Reward Function Definition
rl.Reward.DeltaReward
(revised)

UAV Design Project

October 12, 2022

1 Definitions

Let S_t be a state representation defining a UAV design and $S(S_t)$ define a simulator function that operates on this representation to produce the vector-string tuple as output:

$$S(s) = (\mathbf{x}, r)$$

where $\mathbf{x} = [x_r, x_c, x_v]^T$ is a vector defining the UAV's *metrics* range, cost and velocity (respectively). Let the metrics be clipped to within the ranges:

$$x_r \in [x_r^{min}, x_r^{max}]$$

$$x_c \in [x_c^{min}, x_c^{max}]$$

$$x_v \in [x_v^{min}, x_v^{max}]$$

.

and r is a categorical variable defining if the drone is stable (r = 'Success'), or unstable ($r \in ['HitBoundary','CouldNo$

Let the min-max normalized metric vector $\hat{\mathbf{x}} = [\hat{x_r}, \hat{x_c}, \hat{x_v}]$ be defined as:

$$\hat{\mathbf{x}} = \left[\left(\frac{x_r - x_r^{min}}{x_r^{max} x_r^{min}} \right), \quad \left(\frac{x_c - x_c^{min}}{x_c^{max} - x_c^{min}} \right), \quad \left(\frac{x_v - x_v^{min}}{x_v^{max} - x_v^{min}} \right) \quad \right]^T$$

For convenience, let the indicator function $\mathbb{1}_{stable}(S_t)$ be 1 (true) if the UAV state S_t is stable and 0 (false) otherwise:

$$\mathbb{1}_{stable}(S_t) = \begin{cases} 1 \text{ iff r='Success'} \\ 0 \quad \text{otherwise} \end{cases}$$

Then let $\mathbf{w} = [w_r, w_c, w_v]^T$ s.t $w_r + w_c + w_v = 1$ be a vector defining the weights for each simulator metric element, used for assigning importance to each metric, and let $\mathbf{t} = [t_r, t_c, t_v]$ be a set of threshold scalars for range, cost and velocity (respectively) and are defined by one of the objective definitions in [1]. Note that $\mathbf{t}$ will have zero elements for objectives where certain thresholds are not defined.

2 Reward function

The Delta Reward function for a state-action pair $\mathcal{R}(S_t, A_t)$ is defined as the difference between the 'state-value' functions for the current state and the next state:

$$R(S_t, A_t) = v(S_{t+1}) - v(S_t)$$

where the current state is S_t and the next state S_{t+1} is produced by taking action A_t with $p(S_{t+1}|S_t, A_t) = 1$). Here, $v(S)$ is defined as the scalar produced by the sum of the elementwise product between a weighted 'drone quality' -function and an 'objective penalty' -function if and only if the drone is stable:

$$v(S_t) = \mathbb{1}_{stable}(s) * (\mathbf{1} \cdot [p(S_t) \odot q(S_t)]) \,, \text{ where } v(S_t) \in [0, 1]$$

Here the function q is defined as:

$$q(S_t) = \mathbf{w} \odot \hat{x}$$

i.e the drone's quality is its normalized metrics (normalized complement in the case of cost) weighted.

The objective penalty function p is defined as a sigmoid function centered at the given metrics objective threshold value:

$$p(S_t) = \begin{bmatrix} p_r & p_c & p_v \end{bmatrix}^T$$

where the components are defined as:

$$p_r = \begin{cases} \frac{1}{1+exp(\gamma(t_r - \hat{x_r}))}, & \text{iff } t_r \neq 0 \\ 1 & \text{otherwise} \end{cases}$$

$$p_c = \begin{cases} \frac{1}{1+exp(\gamma(\hat{x_c} - t_c))}, & \text{iff } t_c \neq 0 \\ 1 & \text{otherwise} \end{cases}$$

$$p_r = \begin{cases} \frac{1}{1+exp(\gamma(t_v - \hat{x_v}))}, & \text{iff } t_v \neq 0 \\ 1 & \text{otherwise} \end{cases}$$

where γ is a "penalty strictness" parameter defining the sharpness of the sigmoid function at the threshold. Note that the penalty component is one when a threshold is not defined by the objective (e.g $t_r = 0$), meaning that an unpenalized UAV quality score for that metric is considered. Note also that state S_t has vector $\hat{x}$ associated with it, as calculated by $S(S_t)$.

3 Visualizations

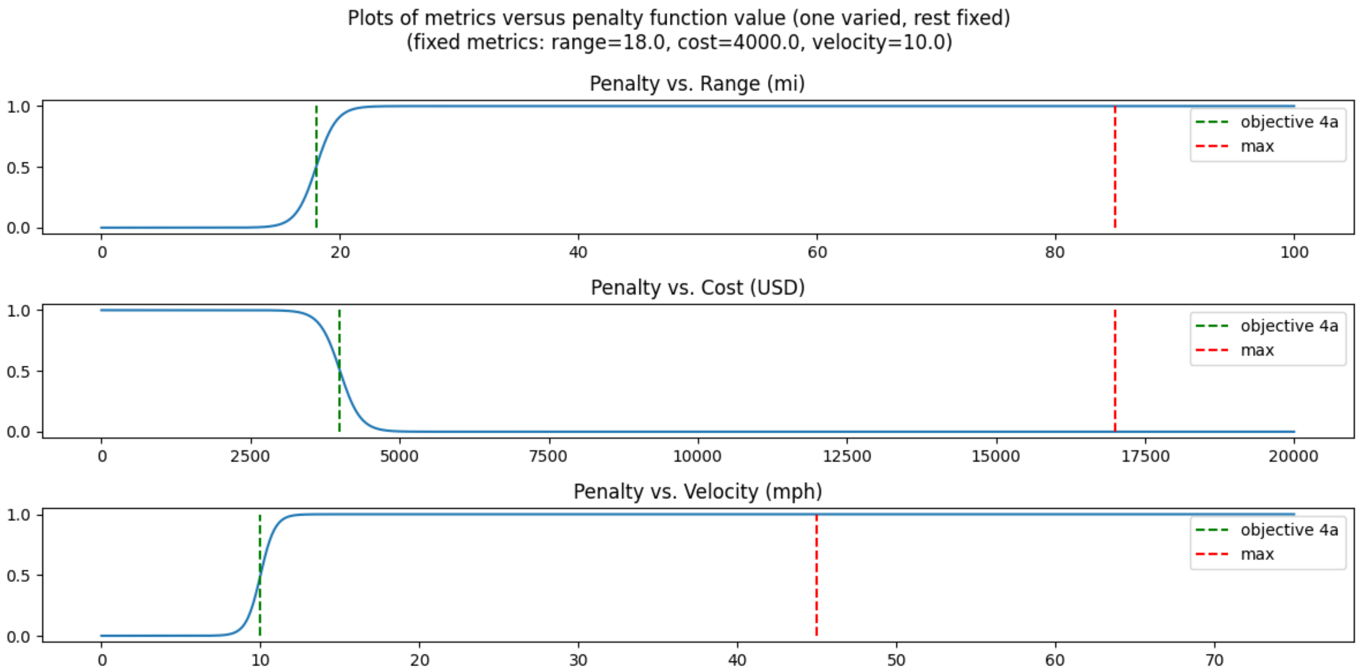

Figure 1:

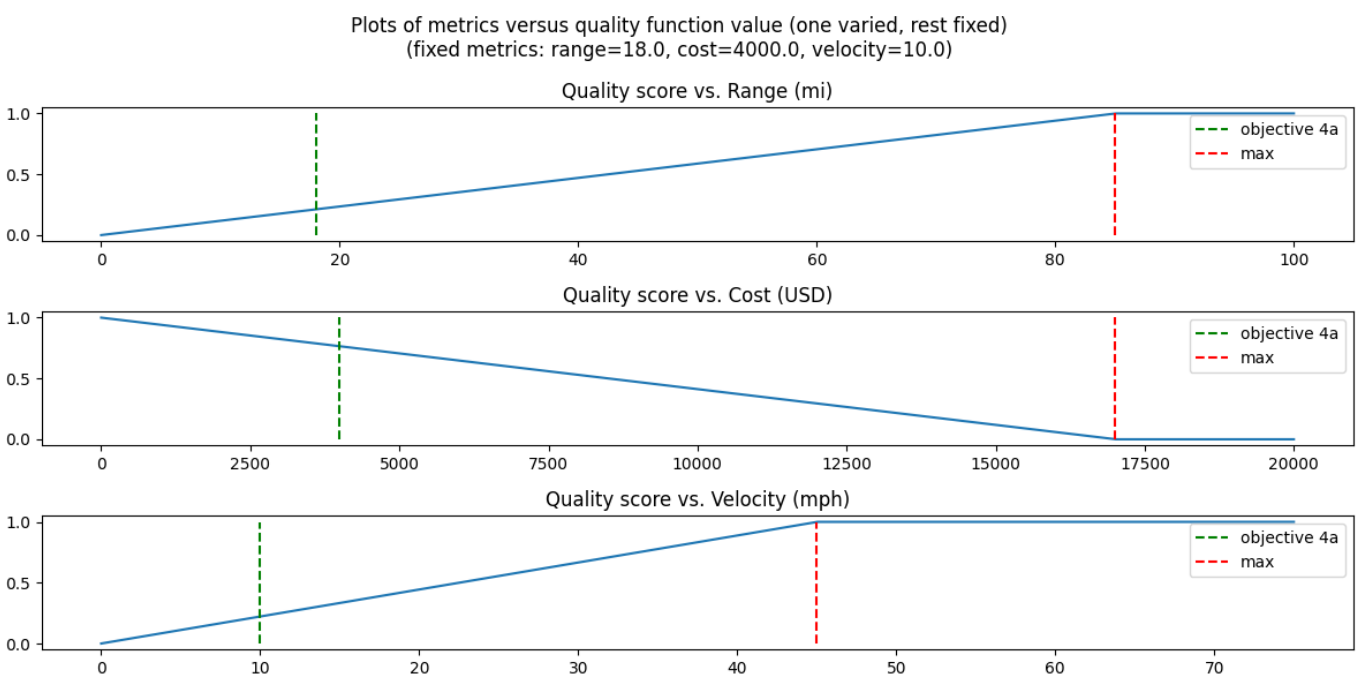

Figure 2:

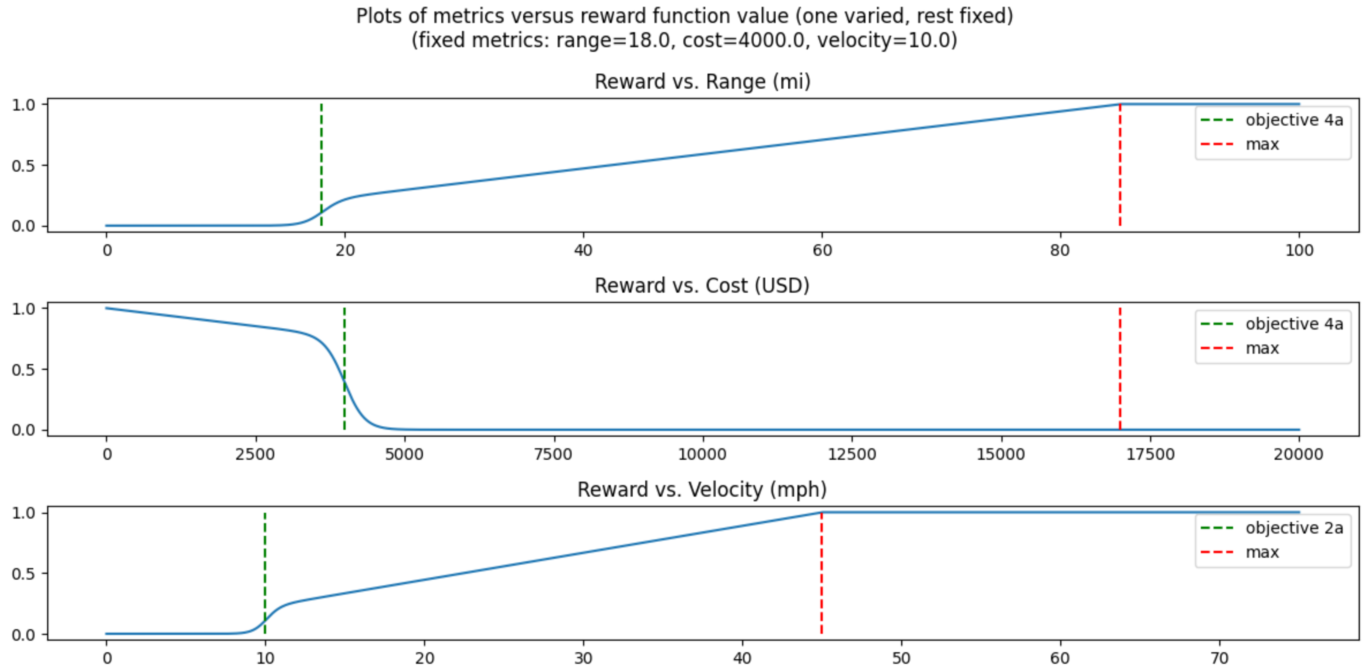

Figure 3:

References

[1] B. Song, N. F. Soria Zurita, H. Nolte, H. Singh, J. Cagan, and C. McComb, "When Faced With Increasing Complexity: The Effectiveness of Artificial Intelligence Assistance for Drone Design," *Journal of Mechanical Design*, vol. 144, 09 2021. 021701.