0.1 Set Up

When the reward/cost function has state dependent values, the objective of the optimiztaion problem becomes quadratic. Suppose each road has state dependent waiting time $r_t(s) = c(s)\sum_a y_t(s,a) + d(s)$. Then the MDP problem can be written as:

$$
\begin{aligned}
\min_{y_t(s,a)} \quad & \sum_0^T \sum_s \sum_a \frac{1}{2} c(s) y_t^2(s,a) + d(s) y_t(s,a) \\
\text{s.t.} \quad & \sum_a y_t(s',a) = \sum_s \sum_a p(s'|s,a) y_{t-1}(s,a) \quad (a) \\
& \sum_a y_0(s,a) = p_0(s) \quad (b) \\
& y_t(s,a) \geq 0 \quad (c)
\end{aligned} \tag{1}
$$

Taking dual variables $V_t(s)$ for constraint a, $V_0(s)$ for constraint b, and $\mu_t(s,a)$ for constraint c, the Lagrangian is:

$$
\begin{aligned}
L(y_t(s,a), V_t(s), \mu_t(s,a)) = & \sum_0^T \sum_s \sum_a \frac{1}{2} c(s) y_t^2(s,a) + d(s) y_t(s,a) \\
& + \sum_1^T \sum_{s'} V_t(s') \Big(\sum_s \sum_a p(s'|s,a) y_{t-1}(s,a) - \sum_a y_t(s',a) \Big) \\
& + \sum_s V_0(s) \Big(p_0(s) - \sum_a y_0(s,a) \Big) \\
& + \sum_0^T \sum_s \sum_a \mu_t(s,a) y_t(s,a)
\end{aligned} \tag{2}
$$

The KKT conditions are:

$$
\begin{aligned}
t = T: \quad & c(s) y_T(s,a) + d(s) - V_T(s) + \mu_T(s,a) && = 0 \\
t = T-1 \cdots 0: \quad & c(s) y_t(s,a) + d(s) - V_t(s) + \sum_{s'} V_{t+1} p(s'|s,a) + \mu_t(s,a) && = 0
\end{aligned} \tag{3}
$$

This gives the following dual problem:

$$
\begin{aligned}
\max_{y_t(s,a), V_t(s)} \quad & \sum_s V_0(s) p_0(s) - \sum_0^T \sum_a \sum_s \frac{1}{2} c(s) y_t^2(s,a) \\
\text{s.t.} \quad & V_T(s) = c(s) y_T(s,a) + d(s) + \mu_T(s,a) \\
& V_t(s) = c(s) y_t(s,a) + d(s) + \mu_t(s,a) - \sum_{s'} V_{t+1} p(s'|s,a)
\end{aligned} \tag{4}
$$

This is very similar to the dynamic programming problem in the linear case, except for the second term in the objective. Suppose now $W_T(s,a) = \frac{1}{2} c(s) y_T^2(s,a)$, and $W_t(s,a) =$

$\frac{1}{2}c(s)y_t^2(s,a) + \sum_{s'} W_{t+1}(s',a)p(s'|s,a)$. The dual becomes:

$$\begin{aligned}
\max_{y_t(s,a),V_t(s)} \quad & \sum_s V_0(s)p_0(s) - \sum_a \sum_s W_0(s,a) \\
\text{s.t.} \quad & V_T(s) = c(s)y_T(s,a) + d(s) + \mu_T(s,a) \\
& V_t(s) = c(s)y_t(s,a) + d(s) + \mu_t(s,a) - \sum_{s'} V_{t+1}(s')p(s'|s,a) \\
& W_T(s) = \frac{1}{2}c(s)y_T^2(s,a) \\
& W_t(s,a) = \frac{1}{2}c(s)y_t^2(s,a) + \sum_{s'} W_{t+1}(s',a)p(s'|s,a)
\end{aligned} \tag{5}$$