V_H(s_0, b') = \alpha V_H(s_0, b) + (1 - \alpha) V_H(s_0, b'').$$

Hence, the value function is a concave function of the budget.

A detailed version of this proof can be found in this letter's extended version.[1] The proof can be easily extended to any general $c_m \in \mathbb{R}^+$ by scaling the costs and budget with $1/c_m$. However, the proof works under heavy assumptions of full observability and deterministic transitions. Nonetheless, we empirically observe that same property is often true for general systems, partially observable and stochastic, and we believe the same proof approach might work for those systems, but we leave it for future work.

We will now use this concavity property to obtain an approximately optimal policy for an n-component POMDP by optimally allocating the budget to each component POMDP.

C. Optimal Policy Synthesis for Multi-Component POMDP

Consider a multi-component POMDP with n components and budget B as described in Section III-A. The size of the state space is $(|\mathcal{S}|)^n$ where $|\mathcal{S}|$ is the size of the state space of each component POMDP. Also, the size of the total action space is 3^n. Directly applying a POMDP solver to such a large state and action space may not be computationally feasible. Hence, we propose an algorithm which decouples the n component POMDPs by allocating a portion of the total budget to each of them prior to the beginning of the system run. We then compute the approximate value function for each component POMDP as a function of the budget and then using that, obtain the optimal split of the total budget.

Given a total budget B, we assume that the ith component POMDP is alloted a budget b_i from the total budget. Hence,

$$b_1 + b_2 + \ldots + b_n = B \qquad (3)$$

We now have n independent POMDPs, where each POMDP has its own total budget. We formulate each of them as a b-POMDP and solve each b-POMDP using the POMCP algorithm as discussed in Section IV-A. Let the maximal collected reward, for component i, obtained using the POMCP algorithm, for a given initial state s_0 and horizon H, be denoted by $V_H^i(s_0, b_i)$. We can then find the optimal budget split among the n b-POMDPs by solving a welfare maximization problem. Welfare maximization is the concept of maximizing the overall well-being or welfare of a society, and is achieved by maximizing some measure of social welfare (e.g., maximal collected

[1] https://arxiv.org/pdf/2303.10302.pdf

reward). We thus maximize the total maximal collected reward, for all components, with respect to b_i while adhering to the constraint in (3), i.e.,

$$\max_{b_i} \sum_{i=1}^n V_H^i(s_0, b_i)$$

$$\text{s.t.} \sum_{i=1}^n b_i = B. \qquad (4)$$

Using the results we proved in Section IV-B, we know that $V_H(s_0, b_i)$ is a concave function of b_i in the special case mentioned in Section IV-B and emperically observe it to be concave in general. The welfare maximization problem then becomes a constrained convex optimization problem. Hence, it can be solved easily and is guaranteed to have a global optimum. The solution to (4) provides the optimal budget allocation for each b-POMDP which in-turn gives us the optimal policy for all n component POMDPs. Let $\pi_i : \mathcal{S}^i \to A^i$ be a policy for component i with budget b_i obtained using the POMCP algorithm. Then, we define the overall policy, for the n-component POMDP, $\pi : \mathcal{S} \to A$ by

$$\pi = \prod_{i=1}^n \pi_i.$$

While such a policy is naturally not guaranteed to be generally optimal on the multi-component POMDP, it provably satisfies the budgetary constraints and performs well in practice. To illustrate its performance on real data, we now move to the implementation and evaluation section.

V. IMPLEMENTATION AND EVALUATION

In this section we illustrate the utility of the proposed approach for multi-component decision making with budgetary-constraints. In particular, we compare the policy described above with existing approaches on a scenario of multi-component building management. Our implementation utilizes the POMDP.jl [21] Julia package for efficiently solving the budgeted-POMDPs using POMCP, as well as CVXPY [22] for solving the convex optimization formulation of the budget allocation problem. The initial budget-split for solving the budget allocation problem is chosen randomly while satisfying the constraint of (3).

We model the components that comprise a typical administration building with a total size of 10,000 sq. ft. The building comprises multiple components such as lighting systems, roofing components, boilers, and carpeting, where each component's cost of replacement and inspection are based on empirically derived industry averages. Each component's health is defined by the Condition Index (CI) [23], which takes values between 0 and 100. The condition deteriorates stochastically over time, depending on various factors, and can only be observed through explicit inspections, which incur a cost. The component fails when the CI reaches below a *failure threshold*, which we assume to be 0. Components can also be replaced, restoring their CI to its full value.

The building is associated with an average maintenance budget of \$2,200,000 for a given period of interest. Using historic CI data for each component, we synthesize the transition probabilities of their corresponding Partially Observable
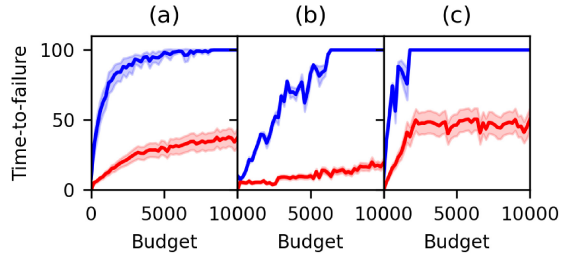
Fig. 1. Comparison of the proposed (blue) and baseline (red) approaches using time-to-failure for a range of budget values. (a) Overall results obtained by averaging over all components. (b) Results for the Air Handling Unit component with a replacement cost of 250 units. (c) Results for the Lighting Equipment component with a replacement cost of 24 units.

Markov Decision Processes (POMDPs). We consider 20 sustainable components from the building, with replacement costs ranging from 0.15% to 3% of the total budget and inspection costs ranging from 0.01% to 0.03% of the total budget. We scale the total budget to 10,000 units and appropriately scale the replacement and inspection costs of all components while ensuring that they are rounded to the nearest integers. The decision-maker's objective is to maximize the time until failure of the components by effectively allocating the budget among the components and taking replacement and inspections when needed. As in Section IV-C, we model this objective as a POMDP by assigning a reward of 1 when the CI is greater than the failure threshold and 0 otherwise, and modeling the state of 0 health and budget exhaustion as absorbing states. For our experiments, we consider simulations with a horizon of up to 100 decision steps with a 1 year step size.

A. Maintenance Policy Synthesis

In this section we compare the maintenance and inspection policies obtained from the proposed POMDP-based model with a realistic baseline approach. In the baseline approach, a building manager typically schedules component inspection at a regular interval and the true health of the component is only obtained at these regular intervals. In the absence of an inspection, the CI of a component at a given time step is estimated to be the most probable CI state as determined by its CI transition dynamics. The baseline policy used in this section replaces the component if its estimated CI is less than a pre-determined threshold.

We use *time-to-failure* (TTF), defined as the number of simulation steps until failure, as the performance metric. We run experiments to calculate the TTF for each component by averaging the values obtained over 5 independent simulations with 100 maximum possible simulation steps. We set the maximum tree depth for POMCP rollouts to 50 and use a UCB exploration constant of 10. In the baseline policy, we inspect the CI every 5 steps and replace the component if the estimated CI is below 15. Figure 1 shows the simulation results comparing the TTF obtained for different budget values using the baseline and the proposed approach. The proposed approach provides a clear advantage over baseline strategy over the entire range of budget values for all 20 components, irrespective of the replacement costs.
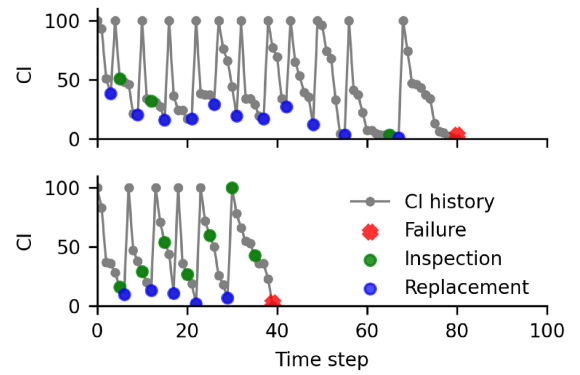

Fig. 2. Sample condition index (CI) histories illustrating the performance of the proposed policy when compared to the baseline for the Boiler component with a replacement cost of 45 units, an inspection cost of 1 unit, and a total budget of 500 units. (a) CI history using proposed approach showing failure at 80 time steps. (b) Baseline approach failing at 39 time steps.

Figure 2 shows sample CI histories for the same component obtained from simulations using the proposed approach and the baseline policy. The proposed approach takes inspection and replacement actions only when deemed necessary based on the latest belief estimate and the potential loss of value due to an inaccurate estimate or due to not taking a replacement action. Such a behavior holds true for every component without any component specific parameter tuning. On the other hand, the estimated state from the baseline policy based on the most-probable transition may not always be the same as the real transition, ultimately resulting in early failures. Although it is possible to enhance the baseline by incorporating component-specific parameters and budget-aware heuristics, our experiments indicate that its performance still lags behind the proposed approach, particularly when the budget is tightly constrained.

B. Budget Allocation

We demonstrate the effectiveness of our proposed budget allocation approach by comparing it to a baseline that depends on two component properties: (i) mean-time-to-failure (MTTF) which is the expected number of steps a component takes for its condition index to go below the failure threshold when starting from maximum possible condition index, and (ii) the replacement cost of the component. The baseline allocation is proportional to ratio of the component's replacement cost and MTTF.

We quantify the performance of the budget allocation algorithms by running 20 independent simulations over all the components using the allocated budgets and calculating the overall TTF for the building. To ensure fairness, we compare both budget allocation algorithms by running simulations using policies obtained by the same decision making strategy: the POMCP-based approach. Figure 3 summarizes the TTF results from the baseline and the proposed approaches for all components in the building. The proposed allocation approach achieves an overall TTF of 1510, outperforming the baseline that achieves an overall TTF of 1355. Hence the proposed approach maximizes the overall TTF in accordance with the objective defined by (4). Analyzing individual component data
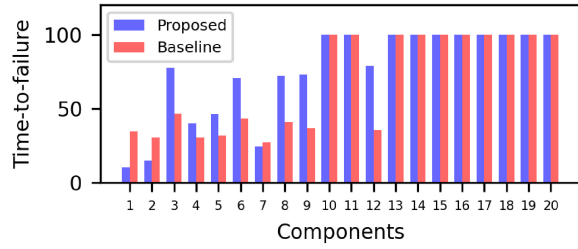
Fig. 3. Comparison of baseline and proposed budget allocation approaches for the all 20 components for an overall budget of 10,000 units.

TABLE I

COMPARISON OF TIME TAKEN TO FIND OPTIMAL BUDGET SPLIT AMONG 5, 10 AND 20 COMPONENTS RESPECTIVELY, FOR A TOTAL BUDGET OF 10000 UNITS

Number of Components	Time (mean $\pm$ std. dev. of 7 runs)
5	$333ms \pm 21.4ms$
10	$412ms \pm 37.1ms$
20	$552ms \pm 31.6ms$

we observe that instances where the proposed approach underperforms the baseline exhibit only slight differences in TTF values. In contrast, when the proposed approach outperforms the baseline, we observe a significant improvement. Note that the maximal TTF of 100 is achieved by both strategies for 50% of components, the proposed strategy performs better for 35% of the components, and the baseline performs better only for 15% of the components.

The results in Table I present the time taken to solve the optimization problem given by (4). As can be clearly observed, the solution time increases with an increase in the number of components. However, the values are always of the order of milliseconds.

VI. FUTURE WORK

There are two possible directions of future work. While the proposed approach makes it possible to compute an approximately optimal policy in a feasible amount of time, the computational cost of using the POMCP algorithm is still high. Our first direction of work is to reduce this cost by incorporating a learning framework so as eliminate repeated runs of POMCP. Finally, the budget allocation scheme is fixed in the sense that the budget-split is done before the start of the planning horizon. The second direction of work is to consider other efficient budget allocation methods. A sequential algorithm for optimal computing budget allocation is presented in [24]. Applying this algorithm to our problem may result in more accurate budget allocation. Similarly, another method which can be explored is the allocating method presented in [25]. Also, following these methods may allow us to generalize our algorithm for cases where the value function does not satisfy the concavity property. Furthermore, another direction is to derive an optimal dynamic budget allocation scheme to account for change in transition probabilities of the component states during the planning horizon and also account for a cyclical budget.

REFERENCES

[1] A. R. Cassandra, "A survey of POMDP applications," in *Proc. Fall Symp. Planning Partially Observable Markov Decis. Process. (AAAI)*, 1998, pp. 1–9.

[2] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, 1st ed. Hoboken, NJ, USA: Wiley, 1994.

[3] K. J. Åström, "Optimal control of Markov processes with incomplete state information," *J. Math. Anal. Appl.*, vol. 10, no. 1, pp. 174–205, 1965.

[4] C. H. Papadimitriou and J. N. Tsitsiklis, "The complexity of Markov decision processes," *Math. Oper. Res.*, vol. 12, no. 3, pp. 441–450, 1987.

[5] A. Cassandra, M. L. Littman, and N. L. Zhang, "Incremental pruning: A simple, fast, exact method for partially observable Markov decision processes," in *Proc. 13th Conf. Uncertainty Artif. Intell.*, 1997, pp. 54–61.

[6] T. Smith and R. Simmons, "Heuristic search value iteration for POMDPs," in *Proc. 20th Conf. Uncertainty Artif. Intell.*, 2004, pp. 520–527.

[7] D. Silver and J. Veness, "Monte-Carlo planning in large POMDPs," in *Proc. 23rd Int. Conf. Neural Inf. Process. Syst.*, vol. 2, 2010, pp. 2164–2172.

[8] M. Mishra, P. B. Lourenço, and G. Ramana, "Structural health monitoring of civil engineering structures by using the Internet of Things: A review," *J. Build. Eng.*, vol. 48, May 2022, Art. no. 103954.

[9] Y. Zhou, B. Li, and T. R. Lin, "Maintenance optimisation of multicomponent systems using hierarchical coordinated reinforcement learning," *Rel. Eng. Syst. Safety*, vol. 217, Sep. 2022, Art. no. 108615.

[10] X. Lei, Y. Xia, L. Deng, and L. Sun, "A deep reinforcement learning framework for life-cycle maintenance planning of regional deteriorating bridges using inspection data," *Struct. Multidiscipl. Optim.*, vol. 65, no. 5, p. 149, 2022.

[11] O. E. Hamshary, M. Abouhamad, and M. Marzouk, "Integrated maintenance planning approach to optimize budget allocation for subway operating systems," *Tunnelling Underground Space Technol.*, vol. 121, Mar. 2022, At. no. 104322.

[12] M. Memarzadeh and M. Pozzi, "Integrated inspection scheduling and maintenance planning for infrastructure systems," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 31, no. 6, pp. 403–415, 2016.

[13] C. P. Andriotis and K. G. Papakonstantinou, "Deep reinforcement learning driven inspection and maintenance planning under incomplete information and constraints," *Rel. Eng. Syst. Safety*, vol. 212, Aug. 2021, Art. no. 107551.

[14] J. Lee, G.-H. Kim, P. Poupart, and K.-E. Kim, "Monte-carlo tree search for constrained POMDPs," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 7934–7943.

[15] P. G. Morato, K. G. Papakonstantinou, C. P. Andriotis, J. Nielsen, and P. Rigo, "Optimal inspection and maintenance planning for deteriorating structural components through dynamic Bayesian networks and Markov decision processes," *Struct. Safety*, vol. 94, Jan. 2022, Art. no. 102140.

[16] Y. Peng, E. K. P. Chong, C.-H. Chen, and M. C. Fu, "Ranking and selection as stochastic control," *IEEE Trans. Autom. Control*, vol. 63, no. 8, pp. 2359–2373, Aug. 2018.

[17] C. Boutilier and T. Lu, "Budget allocation using weakly coupled, constrained Markov decision processes," in *Proc. 32nd Conf. Uncertainty Artif. Intell.*, 2016, pp. 52–61.

[18] E. Altman, *Constrained Markov Decision Processes*. London, U.K.: Chapman and Hall, 1999.

[19] A. R. Cassandra, L. P. Kaelbling, and M. L. Littman, "Acting optimally in partially observable stochastic domains," in *Proc. 12th AAAI Nat. Conf. Artif. Intell.*, 1994, pp. 1023–1028.

[20] D. Braziunas, "POMDP solution methods: A survey," Dept. Comput. Sci., Univ. Toronto, Toronto, ON, Canada, Rep. TR-2003, 2003.

[21] M. Egorov, Z. N. Sunberg, E. Balaban, T. A. Wheeler, J. K. Gupta, and M. J. Kochenderfer, "POMDPs.jl: A framework for sequential decision making under uncertainty," *J. Mach. Learn. Res.*, vol. 18, no. 26, pp. 1–5, 2017.

[22] S. Diamond and S. Boyd, "CVXPY: A python-embedded modeling language for convex optimization," *J. Mach. Learn. Res.*, vol. 17, no. 83, pp. 1–5, 2016.

[23] M. N. Grussing, D. R. Uzarski, and L. R. Marrano, "Condition and reliability prediction models using the Weibull probability distribution," in *Proc. Appl. Adv. Technol. Transp.*, 2006, pp. 19–24.

[24] C.-H. Chen, J. Lin, E. Yücesan, and S. E. Chick, "Simulation budget allocation for further enhancing the efficiency of ordinal optimization," *Discr. Event Dyn. Syst.*, vol. 10, no. 3, pp. 251–270, 2000.

[25] C.-H. Chen, "A lower bound for the correct subset-selection probability and its application to discrete-event system simulations," *IEEE Trans. Autom. Control.*, vol. 41, no. 8, pp. 1227–1231, Aug. 1996.