Resource Allocation in Schedule Constrained Research and Development Programs

Jacob T. Needels
Department of Aeronautics and Astronautics, Stanford University

Resource allocation in schedule constrained research and development programs is formulated as an explicit Partially Observable Markov Decision Process, and solved using both offline and online solution methods. An offline fast informed bound (FIB) method performed best, with comparable run times to an online modified Monte-Carlo Tree Search method for the small state and action spaces assessed.

I. Nomenclature

a	=	Action
$\mathcal{A}$	=	Action Space
$\mathcal{B}(\cdot,\cdot)$	=	Binomial Distribution
o	=	Observation
$\mathcal{O}$	=	Observation Space
$O(\cdot\|\cdot)$	=	Observation Function
s	=	State
$\mathcal{S}$	=	State Space
$T(\cdot\|\cdot)$	=	Transition Function
$R(\cdot,\cdot)$	=	Reward Function

II. Introduction

Research and development (R&D) activity is necessary to develop and improve products and technology in support of both public and private sector objectives. Effective research and development programs are increasingly necessary due to the continual reduction in project development timelines and product life-cycles[1]. In 2016, combined public and private investment in R&D exceeded 500 billion dollars, meaning that effective management of this funding has a significant impact on organizational budgeting[2].

In general, the goal of an R&D program is to develop a product or service to a specified level of technical maturity. Here we consider the case of a program with externally imposed schedule milestones specified by stakeholders. An example of such an arrangement is a cost-sharing agreement between a public agency and a private corporation, where the corporation must achieve technical milestones by deadlines specified by the public agency in order to continue to receive funding. Unbounded increase in program resource allocation is typically not an option, due to the finite resource budget which must be shared among other programs and departments. Therefore, the program must be selective regarding resources, expending the minimal resources necessary to achieve technical maturity by the required milestones.

The responsibility of meeting milestone requirements typically falls on program management. Here we consider the case where management's primary control of program performance is resource allocation. Increasing resource allocation increases the likelihood that the program will move from behind schedule to on schedule. However, due to constraints on program resources, there is a penalty associated with increasing program resource allocation, such that program budgeting is a trade-off between minimizing allocated resources and ensuring that the program meets schedule requirements. Determining whether a program is on track to meet milestones is made difficult since a manager is typically not able to review every technical detail, and instead must rely on status updates provided by subordinates. In general, we expect these status reports to include "noise": inaccurate reflection of the true status of the project due to

career risk , bias, lack of knowledge, etc.

A. Objective

This paper proposes a method of formulating and solving optimal resource allocation for research and development programs an explicit Partially Observable Markov Decision Process (POMDP) that accounts for the stochastic nature of program progress and observation of program status. A primary objective is developing a tractable model that is also representative of system characteristics. Both online and offline solution methods are evaluated and compared for this application.

B. Relevant Literature

The formulation of optimal resource allocation problems as POMDPs has been treated in the literature for several applications. Firestone writes about the formulation of battle management in military applications as a POMDP, solved using a combination of offline and online methods for efficient calculation of resource assignment and action allocation [3]. Firestone introduces a novel method augmenting the state vector to incorporate constraints on resource depletion into policy generation[3]. McDonald-Madden demonstrate online solution of a POMDP, as applied to optimal resource allocation in conservation programs [4]. While these works deal with the general topic of resource allocation in program management, they do not explicitly examine the effect of a milestone goal state in a finite time horizon, which is the focus of this project.

III. Methods

A. Problem Formulation

A POMDP is defined over a set of states $\mathcal{S}$, actions $\mathcal{A}$, and observations O. In order to reduce computational complexity, discretized state, action, and observation spaces are used in this study. The state space consists of a Boolean variable indicating whether the program is on schedule or behind schedule at a given time step. The space of actions include increasing, decreasing, or not changing program budget, and the set of possible observations is the same as the set of possible states, as shown below.

$$\mathcal{S} = [On\ Schedule, Behind\ Schedule]$$
$$O = [On\ Schedule, Behind\ Schedule]$$
$$\mathcal{A} = [Increase, None, Decrease]$$

A transition function, $T(s'|s,a)$, is defined that captures the stochastic state dynamics for different state-action pairs. A stationary transition function was used to simplify the model. Values for state transition probabilities were defined arbitrarily for this project, although in general it might be informed by experience or data. Transition probabilities are defined in terms of a binomial distribution, $\mathrm{B}(n,p)$, parameterized by a number of trials, n, and probability of transition to a Boolean True value, p.

$$T(s'|s, Increase) \sim \mathcal{B}(1, 0.7) \ \ \forall \ \ s, s' \in \mathcal{S}$$
$$T(s'|s, Decrease) \sim \mathcal{B}(1, 0.2) \ \ \forall \ \ s, s' \in \mathcal{S}$$
$$T(s'|s, None) \sim \mathcal{B}(1, 0.75) \ \ \forall \ \ s, s' \in \mathcal{S}$$

A reward function, $R(s,a)$, is defined that rewards both being ahead of schedule and decreasing budget, while penalizing being behind schedule and increasing program budget. Values for the reward function can be tuned to program requirements. Reward values chosen for the simulations presented in this paper are given in Table 1 below.

Reward	Value
Behind Schedule	-200
On Schedule	+100
Increase Resources	-50
Decrease Resources	+25

Table 1 Reward Function Values

Observation of system state are modeled by a stationary, noisy observation function. This function models the fact that reports of program status may not accurately reflect system state.

$$O(o|s',a) \sim \mathcal{B}(1,0.9) \;\; \forall \;\; s',a \in \mathcal{S},\mathcal{A}$$

B. Solution Methods

The explicit POMDP described in the previous section was implemented using the POMDPs.jl package in the Julia programming language[5]. Both offline and online solution methods were implemented to solve the formulated POMDP. For offline solution, a fast informed bound (FIB) method was utilized. FIB calculates a single alpha vector for each action, using therupdate iteration as shown in Equation 1 below, where a discount factor, γ, of 0.95 was used for this study. [6].

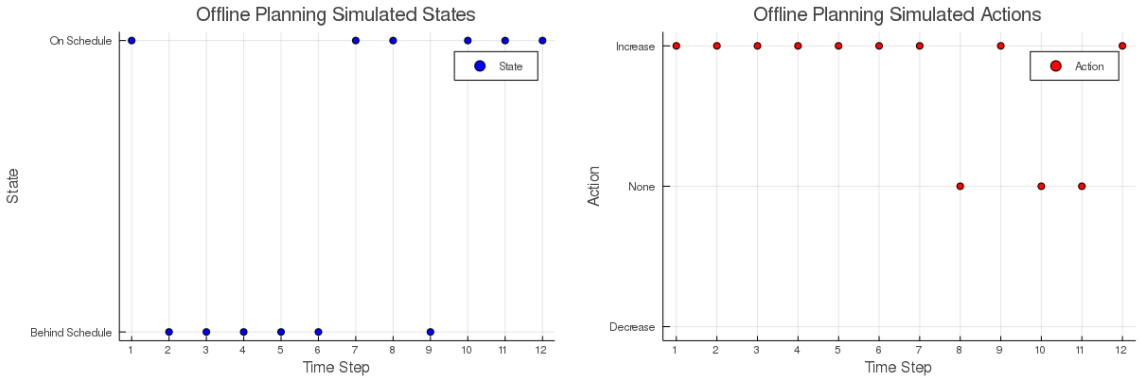
$$\alpha^{k+1} = R(s,a) + \gamma \sum \max_{a'} \sum O(o|s',a)T(s'|s,a)\alpha^{k}(s') \tag{1}$$

A Monte Carlo tree search algorithm, Partially Observable Monte Carlo Planning (POMCP), was used for online planning. POMCP is combines a Monte-Carlo update of the agent's belief state with a Monte-Carlo tree search from the current belief state, making it highly scalable to large discrete state spaces [7].

IV. Simulation Results

A. Offline and Online Simulation Results

Both offline and online solvers were simulated using a time horizon of 12 steps to model performance through one fiscal year with monthly milestone requirements. Results for a single simulation of the offline FIB solver are given in Figure 1 below.



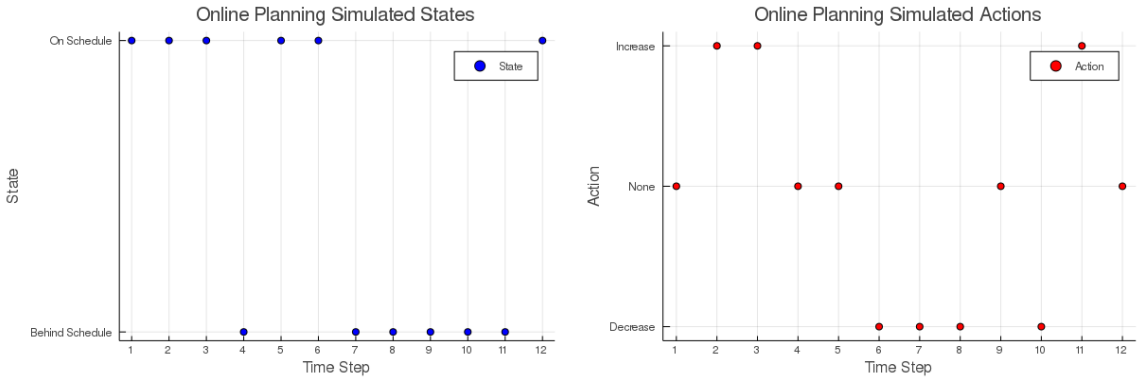
(a) Offline Policy Results - States **(b) Offline Policy Results - Actions**

Fig. 1 Offline Policy Results

From the plot above, it is evident the solver exhibits expected behavior based on the state. When the system is in a state that is behind schedule, the the policy increases program funding to increase the likelihood that the

system will transition to an on schedule state. To minimize the penalty associated with increasing funding, the policy generally selects no action when the state is on schedule. In this case funding is never decreased, due to the relatively small value of the reward associated with decreasing funding relatively to being in a state that is behind schedule.

Results for the online POMCP solver are shown in Figure 2 below.



(a) Online Policy Results - States **(b) Online Policy Results - Actions**

Fig. 2 Online Policy Results

In general, the online solver tended to generate greedier policies, evidenced by a higher incidence of decisions to decreased program budget. This is likely due to the stochastic exploration strategy employed by the POMCP algorithm.

B. Comparison of Offline and Online Solvers

The FIB and POMCP solvers, as well as a baseline random policy generator, were simulated over 1000 runs and the averaged rewards are reported in Table 2 below.

Offline (FIB)	Online (POMCP)	Random
-7.86	-49.10	-58.01

Table 2 Reward Values Averaged Over 1000 Simulations

Both the offline and online policy search score higher than the random policy generator, though the FIB solver performs better than the POMCP solver. This is due to the capability of the FIB solver to exhaustively evaluate the value function over the whole action space, in contrast to the POMCP solver which uses a stochastic exploration strategy and is not guaranteed to converge on an optimal solution.

V. Conclusions and Future Work

A simplified model of resource allocation for research and development programs was formulated as a POMDP. A reward model was developed associated with state-action pairs, that incentivizes meeting schedule milestones and minimizing allocated resources. A stochastic state transition model was developed to model the effect of increasing/decreasing programs budget on system dynamics, and a stationary observation model was implemented to model the bias associated with soliciting state information. The model was solved using both online and offline methods, and the results were compared against a baseline random policy generator.

The offline FIB solver performed best due to its exhaustive search of the space of possible state actions. While POMCP performed significantly worse than FIB for this application, online methods may be preferred for an expanded state space where application of an offline solver might be computationally intractable.

While this project has demonstrated the viability of program resource allocation as a POMDP, there are significant opportunities for future development. One significant modeling improvement would be implementation of terminal rewards, reflective of reaching the end of a milestone cycle, in the POMDP formulation. Expanded (and potentially continuous) state, action, and observation spaces would offer greater model fidelity. An additional improvement would be use of actual program data to information state transition and observation models in order, allowing validation against a relevant baseline case.

References

[1] Heidenberger, K., and Stummer, C., "Research and Development Project Selection and Resource Allocation: A Review of Quantitative Modelling Approaches," *International Journal of Management Reviews*, 1999, pp. 197–224.

[2] Service, C. R., "US Research and Development Funding Performance: Fact Sheet," , 2019.

[3] Firestone, S., "A Partially Observable Approach to Allocating Resources in a Dynamic Battle Scenario," Ph.D. thesis, Massachusetts Institute of Technology, 2002.

[4] Mcdonald-Madden, E., Chades, I., McCarthy, M., Linkie, M., and Possingham, H., "Allocating Conservation Resources Between Areas Where Persistence of a Species is Uncertain," *Ecological Applications*, 2011.

[5] Egorov, M., Sunberg, Z. N., Balaban, E., Wheeler, T. A., Gupta, J. K., and Kochenderfer, M. J., "POMDPs.jl: A Framework for Sequential Decision Making under Uncertainty," *Journal of Machine Learning Research*, Vol. 18, No. 26, 2017, pp. 1–5. URL `http://jmlr.org/papers/v18/16-300.html`.

[6] Kochenderfer, M., *Decision Making under Uncertainty*, The MIT Press, 2017.

[7] Silver, D., and Veness, J., "Monte-Carlo Planning in Large POMDPs," *Advances in Neural Information Processing Systems 23*, edited by J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta, Curran Associates, Inc., 2010, pp. 2164–2172. URL `http://papers.nips.cc/paper/4031-monte-carlo-planning-in-large-pomdps.pdf`.

[8] Golovin, D., Krause, A., Gardner, B., Converse, S., and Morey, S., "Dynamic Resource Allocation in Conservation Planning," *25th AAAI Conference on Artificial Intelligence*, 2011.