Additional details for the ALP method applied to controlling wildfires

Ravi N. Haksar

This document provides additional details for the approximate linear program (ALP) approach for controlling a GMDP model of wildfires presented in [1]. The approach requires solving the following ALP,

$$
\begin{aligned}
\min_{w_i \in \mathbb{R}^{k_i}, \phi_i \in \mathbb{R}} \quad & \phi_i \\
\text{subject to} \quad & \phi_i \geq w_i^T h_i(x^t_{O(i)}) - g_i(x^t_{\Gamma(O(i))}, \overline{a}^t_{\Gamma(O(i))}), \\
& \quad \forall x^t_{\Gamma(O(i))} \in \mathcal{X}_{\Gamma(O(i))}. \\
& \phi_i \geq -w_i^T h_i(x^t_{O(i)}) + g_i(x^t_{\Gamma(O(i))}, a^t_{\Gamma(O(i))}), \\
& \quad \forall x^t_{\Gamma(O(i))} \in \mathcal{X}_{\Gamma(O(i))}, a^t_{\Gamma(O(i))} \in \mathcal{A}_{\Gamma(O(i))}.
\end{aligned}
\tag{1}
$$

where g_i is the expected future reward and discounted value,

$$
g_i(x^t_{\Gamma(O(i))}, a^t_{\Gamma(O(i))}) = \mathbb{E}_p \left[r_i(x^t_{\Gamma(O(i))}, a^t_{\Gamma(O(i))}, x^{t+1}_{O(i)}) + \gamma w_i^T h_i(x^{t+1}_{O(i)}) \right].
$$

For the wildfire model, Table 1 describes the dynamics of each tree, where $f_i^t = \sum_{j \in N(i)} \mathbf{1}_F(x_j^t)$ describes the number of neighboring trees on fire. Each tree is in one of three states, $x_i^t = \{\text{healthy}, \text{on fire}, \text{burnt}\} = \{H, F, B\}$. The indicator function is denoted by $\mathbf{1}_A(z)$ and is equal to one when $z = A$ and is zero otherwise. We also use the shorthand notation $h_i^t = \sum_{j \in N(i)} \mathbf{1}_H(x_j^t)$ to describe the number of healthy neighbor trees.

To simplify the problem, we assume each tree i has the local structure shown in Figure 1. This assumption leads to a single equivalence class in the GMDP. The derived control law is then applied to the original problem, with no additional structure assumptions.

Table 1: Single tree dynamics for wildfire model

x_i^t \ x_i^{t+1}	H	F	B
H	$1-\alpha f_i^t$	αf_i^t	0
F	0	$\beta - \Delta\beta a_i^t$	$1-\beta+\Delta\beta a_i^t$
B	0	0	1

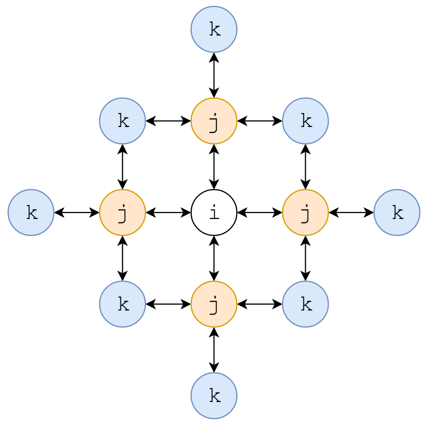

Figure 1: The 1-hop and 2-hop neighbors of tree i. The circles denoted j indicate the 1-hop neighbors, $\{x_j \mid j \in N(i)\}$. The circles denoted k indicate the 2-hop neighbors (also equivalently the 1-hop neighbors of j), $\{x_k \mid k \in N(j) \forall j \in N(i)\}$.

The reward function associated with each tree is,

$$\begin{aligned} r_i(x_{\Gamma(i)}^t) &= \mathbf{1}_H(x_i^t) - \mathbf{1}_F(x_i^t) \sum_{j \in N(i)} \mathbf{1}_H(x_j^t) \\ &= \mathbf{1}_H(x_i^t) - \mathbf{1}_F(x_i^t) h_i^t. \end{aligned} \tag{2}$$

The value function approximation for each tree is,

$$\begin{aligned} w_i^T h_i(x_{\Gamma(i)}^t) &= w_0 + w_1 \mathbf{1}_H(x_i^t) + w_2 \mathbf{1}_F(x_i^t) \sum_{j \in N(i)} \mathbf{1}_H(x_j^t) \\ &= w_0 + w_1 \mathbf{1}_H(x_i^t) + w_2 \mathbf{1}_F(x_i^t) h_i^t. \end{aligned}$$

To implement Program (1), the following expectation must be computed,

$$\mathbb{E}_p\left[w_i^T h_i(x_{\Gamma(i)}^{t+1})\right] = \mathbb{E}_p\left[w_0 + w_1\mathbf{1}_H(x_i^{t+1}) + w_2\mathbf{1}_F(x_i^{t+1})\sum_{j\in N(i)}\mathbf{1}_H(x_j^{t+1})\right].$$

Note that since the reward does not rely on information at the next time step, it does not need to be included in the expectation. Using the structure of the tree dynamics in Table 1, this expectation is,

$$\begin{aligned}\mathbb{E}_p\left[w_i^T h_i(x_{\Gamma(i)}^{t+1})\right] &= w_0 + w_1 p(x_i^{t+1} = H \mid x_i^t, x_{N(i)}^t) \\ &\quad + w_2(p(x_i^{t+1} = F \mid x_i^t, x_{N(i)}^t) + p(x_i^{t+1} = F \mid x_i^t, a_i^t))\times \\ &\quad \sum_{j\in N(i)} p(x_j^{t+1} = H \mid x_j^t, x_{N(j)}^t).\end{aligned} \tag{3}$$

The resulting approximate linear program is therefore,

$$\begin{aligned}\min_{w\in\mathbb{R}^3,\phi_i\in\mathbb{R}} \quad & \phi_i \\ \text{subject to} \quad & \phi_i \geq w_0 + w_1\mathbf{1}_H(x_i^t) + w_2\mathbf{1}_F(x_i^t)h_i^t - \mathbf{1}_H(x_i^t) - \mathbf{1}_F(x_i^t)h_i^t \\ & \quad - \gamma\,(\text{Equation (3)})\,, a_i^t = 0, \forall\ x_i^t, x_j^t, x_k^t \\ & \phi_i \geq -w_0 - w_1\mathbf{1}_H(x_i^t) - w_2\mathbf{1}_F(x_i^t)h_i^t + \mathbf{1}_H(x_i^t) + \mathbf{1}_F(x_i^t)h_i^t \\ & \quad + \gamma\,(\text{Equation (3)})\ \forall\ x_i^t, x_j^t, x_k^t, a_i^t\end{aligned}$$

As a result of the local structure (Figure 1), the above program requires on the order of 10^6 constraints. This is because it is necessary to consider all combinations of 1-hop and 2-hop neighbors of tree i. However, Anonymous Influence can significantly reduce the number of required constraints.

The constraints require the number of healthy neighbors, $h_i^t \in [0, 4]$, and the number of on fire neighbors, $f_i^t \in [0, 4]$. In addition, each healthy neighbor requires its number of on fire neighbors, $f_j^t \in [0, 4]$. Exploiting this property reduces the number of constraints to on the order of 10^3.

After solving the linear program, the control policy is generated by computing the coefficient of the tree action a_i^t,

$$\begin{aligned}
&\mathbb{E}_p\left[r_i + \gamma w_i^T h_i(x_{\Gamma(i)}^{t+1})\right] \\
&= \mathbf{1}_H(x_i^t) - \mathbf{1}_F(x_i^t)h_i^t + \gamma w_0 + \gamma w_1 \mathbf{1}_H(x_i^t)(1-\alpha f_i^t) \\
&\quad + \gamma w_2(\mathbf{1}_H(x_i^t)\alpha f_i^t + \mathbf{1}_F(x_i^t)(\beta - \Delta\beta a_i^t)) \sum_{j\in N(i)} \mathbf{1}_H(x_j^t)(1-\alpha f_j^t) \\
&= \mathbf{1}_H(x_i^t) - \mathbf{1}_F(x_i^t)h_i^t + \gamma w_0 + \gamma w_1 \mathbf{1}_H(x_i^t)(1-\alpha f_i^t) \\
&\quad + \gamma w_2(\mathbf{1}_H(x_i^t)\alpha f_i^t + \mathbf{1}_F(x_i^t)\beta) \sum_{j\in N(i)} \mathbf{1}_H(x_j^t)(1-\alpha f_j^t) \\
&\quad - a_i^t \gamma w_2 \mathbf{1}_F(x_i^t)\Delta\beta \sum_{j\in N(i)} \mathbf{1}_H(x_j^t)(1-\alpha f_j^t).
\end{aligned}$$

The above expression shows that each action is weighted by the quantity $\gamma w_2 \mathbf{1}_F(x_i^t)\Delta\beta \sum_{j\in N(i)} \mathbf{1}_H(x_j^t)(1-\alpha f_j^t)$. Therefore, for each time step of a simulation, the weight is calculated for each tree on fire. Then, $a_i^t = 1$ for the top C weights and $a_i^t = 0$ otherwise; C indicates the control effort capacity. This policy is an approximation of the capacity constrained linear program formulation presented in [1].

An ALP approach from literature [2] is also implemented. The following basis functions are used,

$$w_i^T h_i^{\text{prior}}(x_i^t) = w_0 \mathbf{1}_H(x_i^t) + w_1 \mathbf{1}_F(x_i^t) + w_2 \mathbf{1}_B(x_i^t),$$

with the same reward function as before. The required expectation is then,

$$\begin{aligned}
\mathbb{E}_{p_i}\left[w_i^T h_i^{\text{prior}}(x_i^{t+1})\right] &= \mathbb{E}_{p_i}\left[w_0 \mathbf{1}_H(x_i^{t+1}) + w_1 \mathbf{1}_F(x_i^{t+1}) + w_2 \mathbf{1}_B(x_i^{t+1})\right] \\
&= w_0 p(x_i^{t+1} = H \mid x_i^t, x_{N(i)}^t) \\
&\quad + w_1\left(p(x_i^{t+1} = F \mid x_i^t, x_{N(i)}^t) + p(x_i^{t+1} = F \mid x_i^t, a_i^t)\right) \\
&\quad + w_2\left(p(x_i^{t+1} = B \mid x_i^t, a_i^t) + p(x_i^{t+1} = B \mid x_i^t)\right).
\end{aligned} \tag{4}$$

The ALP from [2] is thus,

$$\begin{aligned}
\min_{w\in\mathbb{R}^3, \phi_i\in\mathbb{R}} \quad & \phi_i \\
\text{subject to} \quad & \phi_i \geq w_0 \mathbf{1}_H(x_i^t) + w_1 \mathbf{1}_F(x_i^t) + w_2 \mathbf{1}_B(x_i^t) - \mathbf{1}_H(x_i^t) - \mathbf{1}_F(x_i^t)h_i^t \\
& \quad - \gamma\,(\text{Equation (4)})\ \forall\ x_i^t, x_j^t, a_i^t \\
& \phi_i \geq -w_0 \mathbf{1}_H(x_i^t) - w_1 \mathbf{1}_F(x_i^t) - w_2 \mathbf{1}_B(x_i^t) + \mathbf{1}_H(x_i^t) + \mathbf{1}_F(x_i^t)h_i^t \\
& \quad + \gamma\,(\text{Equation (4)})\ \forall\ x_i^t, x_j^t, a_i^t
\end{aligned}$$

Anonymous Influence is again used to reduce the number of required constraints; on the order of 10^2 constraints are required. The resulting control policy is found by expanding the expected reward and future value,

$$\begin{aligned}
\mathbb{E}_{p_i}\left[r_i + \gamma w_i^T h_i^{\text{prior}}(x_i^{t+1})\right] &= \mathbf{1}_H(x_i^t) - \mathbf{1}_F(x_i^t)h_i^t + \gamma w_0 \mathbf{1}_H(x_i^t)(1-\alpha f_i^t) \\
&\quad + \gamma w_2(\mathbf{1}_H(x_i^t)\alpha f_i^t + \mathbf{1}_F(x_i^t)(\beta - \Delta\beta a_i^t)) \\
&\quad + \gamma w_3(\mathbf{1}_F(x_i^t)(1-\beta+\Delta\beta a_i^t) + \mathbf{1}_B(x_i^t)) \\
&= \mathbf{1}_H(x_i^t) - \mathbf{1}_F(x_i^t)h_i^t + \gamma w_0 \mathbf{1}_H(x_i^t)(1-\alpha f_i^t) \\
&\quad + \gamma w_2(\mathbf{1}_H(x_i^t)\alpha f_i^t + \mathbf{1}_F(x_i^t)\beta) \\
&\quad + \gamma w_3(\mathbf{1}_F(x_i^t)(1-\beta) + \mathbf{1}_B(x_i^t)) \\
&\quad + a_i^t \gamma \Delta\beta \mathbf{1}_F(x_i^t)(w_3 - w_2).
\end{aligned}$$

Each tree action is weighted by the quantity $\gamma\Delta\beta\mathbf{1}_F(x_i^t)(w_3 - w_2)$. Since this weight is constant once the weights w are determined, the control policy is equivalent to randomly choosing C trees on fire to apply action.

References

[1] R. N. Haksar and M. Schwager, "Controlling large, graph-based mdps with global control capacity constraints: An approximate lp solution," in *2018 IEEE Conference on Decision and Control (CDC)*, Dec 2018, pp. 35–42.

[2] N. Forsell and R. Sabbadin, "Approximate linear-programming algorithms for graph-based markov decision processes," in *Proceedings of the 2006 conference on ECAI 2006: 17th European Conference on Artificial Intelligence*. IOS Press, 2006, pp. 590–594.