A MARKOV DECISION PROCESS APPROACH TO OPTIMIZING CANCER THERAPY USING MULTIPLE TREATMENT MODALITIES

KELSEY MAASS, UW DEPARTMENT OF APPLIED MATHEMATICS
MINSUN KIM, UW DEPARTMENT OF RADIATION ONCOLOGY

1. INTRODUCTION

There are several different modalities, e.g., surgery, chemotherapy, and radiotherapy, that are currently used to treat cancer. It is common practice to use a combination of these modalities to maximize clinical outcomes, a balance between maximizing tumor damage while minimizing normal tissue toxicity due to treatment. However, multi-modality treatment policies are mostly empirical in current medical practice, and therefore subject to individual clinicians' experiences. We present a novel formulation of optimal multi-modality cancer management using a finite-horizon Markov decision process approach.

2. MODEL FORMULATION

For a fixed number of treatment periods, our model determines the optimal treatment modality at time $t = 1, 2, \ldots, T$ based upon the patient's observed state to maximize the expected patient utility.

Treatment modalities

Treatment modalities are categorized into three types:

	Treatment modality	Restrictions
M_1	High risk, high reward	May be implemented only once
M_2	Lower risk, lower reward than M_1	May be repeated
M_3	Surveillance (no treatment)	Possibility of reducing normal tissue side effect at the risk of worsening tumor progression

Patient state

The patient state $s = (h, \phi, \tau)$ consists of three state variables:

	State variable	Domain description
$h \in \{0, 1\}$	M_1 history	$h = 0 \rightarrow M_1$ has not yet been chosen
$\phi \in \{0, 1, \ldots, m\}$	Normal tissue side effect	$\phi = 0 \rightarrow$ no side effect, $\phi = m \rightarrow$ death
$\tau \in \{0, 1, \ldots, n\}$	Tumor progression	$\tau = 0 \rightarrow$ tumor remission, $\tau = n \rightarrow$ death

State transition probabilities

When treatment modality a is implemented in the t-th treatment period for a patient in state s_t, the patient's state at the $(t+1)$-th treatment period is assumed to be s_{t+1} with probability $P_t(s_{t+1}|s_t, a)$.

- Transition probabilities for h are deterministic:

$$P_t(h_{t+1} = 1|h_t, M_1) = 1 \quad \text{and} \quad P_t(h_{t+1} = h_t|h_t, a) = 1, \quad a \in \{M_2, M_3\}$$

- If M_1 is chosen more than once, the patient will transition to the worst possible state:

$$P_t\Big(s_{t+1} = (1, m, n)|s_t = (1, \phi_t, \tau_t), M_1\Big) = 1$$

- We impose absorbing boundary conditions to simulate patient death and tumor remission:

Absorbing boundary conditions	
Death (side effect)	$P_t\Big(s_{t+1} = (h_t, m, \tau_t)|s_t = (h_t, m, \tau_t), a\Big) = 1$
Death (tumor progression)	$P_t\Big(s_{t+1} = (h_t, \phi_t, n)|s_t = (h_t, \phi_t, n), a\Big) = 1$
Tumor remission	$P_t(\tau_{t+1} = 0|\tau_t = 0, a) = 1$

3. BACKWARD INDUCTION

For each patient state and treatment period, our goal is to maximize the expected patient utility,

$$V_t(s) = \sum_{s'} P_t(s'|s, a)\Big(r_t(s, a, s') + V_{t+1}(s')\Big).$$

After each treatment period the patient receives an intermediate reward, $r_t(s, a, s')$, and the boundary condition $V_{T+1}(s) = r_{T+1}(s)$ quantifies the patient's outcome based on their final state. These are the *intermediate and terminal reward functions*. Maximum patient utility and optimal treatment policy can be solved recursively for all s and t with the well-known *backward induction algorithm*:

Set $V^*_{T+1}(s) = r_{T+1}(s)$ for all s
for $t = T, T-1, \ldots, 1$ **do**
  $V^*_t(s) = \max_a \sum_{s'} P_t(s'|s, a)\Big(r_t(s, a, s') + V_{t+1}(s')\Big)$
  $a^*_t(s) = \arg\max_a \sum_{s'} P_t(s'|s, a)\Big(r_t(s, a, s') + V_{t+1}(s')\Big)$
end for

4. NUMERICAL SIMULATIONS

Using the state transition probabilities below, we demonstrate how the structure of optimal treatment policies change as we vary parameters of the reward functions on the left.

(a) With no intermediate reward function and a terminal reward function composed of functions for side effect and tumor progression, we see that treatment policies become more aggressive as the relative importance of tumor progression increases.

(b) With no intermediate reward function and a terminal reward function composed of functions for side effect and tumor progression, we observe the effects of function shape. First we see that linear reward functions do not result in intuitive treatment policies. Second, we see that increasing d increases the amount of surveillance in the treatment policy.

(c) Finally, we experiment with different intermediate rewards. In the first row of (c), we see that adding an intermediate reward consisting of only our side effect function $f(\phi)$ produces a more conservative optimal treatment policy that saves treatment modality M_1 until the last treatment period. On the other hand, in the third row of (c) we see that adding an intermediate reward consisting only of our tumor progression function $g(\tau)$ produces an optimal treatment policy that is much more aggressive, with no ties between the M_1 and M_2 modalities.

State transition probabilities

We utilize stationary transition probabilities that depend only upon the changes between states rather than on the actual value of the state. We also assume that state variables can only change by one increment between two successive treatment periods for simplicity.

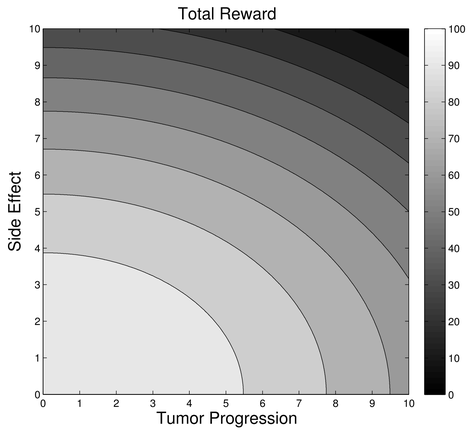
Modality	Side Effect			Tumor Progression		
	$P(\phi_{t+1} < \phi_t)$	$P(\phi_{t+1} = \phi_t)$	$P(\phi_{t+1} > \phi_t)$	$P(\tau_{t+1} < \tau_t)$	$P(\tau_{t+1} = \tau_t)$	$P(\tau_{t+1} > \tau_t)$
M_1	0	0.4	0.6	0.7	0.3	0
M_2	0	0.6	0.4	0.6	0.4	0
M_3	0.6	0.4	0	0	0.3	0.7

Reward functions

We utilize concave reward functions, which reward improvements made in worse states higher than improvements made in better states.

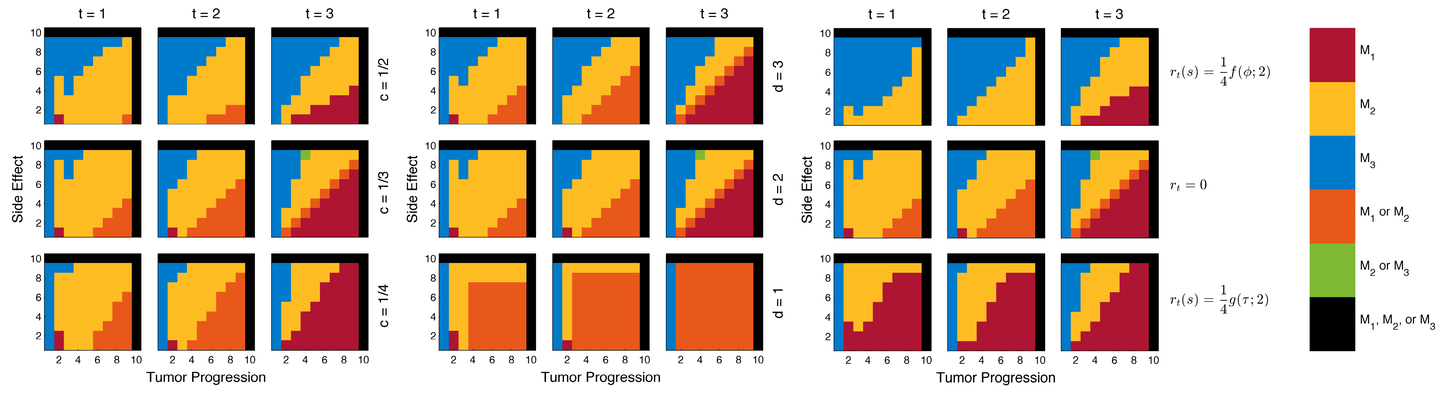
Side effect: $f(\phi; d) = \frac{100}{m^d}\left(m^d - \phi^d\right)$

Tumor progression: $g(\tau; d) = \frac{100}{n^d}\left(n^d - \tau^d\right)$



$r(s) = \frac{1}{3}f(\phi; 2) + \frac{2}{3}g(\tau; 2)$

Optimal treatment policies



(a) Change in relative importance of ϕ and τ
$r_{T+1}(s) = cf(\phi; 2) + (1-c)g(\tau; 2)$

(b) Change in the terminal reward shape
$r_{T+1}(s) = \frac{1}{3}f(\phi; d) + \frac{2}{3}g(\tau; d)$

(c) Change in intermediate reward function
$r_{T+1}(s) = \frac{1}{3}f(\phi; 2) + \frac{2}{3}g(\tau; 2)$

5. CONCLUSION

With diverse patient characteristics and numerous possible outcomes, making treatment decisions is extremely complex, and it may no longer be practical to make decisions based solely on individual experiences and empirical intuition. We proposed a novel mathematical framework to model optimal treatment policies for cancer therapy using a finite-horizon Markov decision process. Numerical simulations using simplified patient states and clinically intuitive reward functions have shown the potential application of our model to aid in treatment decision-making. Further study using clinical treatment-outcome data with custom utility functions will bring the problem to a higher dimensional state and outcome space, subject to "the curse of dimensionality". These high-dimensional problems will require an approximate dynamic programming approach which we leave for future work.

ACKNOWLEDGEMENTS

This work is supported in part by NSF grant CMMI 1560476 and an ARCS Foundation Fellowship.