ADVANCED OPTIMISATION TOPICS

You have seen several heuristic search algorithms in the AI class to solve unconstrained combinatorial problems. You have seen that A* using an admissible heuristic is guaranteed to find the optimal solution to the problem.

Recall that A* works by minimising the cost of reaching the goal and an admissible heuristic for A* is a function $h(s)$ such that $h(s) \leq h^*(s)$ for all states s of the problem where $h^*(s)$ represents the optimal cost to reach the goal from s. In other words, an admissible heuristic $h(s)$ is a lower bound on the optimal solution for all s.

In the multi-objective version of A*, a multi-objective cost function $\vec{c}(x) = [c_1(x), c_2(x), \ldots, c_m(x)]$ is used for $m \geq 2$ and $h^*(s)$ now represents the **Pareto Front** for the state s.

Explain how to generalize the concept of an admissible heuristic for the multi-objective case and describe the property an admissible multi-objective heuristic should satisfy.

Furthermore, use your definition to provide an example of an admissible heuristic value for the Pareto front $\left\{ \begin{pmatrix} 162 \\ 7.2 \\ 34 \end{pmatrix}, \begin{pmatrix} 149 \\ 7.0 \\ 38 \end{pmatrix}, \begin{pmatrix} 140 \\ 7.5 \\ 45 \end{pmatrix}, \begin{pmatrix} 152 \\ 8.2 \\ 29 \end{pmatrix} \right\}$

Multi-Objective Optimisation

The formal definition of an admissible heuristic for multi-objective A* is:

A multi-objective heuristic h is a set of vectors and it is admissible if and only if, for all states s, for all vector $\vec{u}^*$ in the Pareto Front of s there exists a vector $\vec{v} \in h(s)$ such that $\vec{v} \preceq \vec{u}^*$.

The key differences from the single objective case are:

- replace $\leq$ by $\preceq$
- make sure that every value in the Pareto Front has at least 1 vector in the heuristic that is "better than or equal to" it, i.e., $\vec{v} \preceq \vec{u}^*$.

Also, keep in mind that the best admissible heuristic possible is the optimal solution, i.e., a lower bound that is equal to the function being bounded. For multi-objective, this is the same thing, i.e., the Pareto-Front is the best possible admissible heuristic. The ideal-point or using an individual admissible for each direction is an **example** of admissible multi-objective heuristic and it can be greatly improved in practice.

<u>Stochastic Programming</u>

In this question, let's consider that our random variable ϵ represents all the uncertainty in our stochastic problem and is "discrete", meaning $P(\epsilon = i) > 0$ for $i \in \{o_1, \ldots, o_k\}$. This is exactly the case we discussed in the lecture.

An important property of the expected value operator is its "<u>linearity</u>": given a random variable ϵ, $E[a\epsilon + b] = aE[\epsilon] + b$ for $a \in \mathbb{R}$ and $b \in \mathbb{R}$. It is easy to see this when you expand the expected value operator, i.e.,

$E[a\epsilon + b] = \sum_i P(\epsilon = i)(a \times i + b) = a(\sum_i P(\epsilon = i)) + b = aE[\epsilon] + b$

In class, we learned that the optimal solution of the recourse-problem is less than or equal to the optimal solution of the expected-value problem. Formally, as we saw in class, let $g(x, \hat{\epsilon})$ represent the stochastic problem for a fixed of the value $\hat{\epsilon} \in \{o_1, \ldots, o_k\}$ for the random variable ϵ. Moreover, for this question, assume that this stochastic problem takes the form of a <u>Linear Program</u>, that is, the objective function is a linear function of x and $\hat{\epsilon}$ and the feasible region Ω is defined by a set of linear constraints. Then we have $\min_{x \in \Omega} E_\epsilon[g(x, \epsilon)] \leq \min_{x \in \Omega} g(x, E_\epsilon[\epsilon])$.

Explain why the recourse-problem $\min_{x \in \Omega} E_\epsilon[g(x, \epsilon)]$ can potentially be lower than $\min_{x \in \Omega} g(x, E_\epsilon[\epsilon])$ for a <u>Linear Program</u>, despite the <u>linearity</u> of the expected value operator.

Using the same notation as the L-Shape Method slides, we have that $\min_{x \in \Omega} g(x, E_\epsilon[\epsilon])$ is:

$$\min_{x,z} f^\top x + (\sum_i P(\epsilon = i)c_i)^\top z$$
$$\text{s.t. } (\sum_i P(\epsilon = i)B_i)x + Dz = (\sum_i P(t = i)d_i)$$

where all the occurrences of "$(\sum_i P(\epsilon = i) \cdots)$" are computing the expectation regarding to the uncertainty. In other words, they will become the average vector or matrix depending on what the target of the expectation is.

Linearity of expectation still holds, e.g., we can move z in the objective function inside the summation computing the expectation and value is still the same: $(\sum_i P(\epsilon = i)c_i)^\top z = \sum_j (\sum_i P(\epsilon = i)c_i)_j * z_j = \sum_i (\sum_j P(\epsilon = i)c_{ij} * z_j)$. Using this, we can create an specific value of z (the recourse action) for each possible value of the uncertainty, i.e., change z to z_i. But this alone is not enough. We need to specialize the constraints, i.e., change $(\sum_i P(\epsilon = i)B_i)x + Dz = (\sum_i P(t = i)d_i)$ to $B_i x + Dz_i = d_i \quad \forall i$ and this change in the feasible space is what allows us to obtain a potential smaller objective function.