Electronic companion for "Decomposable Markov decision processes: a fluid optimization approach" by D. Bertsimas and V. V. Mišić

EC.1. Proofs

EC.1.1. Proof of Proposition 1

This result follows directly from observations in Romeijn et al. (1992). To see this, observe that problem (2) can be expressed as a type of doubly infinite linear optimization problem (problem (P) in Romeijn et al. 1992) where each variable appears in finitely many constraints and each constraint involves only finitely many variables. Note that problem (P) in Romeijn et al. (1992) contains explicit upper bounds on all of the variables, which problem (2) does not appear to have; however, note that the variables in problem (2), by virtue of the constraints, are upper bounded by 1. In this way, problem (2) can be cast as problem (P) from Romeijn et al. (1992). Assumption A of Romeijn et al. (1992) holds (that the feasible region of problem (2) is nonempty; this will be established in the proof of Proposition 2, where we will directly construct a feasible solution to problem (2)). Assumption B of Romeijn et al. (1992) also holds (that the objective function is uniformly convergent on the feasible region; this holds due to the discounting in the objective function). With these two assumptions and following the argument in Section 2 of Romeijn et al. (1992), it follows that there exists an optimal solution to problem (2). $\square$

EC.1.2. Proof of Proposition 2

To prove this, we will construct a feasible solution to problem (2) whose objective value in (2) is the same as $J^*(\mathbf{s})$. For each $t \in \{1, 2, \dots\}$, set

$$\begin{aligned} x^m_{ka}(t) &= \mathbb{P}(s^m(t) = k,\ \pi^*(\mathbf{s}(t)) = a) \\ A_a(t) &= \mathbb{P}(\pi^*(\mathbf{s}(t)) = a), \end{aligned}$$

where $\{\mathbf{s}(t)\}_{t=1}^{\infty}$ is the stochastic process of the system state that starts in state $\mathbf{s} = (s^1, \dots, s^M)$, operated according to the optimal policy π^* that yields the optimal value function $J^*(\cdot)$. We now verify that the solution $(\mathbf{x}, \mathbf{A})$ is indeed feasible.

For constraint (2b), we have for any $m \in \{1, \dots, M\}$, $j \in \mathcal{S}^m$ and $t \in \{2, 3, 4, \dots\}$,

$$\begin{aligned} \sum_{a \in \mathcal{A}} x^m_{ja}(t) &= \sum_{a \in \mathcal{A}} \mathbb{P}(s^m(t) = j,\ \pi^*(\mathbf{s}(t)) = a) \\ &= \mathbb{P}(s^m(t) = j) \end{aligned}$$

$$= \sum_{k \in \mathcal{S}^m} \sum_{\tilde{a} \in \mathcal{A}} \mathbb{P}(s^m(t) = j,\ s^m(t-1) = k,\ \pi^*(\mathbf{s}(t-1)) = \tilde{a})$$
$$= \sum_{k \in \mathcal{S}^m} \sum_{\tilde{a} \in \mathcal{A}} \mathbb{P}(s^m(t) = j \mid s^m(t-1) = k,\ \pi^*(\mathbf{s}(t-1)) = \tilde{a}) \cdot \mathbb{P}(s^m(t-1) = k,\ \pi^*(\mathbf{s}(t-1)) = \tilde{a})$$
$$= \sum_{k \in \mathcal{S}^m} \sum_{\tilde{a} \in \mathcal{A}} p^m_{kja} \cdot x^m_{ka}(t-1)$$

where the first equality follows by the definition of $x^m_{ka}(t)$, the second and third equalities follow by the countable additivity of probability and the law of total probability, the fourth by conditioning on the state and action at $t-1$ and the final step by using the definition of p^m_{kja} and $x^m_{ka}(t)$.

Similarly, for constraint (2c), for any $a \in \{1, \dots, M\}$ and $t \in \{1, 2, \dots\}$ we have

$$\sum_{k \in \mathcal{S}^m} x^m_{ka}(t) = \sum_{k \in \mathcal{S}^m} \mathbb{P}(s^m(t) = k,\ \pi^*(\mathbf{s}(t)) = a) = \mathbb{P}(\pi^*(\mathbf{s}(t)) = a) = A_a(t).$$

For constraint (2d), our earlier reasoning gives us that for any $m \in \{1, \dots, M\}$, $k \in \mathcal{S}^m$,

$$\sum_{a=1}^{M} x^m_{k,a} = \sum_{a \in \mathcal{A}} \mathbb{P}(s^m(1) = k,\ \pi^*(\mathbf{s}(1)) = a) = \mathbb{P}(s^m(1) = k) = \alpha^m_k(\mathbf{s}).$$

Lastly, since $\mathbf{x}$ and $\mathbf{A}$ are defined as probabilities, it follows that

$$x^m_{ka}(t) \geq 0, \quad \forall\ m \in \{1, \dots, M\}, a \in \mathcal{A},\ k \in \mathcal{S}^m,\ t \in \{1, 2, \dots\},$$
$$A_a(t) \geq 0, \quad \forall\ a \in \mathcal{A},\ t \in \{1, 2, \dots\}.$$

This establishes that $(\mathbf{x}, \mathbf{A})$ is a feasible solution to (2). We will now verify that its objective value is identical to $J^*(\mathbf{s})$. By the definition of the optimal value function, we have

$$J^*(\mathbf{s}) = \mathbb{E}\left[\sum_{t=1}^{\infty} \sum_{m=1}^{M} \beta^{t-1} \cdot g^m_{s^m(t), \pi^*(\mathbf{s}(t))}\right] = \sum_{t=1}^{\infty} \sum_{m=1}^{M} \beta^{t-1} \cdot \mathbb{E}\left[g^m_{s^m(t), \pi^*(\mathbf{s}(t))}\right],$$

where the second step follows by the monotone convergence theorem (since β and all the g^m_{ka} values are nonnegative) and the linearity of expectation. Observe now that

$$\mathbb{E}[g^m_{s^m(t), \pi^*(\mathbf{s}(t))}] = \sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} g^m_{ka} \cdot \mathbb{P}(s^m(t) = k,\ \pi^*(\mathbf{s}(t)) = a) = \sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} g^m_{ka} \cdot x^m_{ka}(t).$$

We therefore have

$$J^*(\mathbf{s}) = \sum_{t=1}^{\infty} \sum_{m=1}^{M} \beta^{t-1} \cdot \mathbb{E}\left[g^m_{s^m(t), \pi^*(\mathbf{s}(t))}\right] = \sum_{t=1}^{\infty} \sum_{m=1}^{M} \sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} \beta^{t-1} \cdot g^m_{ka} \cdot x^m_{ka},$$

in other words, the value function evaluated at the starting state $\mathbf{s}$ is exactly the objective value of the $(\mathbf{x}, \mathbf{A})$ solution that we constructed. Since $(\mathbf{x}, \mathbf{A})$ is a feasible solution for (2), it follows that $J^*(\mathbf{s}) \leq Z^*(\mathbf{s})$ which is the required result. $\square$

EC.1.3. Proof of Proposition 3

Using the same type of reasoning that we used in the proof of Proposition 2 to show that the objective function of the constructed solution was equal to $J^*(\mathbf{s})$, one can show that the objective value of the optimal fluid solution $(\mathbf{x}, \mathbf{A})$ is equal to $\mathbb{E}\left[\sum_{t=1}^{\infty}\sum_{m=1}^{M} \beta^{t-1} \cdot g^m_{s^m(t),\pi(t,\mathbf{s}(t))}\right]$, i.e., the expected discounted reward when the system is operated according to the policy π and the system starts in state $\mathbf{s}$. Since $J^*(\mathbf{s})$ is the maximum over all policies of the expected discounted reward of the system, it follows that $J^*(\mathbf{s})$ is an upper bound on $\mathbb{E}\left[\sum_{t=1}^{\infty}\sum_{m=1}^{M} \beta^{t-1} \cdot g^m_{s^m(t),\pi(t,\mathbf{s}(t))}\right]$, and thus that $J^*(\mathbf{s}) \geq Z^*(\mathbf{s})$. □

EC.1.4. Proof of Theorem 1

Let $\mathbf{s} \in \mathcal{S}$, and let $(\mathbf{x}(\mathbf{s}), \mathbf{A}(\mathbf{s}))$ be the solution of the corresponding fluid problem. To prove the theorem, we will show that for state $\mathbf{s}$, any action a with $A_a(1,\mathbf{s}) > 0$ must be greedy with respect to the optimal value function J^*. By standard results in dynamic programming theory (see, e.g., Bertsekas 1995), this is sufficient to prove that π as we have defined it is an optimal policy.

By Propositions 2 and 3, we have that $Z^*(\mathbf{s}) = J^*(\mathbf{s})$. Since $(\mathbf{x}(\mathbf{s}), \mathbf{A}(\mathbf{s}))$ is achievable, let $\bar{\pi}$ be the (possibly nondeterministic, nonstationary) policy that achieves it and let $\{\mathbf{s}(t)\}_{t=1}^{\infty}$ be the stochastic process of the system state when operated according to $\bar{\pi}$. We then have that

$$\begin{aligned}
J^*(\mathbf{s}) &= Z^*(\mathbf{s}) \\
&= \mathbb{E}\left[\sum_{t=1}^{\infty}\sum_{m=1}^{M} \beta^{t-1} g^m_{s^m(t),\bar{\pi}(t,\mathbf{s}(t))}\right] \\
&= \sum_{m=1}^{M} \mathbb{E}[g^m_{s^m(1),\bar{\pi}(\mathbf{s}(1),1)}] + \mathbb{E}\left[\sum_{t=2}^{\infty}\sum_{m=1}^{M} \beta^{t-1} g^m_{s^m(t),\bar{\pi}(t,\mathbf{s}(t))}\right] \\
&= \sum_{m=1}^{M} \mathbb{E}[g^m_{s^m(1),\bar{\pi}(\mathbf{s}(1),1)}] + \sum_{\tilde{\mathbf{s}}\in\mathcal{S}} \mathbb{E}\left[\sum_{t=2}^{\infty}\sum_{m=1}^{M} \beta^{t-1} g^m_{s^m(t),\bar{\pi}(t,\mathbf{s}(t))} \,\middle|\, \mathbf{s}(2) = \tilde{\mathbf{s}}\right] \cdot \mathbb{P}(\mathbf{s}(2) = \tilde{\mathbf{s}}) \\
&= \sum_{m=1}^{M}\sum_{a\in\mathcal{A}} g^m_{s^m,a} A_a(1,\mathbf{s}) + \beta \cdot \sum_{\tilde{\mathbf{s}}\in\mathcal{S}} \mathbb{E}\left[\sum_{t=2}^{\infty}\sum_{m=1}^{M} \beta^{t-2} g^m_{s^m(t),\bar{\pi}(t,\mathbf{s}(t))} \,\middle|\, \mathbf{s}(2) = \tilde{\mathbf{s}}\right] \cdot \mathbb{P}(\mathbf{s}(2) = \tilde{\mathbf{s}})
\end{aligned}$$

where the second step follows since $(\mathbf{x}(\mathbf{s}), \mathbf{A}(\mathbf{s}))$ is achieved by $\bar{\pi}$; the third step by breaking up the infinite sum; the fourth step by conditioning on the second state; and the fifth step by using the fact that $A_a(1,\mathbf{s}) = \mathbb{P}(\bar{\pi}(\mathbf{s}(1),1) = a)$ and factoring out a β. Now observe that we must have

$$\begin{aligned}
&\sum_{m=1}^{M}\sum_{a\in\mathcal{A}} g^m_{s^m,a} A_a(1,\mathbf{s}) + \beta \cdot \sum_{\tilde{\mathbf{s}}\in\mathcal{S}} \mathbb{E}\left[\sum_{t=2}^{\infty}\sum_{m=1}^{M} \beta^{t-2} g^m_{s^m(t),\bar{\pi}(t,\mathbf{s}(t))} \,\middle|\, \mathbf{s}(2) = \tilde{\mathbf{s}}\right] \cdot \mathbb{P}(\mathbf{s}(2) = \tilde{\mathbf{s}}) \\
&= \sum_{m=1}^{M}\sum_{a\in\mathcal{A}} g^m_{s^m,a} A_a(1,\mathbf{s}) + \beta \cdot \sum_{\tilde{\mathbf{s}}\in\mathcal{S}} J^*(\tilde{\mathbf{s}}) \cdot \mathbb{P}(\mathbf{s}(2) = \tilde{\mathbf{s}}). \qquad \text{(EC.1)}
\end{aligned}$$

This follows because the left-hand side is less than or equal to the right-hand side, which is true because

$$\mathbb{E}\left[\sum_{t=2}^{\infty}\sum_{m=1}^{M}\beta^{t-2}g^m_{s^m(t),\bar{\pi}(t,\mathbf{s}(t))}\,\middle|\,\mathbf{s}(2)=\tilde{\mathbf{s}}\right]\leq J^*(\tilde{\mathbf{s}})$$

(here, the left-hand side is the value of running policy $\bar{\pi}$ from time 2 on at state $\tilde{\mathbf{s}}$, which is clearly at most $J^*(\tilde{\mathbf{s}})$). Note also that the left-hand side in equation (EC.1) cannot be *strictly* less than the right-hand side, because the right-hand side value is achieved by a policy (follow $\bar{\pi}$ at $t=1$, then follow any optimal policy from $t=2$ on); by our earlier manipulation, the left-hand side in equation (EC.1) is equal to $J^*(\mathbf{s})$, so a strict inequality would imply a policy that achieves a better value than the optimal value. This is not possible, so equation (EC.1) must hold.

Thus, so far, we have that

$$J^*(\mathbf{s})=\sum_{m=1}^{M}\sum_{a\in\mathcal{A}}g^m_{s^m,a}A_a(1,\mathbf{s})+\beta\cdot\sum_{\tilde{\mathbf{s}}\in\mathcal{S}}J^*(\tilde{\mathbf{s}})\cdot\mathbb{P}(\mathbf{s}(2)=\tilde{\mathbf{s}}). \tag{EC.2}$$

Since the components are independent, we can write $\mathbb{P}(\mathbf{s}(2)=\tilde{\mathbf{s}})$ as

$$\mathbb{P}(\mathbf{s}(2)=\tilde{\mathbf{s}})=\sum_{a\in\mathcal{A}}A_a(1,\mathbf{s})\cdot\prod_{m=1}^{M}p^m_{s^m\tilde{s}^m a}$$

and simplify equation (EC.2) as follows:

$$\begin{aligned}J^*(\mathbf{s})&=\sum_{m=1}^{M}\sum_{a\in\mathcal{A}}g^m_{s^m,a}A_a(1,\mathbf{s})+\beta\cdot\sum_{\tilde{\mathbf{s}}\in\mathcal{S}}J^*(\tilde{\mathbf{s}})\cdot\left(\sum_{a\in\mathcal{A}}A_a(1,\mathbf{s})\cdot\prod_{m=1}^{M}p^m_{s^m\tilde{s}^m a}\right)\\&=\sum_{a\in\mathcal{A}}A_a(1,\mathbf{s})\cdot\left(\sum_{m=1}^{M}g^m_{s^m(1),a}+\beta\cdot\sum_{\tilde{\mathbf{s}}\in\mathcal{S}}J^*(\tilde{\mathbf{s}})\cdot\prod_{m=1}^{M}p^m_{s^m\tilde{s}^m a}\right).\end{aligned} \tag{EC.3}$$

Now recall that $J^*(\mathbf{s})$ is the optimal value function, so it must satisfy the Bellman equation:

$$J^*(\mathbf{s})=\max_{a\in\mathcal{A}}\left(\sum_{m=1}^{M}g^m_{s^m,a}+\beta\cdot\sum_{\tilde{\mathbf{s}}\in\mathcal{S}}J^*(\tilde{\mathbf{s}})\cdot\prod_{m=1}^{M}p^m_{s^m\tilde{s}^m a}\right).$$

Given these two expressions, it must be that for any $\bar{a}$ with $A_{\bar{a}}(1,\mathbf{s})>0$, we have

$$\sum_{m=1}^{M}g^m_{s^m(1),\bar{a}}+\beta\cdot\sum_{\tilde{\mathbf{s}}\in\mathcal{S}}J^*(\tilde{\mathbf{s}})\cdot\prod_{m=1}^{M}p^m_{s^m\tilde{s}^m\bar{a}}=\max_{a\in\mathcal{A}}\left(\sum_{m=1}^{M}g^m_{s^m,a}+\beta\cdot\sum_{\tilde{\mathbf{s}}\in\mathcal{S}}J^*(\tilde{\mathbf{s}})\cdot\prod_{m=1}^{M}p^m_{s^m\tilde{s}^m a}\right);$$

i.e., that $\bar{a}$ is an action that is greedy with respect to the optimal value function J^*. This must be true; by the definition of the maximum, the right-hand side is clearly greater than or equal to the left (this is true for any $\bar{a}\in\mathcal{A}$, not only those with $A_{\bar{a}}(1,\mathbf{s})>0$). It cannot be strictly greater, because then equation (EC.3) could not hold (the right-hand side in (EC.3) would have to be strictly less than $J^*(\mathbf{s})$, as the $A_a(1,\mathbf{s})$ values are nonnegative and sum to one).

Finally, note that the action $a=\arg\max_{\bar{a}\in\mathcal{A}}A_{\bar{a}}(1,\mathbf{s})$ clearly satisfies $A_a(1,\mathbf{s})>0$, since the $A_a(t)$ variables are nonnegative and have unit sum. Thus, the action $a=\arg\max_{\bar{a}\in\mathcal{A}}A_{\bar{a}}(1,\mathbf{s})$ must be greedy with respect to the optimal value function. This establishes that the policy π, as defined in the statement of Theorem 1, must be an optimal policy, which is the required result. □

EC.1.5. Proof of Proposition 4

EC.1.5.1. Proof of Part (a) To prove (a), we will define solution to problem (3) that (i) is feasible for problem (3) and (ii) achieves an objective value in (3) that is equal to $J^*(\mathbf{s})$. Since $Z_T^*(\mathbf{s})$ is the optimal value of problem (3) and our constructed feasible solution achieves a value of $J^*(\mathbf{s})$, it will follow that $J^*(\mathbf{s}) \leq Z_T^*(\mathbf{s})$.

Let π^* be the optimal policy that solves problem (1) and let us consider the following solution for problem (3):

$$
\begin{aligned}
&x_{ka}^m(t) = \mathbb{P}(s^m(t) = k,\ \pi^*(\mathbf{s}(t)) = a), \quad \forall\, t \in \{1, \ldots, T\}, \\
&A_a(t) = \mathbb{P}(\pi^*(\mathbf{s}(t)) = a), \quad \forall\, t \in \{1, \ldots, T\}, \\
&x_{ka}^m(T+1) = \sum_{t=T+1}^{\infty} \beta^{t-(T+1)} \cdot \mathbb{P}(s^m(t) = k,\ \pi^*(\mathbf{s}(t)) = a), \\
&A_a(T+1) = \sum_{t=T+1}^{\infty} \beta^{t-(T+1)} \cdot \mathbb{P}(\pi^*(\mathbf{s}(t)) = a),
\end{aligned}
$$

where $\{\mathbf{s}(t)\}_{t=1}^{\infty}$ is the stochastic process of the complete system that begins in state $\mathbf{s}$ at $t = 1$ and is operated according to policy π^*.

We first need to show that the proposed $(\mathbf{x}, \mathbf{A})$ is feasible. The same steps used in the proof of Proposition 2 can be used to verify that constraints (3b), (3d) (for $t \in \{1, \ldots, T\}$) and (3e) hold. To verify that constraint (3c) is satisfied, we apply similar reasoning as for constraint (3b), but in a long-run discounted way:

$$
\begin{aligned}
\sum_{a \in \mathcal{A}} x_{ja}^m(T+1) &= \sum_{a \in \mathcal{A}} \sum_{t=T+1}^{\infty} \beta^{t-(T+1)} \cdot \mathbb{P}(s^m(t) = j,\ \pi^*(\mathbf{s}(t)) = a) \\
&= \sum_{t=T+1}^{\infty} \beta^{t-(T+1)} \sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} \mathbb{P}(s^m(t) = j \mid s^m(t-1) = k,\ \pi^*(\mathbf{s}(t)) = a) \\
&\qquad \cdot \mathbb{P}(s^m(t-1) = k,\ \pi^*(\mathbf{s}(t)) = a) \\
&= \sum_{t=T+1}^{\infty} \beta^{t-(T+1)} \sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} p_{kja}^m \cdot \mathbb{P}(s^m(t-1) = k,\ \pi^*(\mathbf{s}(t-1)) = a) \\
&= \sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} p_{kja}^m \cdot \mathbb{P}(s^m(T) = k,\ \pi^*(\mathbf{s}(T)) = a) \\
&\quad + \beta \cdot \sum_{t=T+2}^{\infty} \beta^{t-(T+2)} \sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} p_{kja}^m \cdot \mathbb{P}(s^m(t-1) = k,\ \pi^*(\mathbf{s}(t-1)) = a) \\
&= \sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} p_{kja}^m \cdot \mathbb{P}(s^m(T) = k,\ \pi^*(\mathbf{s}(T)) = a) \\
&\quad + \beta \cdot \sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} p_{kja}^m \left[\sum_{t=T+1}^{\infty} \beta^{t-(T+1)} \cdot \mathbb{P}(s^m(t) = k,\ \pi^*(\mathbf{s}(t)) = a) \right]
\end{aligned}
$$

$$= \sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} p^m_{kja} \cdot x^m_{ka}(T) + \beta \cdot \sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} p^m_{kja} x^m_{ka}(T+1).$$

To verify constraint (3d) for $t = T+1$, observe that

$$\begin{aligned}
\sum_{k \in \mathcal{S}^m} x^m_{ka}(T+1) &= \sum_{k \in \mathcal{S}^m} \sum_{t=T+1}^{\infty} \beta^{t-(T+1)} \cdot \mathbb{P}(s^m(t) = k,\ \pi^*(\mathbf{s}(t)) = a) \\
&= \sum_{t=T+1}^{\infty} \beta^{t-(T+1)} \sum_{k \in \mathcal{S}^m} \mathbb{P}(s^m(t) = k,\ \pi^*(\mathbf{s}(t)) = a) \\
&= \sum_{t=T+1}^{\infty} \beta^{t-(T+1)} \cdot \mathbb{P}(\pi^*(\mathbf{s}(t)) = a) \\
&= A_a(T+1).
\end{aligned}$$

Lastly, by our construction of $\mathbf{x}$ and $\mathbf{A}$, it is clear that both are nonnegative, satisfying the non-negativity constraints of problem (3).

Now, let us consider the objective value of $(\mathbf{x}, \mathbf{A})$; we have

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{\infty}\sum_{m=1}^{M} \beta^{t-1} g^m_{s^m(t),\pi^*(\mathbf{s}(t))}\right] &= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{m=1}^{M} \beta^{t-1} g^m_{s^m(t),\pi^*(\mathbf{s}(t))}\right] + \beta^{T+1} \cdot \mathbb{E}\left[\sum_{t=T+1}^{\infty}\sum_{m=1}^{M} \beta^{t-(T+1)} g^m_{s^m(t),\pi^*(\mathbf{s}(t))}\right] \\
&= \sum_{t=1}^{T}\sum_{m=1}^{M} \mathbb{E}\left[g^m_{s^m(t),\pi^*(\mathbf{s}(t))}\right] + \beta^{T+1} \sum_{t=T+1}^{\infty}\sum_{m=1}^{M} \beta^{t-(T+1)} \mathbb{E}\left[g^m_{s^m(t),\pi^*(\mathbf{s}(t))}\right] \\
&= \sum_{t=1}^{T}\sum_{m=1}^{M} \beta^{t-1} \left(\sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} g^m_{ka} \cdot \mathbb{P}(s^m(t) = k,\ \pi^*(\mathbf{s}(t)) = a) \right) \\
&\quad + \beta^{T+1} \sum_{t=T+1}^{\infty} \beta^{t-(T+1)} \left(\sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} g^m_{ka} \cdot \mathbb{P}(s^m(t) = k,\ \pi^*(\mathbf{s}(t)) = a) \right) \\
&= \sum_{t=1}^{T}\sum_{m=1}^{M} \sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} \beta^{t-1} g^m_{ka} x^m_{ka}(t) \\
&\quad + \beta^{T+1} \sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} g^m_{ka} \left(\sum_{t=T+1}^{\infty} \beta^{t-(T+1)} \cdot \mathbb{P}(s^m(t) = k,\ \pi^*(\mathbf{s}(t)) = a) \right) \\
&= \sum_{t=1}^{T+1}\sum_{m=1}^{M} \sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} \beta^{t-1} g^m_{ka} x^m_{ka}(t)
\end{aligned}$$

i.e., the objective value of the $(\mathbf{x}, \mathbf{A})$ solution is identical to the optimal expected discounted value $\mathbb{E}[\sum_{t=1}^{\infty} \sum_{m=1}^{M} \beta^{t-1} g^m_{s^m(t),\pi^*(\mathbf{s}(t))}]$, which is exactly $J^*(\mathbf{s})$. Since $(\mathbf{x}, \mathbf{A})$ is feasible, it follows that this objective value is less than $Z^*_T(\mathbf{s})$, which establishes that $J^*(\mathbf{s}) \leq Z^*_T(\mathbf{s})$ and concludes the proof of part (a) of the Theorem. □

EC.1.5.2. Proof of Part (b) To prove (b), we will use the optimal solution of problem (3) with a horizon of $T+1$ to construct a solution for problem (3) with a horizon of T that (i) is feasible for problem (3) with T and (ii) achieves an objective value of $Z^*_{T+1}(\mathbf{s})$ in problem (3). As in part (a) of the theorem, it will then follow that $Z^*_T(\mathbf{s}) \geq Z^*_{T+1}(\mathbf{s})$.

Let $(\bar{\mathbf{x}}, \bar{\mathbf{A}})$ be the optimal solution of problem (3) with horizon $T+1$. Define the solution $(\mathbf{x}, \mathbf{A})$ to problem (3) with horizon T as follows:

$$\begin{aligned}
&x^m_{ka}(t) = \bar{x}^m_{ka}(t), && \forall t \in \{1,\dots,T\},\\
&x^m_{ka}(T+1) = \bar{x}^m_{ka}(T+1) + \beta \cdot \bar{x}^m_{ka}(T+2),\\
&A_a(t) = \bar{A}_a(t), && \forall t \in \{1,\dots,T\},\\
&A_a(T+1) = \bar{A}_a(T+1) + \beta \cdot \bar{A}_a(T+2).
\end{aligned}$$

By construction, $(\mathbf{x}, \mathbf{A})$ satisfy constraints (3b), (3e), (3f) and (3g). The solution $(\mathbf{x}, \mathbf{A})$ also satisfies constraint (3d) for $t=1$ to $t=T$. To show that constraint (3c) holds, observe that

$$\begin{aligned}
\sum_{a\in\mathcal{A}} x^m_{ja}(T+1) &= \sum_{a\in\mathcal{A}} \bar{x}^m_{ja}(T+1) + \beta \sum_{a\in\mathcal{A}} \bar{x}^m_{ja}(T+2)\\
&= \sum_{a\in\mathcal{A}} \sum_{k\in\mathcal{S}^m} p^m_{kja}\bar{x}^m_{ka}(T) + \beta\left(\sum_{k\in\mathcal{S}^m}\sum_{a\in\mathcal{S}^m} p^m_{kja}\bar{x}^m_{ka}(T+1) + \beta \sum_{k\in\mathcal{S}^m}\sum_{a\in\mathcal{A}} p^m_{kja}\bar{x}^m_{ka}(T+2)\right)\\
&= \sum_{a\in\mathcal{A}} \sum_{k\in\mathcal{S}^m} p^m_{kja}\bar{x}^m_{ka}(T) + \beta \sum_{k\in\mathcal{S}^m}\sum_{a\in\mathcal{S}^m} p^m_{kja}\left(\bar{x}^m_{ka}(T+1) + \beta\bar{x}^m_{ka}(T+2)\right)\\
&= \sum_{a\in\mathcal{A}} \sum_{k\in\mathcal{S}^m} p^m_{kja}x^m_{ka}(T) + \beta \sum_{k\in\mathcal{S}^m}\sum_{a\in\mathcal{S}^m} p^m_{kja}x^m_{ka}(T+1).
\end{aligned}$$

To show that constraint (3d) for $t=T+1$, we have that

$$\sum_{k\in\mathcal{S}^m} x^m_{ka}(T+1) = \sum_{k\in\mathcal{S}^m} \bar{x}^m_{ka}(T+1) + \beta \sum_{k\in\mathcal{S}^m} \bar{x}^m_{ka}(T+2) = \bar{A}_a(T+1) + \beta\bar{A}_a(T+2) = A_a(T+1)$$

Now, consider the objective of $(\mathbf{x}, \mathbf{A})$:

$$\begin{aligned}
\sum_{t=1}^{T+1}\sum_{m=1}^{M}\sum_{k\in\mathcal{S}^m}\sum_{a\in\mathcal{A}} \beta^{t-1}\cdot g^m_{ka}\cdot x^m_{ka}(t) &= \sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{k\in\mathcal{S}^m}\sum_{a\in\mathcal{A}} \beta^{t-1}\cdot g^m_{ka}\cdot \bar{x}^m_{ka}(t) + \sum_{m=1}^{M}\sum_{k\in\mathcal{S}^m}\sum_{a\in\mathcal{A}} \beta^{T}\cdot g^m_{ka}\bar{x}^m_{ka}(T+1)\\
&\quad + \sum_{m=1}^{M}\sum_{k\in\mathcal{S}^m}\sum_{a\in\mathcal{A}} \beta^{T+1} g^m_{ka}\bar{x}^m_{ka}(T+2)\\
&= \sum_{t=1}^{T+2}\sum_{m=1}^{M}\sum_{k\in\mathcal{S}^m}\sum_{a\in\mathcal{A}} \beta^{t-1} g^m_{ka}\bar{x}^m_{ka}(t)
\end{aligned}$$

which, by definition of $(\bar{\mathbf{x}}, \bar{\mathbf{A}})$, is exactly $Z^*_{T+1}(\mathbf{s})$. Since $(\mathbf{x}, \mathbf{A})$ is a feasible solution for problem (3) with horizon T, it follows that $Z^*_T(\mathbf{s}) \geq Z^*_{T+1}(\mathbf{s})$, which concludes the proof of (b). □

EC.1.6. Proof of Theorem 2

EC.1.6.1. Proof of Part (a) To prove the result, we will show that $Z^*_{ALO}(\mathbf{s}) \leq Z^*_{ALR}(\mathbf{s})$ and $Z^*_{ALO}(\mathbf{s}) \geq Z^*_{ALR}(\mathbf{s})$. The inequality $Z^*_{ALO}(\mathbf{s}) \leq Z^*_{ALR}(\mathbf{s})$ can be established by applying Corollary 1 of Adelman and Mersereau (2008), where the weakly coupled MDP is the one defined via the transformation in Section 4.3. We thus restrict our focus to the inequality $Z^*_{ALO}(\mathbf{s}) \geq Z^*_{ALR}(\mathbf{s})$.

Before proceeding with the proof, let us transform problem (12) (the dual of the Lagrangian relaxation problem (11)) for initial state $\mathbf{s}$ to the following equivalent problem:

$$\begin{aligned}
\underset{\mathbf{z},\mathbf{A}}{\text{maximize}} \quad & \sum_{m=1}^{M} \sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} g^m_{ka} z^m_{ka} && \text{(EC.4a)}\\
\text{subject to} \quad & \sum_{a \in \mathcal{A}} z^m_{ja} = \alpha^m_k(\mathbf{s}) + \beta \cdot \sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} p^m_{kja} z^m_{ka}, \quad \forall\, m \in \{1,\dots,M\},\ j \in \mathcal{S}^m, && \text{(EC.4b)}\\
& \sum_{k \in \mathcal{S}^m} z^m_{ka} = A_a, \quad \forall\, m \in \{1,\dots,M\},\ k \in \mathcal{S}^m, && \text{(EC.4c)}\\
& z^m_{ka} \geq 0, \quad \forall\, m \in \{1,\dots,M\},\ k \in \mathcal{S}^m,\ a \in \mathcal{A}, && \text{(EC.4d)}\\
& A_a \geq 0, \quad \forall\, a \in \mathcal{A}, && \text{(EC.4e)}
\end{aligned}$$

where A_a captures the expected discounted frequency with which action a is taken from $t = 1$ on. It is straightforward to see that problem (EC.4) and (12) are equivalent; we have merely reformulated constraint (12c) in problem (12) through the A_a variables. The dual of problem (EC.4) is

$$\begin{aligned}
\underset{\mathbf{V},\boldsymbol{\phi}}{\text{minimize}} \quad & \sum_{m=1}^{M} \sum_{k \in \mathcal{S}^m} \alpha^m_k(\mathbf{s}) V^m_k && \text{(EC.5a)}\\
\text{subject to} \quad & V^m_k \geq g^m_{ka} + \phi^m_a + \beta \cdot \sum_{j \in \mathcal{S}^m} p^m_{kja} V^m_j, \quad \forall\, m \in \{1,\dots,M\},\ k \in \mathcal{S}^m,\ a \in \mathcal{A}, && \text{(EC.5b)}\\
& \sum_{m=1}^{M} \phi^m_a \geq 0, \quad \forall\, a \in \mathcal{A}. && \text{(EC.5c)}
\end{aligned}$$

Let us now return to proving that $Z^*_{ALO}(\mathbf{s}) \geq Z^*_{ALR}(\mathbf{s})$. Let $\mathbf{J}$ be an optimal solution of problem (5). Consider the following solution to problem (EC.5):

$$\begin{aligned}
V^m_k &= J^m_k, \quad \forall\, m \in \{1,\dots,M\},\ k \in \mathcal{S}^m,\ a \in \mathcal{A},\\
\phi^m_a &= \min_{k' \in \mathcal{S}^m} \left\{ J^m_{k'} - g^m_{k'a} - \beta \cdot \sum_{j \in \mathcal{S}^m} p^m_{k'ja} J^m_j \right\}
\end{aligned}$$

It is straightforward to see that the proposed solution $(\mathbf{V}, \boldsymbol{\phi})$ attains the objective value $Z^*_{ALO}(\mathbf{s})$ in problem (EC.5); therefore, it only remains to prove that $(\mathbf{V}, \boldsymbol{\phi})$ is feasible for problem (EC.5).

We will first verify constraint (EC.5b). Observe that by definition of ϕ^m_a, it satisfies

$$\phi^m_a \leq J^m_k - g^m_{ka} - \beta \cdot \sum_{j \in \mathcal{S}^m} p^m_{kja} J^m_j,$$

for any $m \in \{1, \ldots, M\}$ and $k \in \mathcal{S}^m$, which we can re-arrange to obtain

$$J_k^m \geq g_{ka}^m + \phi_a^m + \beta \cdot \sum_{j \in \mathcal{S}^m} p_{kja}^m J_j^m.$$

This last expression, by our definition of $V_k^m = J_k^m$, is equivalent to

$$V_k^m \geq g_{ka}^m + \phi_a^m + \beta \cdot \sum_{j \in \mathcal{S}^m} p_{kja}^m V_j^m;$$

in other words, the proposed solution $(\mathbf{V}, \boldsymbol{\phi})$ satisfies constraint (EC.5b).

Next, let us verify constraint (EC.5c). For a given $a \in \mathcal{A}$, we have that

$$\begin{aligned} \sum_{m=1}^M \phi_a^m &= \sum_{m=1}^M \min_{k' \in \mathcal{S}^m} \left\{ J_{k'}^m - g_{k'a}^m - \beta \cdot \sum_{j \in \mathcal{S}^m} p_{k'ja}^m J_j^m \right\} \\ &= \min_{\bar{\mathbf{s}} \in \mathcal{S}} \left\{ \sum_{m=1}^M \left(J_{\bar{s}^m}^m - g_{\bar{s}^m a}^m - \beta \cdot \sum_{j \in \mathcal{S}^m} p_{\bar{s}^m ja}^m J_j^m \right) \right\} \end{aligned}$$

where the first step follows by definition of $\boldsymbol{\phi}$; and the second step, which is the most crucial, follows by the fact that the per-component minimizations are independent of each other and taken over each component's state space, allowing them to be merged together into a single minimization over the system state space.

Now, recall that $\mathbf{J}$ is a feasible solution to problem (5) and as such, it satisfies

$$\sum_{m=1}^M J_{\bar{s}^m}^m - \sum_{m=1}^M g_{\bar{s}^m a}^m - \beta \sum_{m=1}^M \sum_{j \in \mathcal{S}^m} p_{\bar{s}^m ja}^m J_j^m \geq 0,$$

for any state $\bar{\mathbf{s}}$ and the action a. Therefore, it follows that

$$\sum_{m=1}^M \phi_a^m = \min_{\bar{\mathbf{s}} \in \mathcal{S}} \left\{ \sum_{m=1}^M J_{\bar{s}^m}^m - \sum_{m=1}^M g_{\bar{s}^m a}^m - \beta \cdot \sum_{m=1}^M \sum_{j \in \mathcal{S}^m} p_{\bar{s}^m ja}^m J_j^m \right\} \geq 0$$

which establishes that $(\mathbf{J}, \boldsymbol{\phi})$ satisfies constraint (EC.5c). It now follows that $(\mathbf{J}, \boldsymbol{\phi})$ is feasible for problem (5) and thus, that $Z_{ALO}^*(\mathbf{s}) \geq Z_{ALR}^*(\mathbf{s})$. Since we have established both $Z_{ALO}^*(\mathbf{s}) \geq Z_{ALR}^*(\mathbf{s})$ and $Z_{ALO}^*(\mathbf{s}) \leq Z_{ALR}^*(\mathbf{s})$, it follows that $Z_{ALO}^*(\mathbf{s}) = Z_{ALR}^*(\mathbf{s})$, which concludes the proof. □

EC.1.6.2. Proof of Part (b) In the proof of part (a), we essentially showed that any optimal solution to problem (EC.5) – the dual of problem (EC.4) (which is the transformation of problem (12)) – can be used to construct an optimal solution to the ALO problem (5), and vice versa. It only remains to show that an optimal solution of the ALR problem (11) can be used to construct an optimal solution to problem (EC.5); this is straightforward and is omitted. □

EC.1.7. Proof of Theorem 3

EC.1.7.1. Proof of Part (a) To prove (a), we will use the optimal solution of problem (3) with a horizon of T to construct a solution for problem (12) that is feasible for problem (12) and achieves an objective value of $Z_T^*(\mathbf{s})$ in problem (12). It will then follow that $Z_T^*(\mathbf{s}) \leq Z_{ALR}^*(\mathbf{s})$.

Let $(\mathbf{x}, \mathbf{A})$ be the optimal solution of problem (3) with a horizon of T. Consider a solution to problem (12) defined as follows:

$$z_{ka}^m = \sum_{t=1}^{T+1} \beta^{t-1} \cdot x_{ka}^m(t).$$

To verify that constraint (12b) holds, observe that

$$\begin{aligned}
\sum_{a \in \mathcal{A}} z_{ja}^m &= \sum_{a \in \mathcal{A}} \sum_{t=1}^{T+1} \beta^{t-1} \cdot x_{ja}^m(t) \\
&= \sum_{a \in \mathcal{A}} x_{ja}^m(1) + \sum_{t=2}^{T} \beta^{t-1} \cdot \sum_{a \in \mathcal{A}} x_{ja}^m(t) + \beta^T \cdot \sum_{a \in \mathcal{A}} x_{ja}^m(T+1) \\
&= \alpha_j^m(\mathbf{s}) + \sum_{t=2}^{T} \beta^{t-1} \cdot \sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} p_{kja}^m x_{ka}^m(t-1) + \beta^T \Big(\sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} p_{kja}^m x_{ka}^m(T) + \beta \sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} p_{kja}^m x_{ka}^m(T+1) \Big) \\
&= \alpha_j^m(\mathbf{s}) + \sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} p_{kja}^m \cdot \left(\sum_{t=2}^{T+2} \beta^{t-1} x_{ka}^m(t-1) \right) \\
&= \alpha_j^m(\mathbf{s}) + \beta \sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} p_{kja}^m \cdot \left(\sum_{t=1}^{T+1} \beta^{t-1} x_{ka}^m(t) \right) \\
&= \alpha_j^m(\mathbf{s}) + \beta \sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} p_{kja}^m \cdot z_{ka}^m.
\end{aligned}$$

To verify that constraint (12c) holds, observe that

$$\sum_{k \in \mathcal{S}^m} z_{ka}^m = \sum_{t=1}^{T+1} \sum_{k \in \mathcal{S}^m} \beta^{t-1} x_{ka}^m(t) = \sum_{t=1}^{T+1} A_a(t) = \sum_{t=1}^{T+1} \sum_{k \in \mathcal{S}^{m+1}} \beta^{t-1} x_{ka}^{m+1}(t) = \sum_{k \in \mathcal{S}^{m+1}} z_{ka}^{m+1}.$$

Finally, if we consider the objective of $\mathbf{z}$ in problem (12), we see that

$$\sum_{m=1}^{M} \sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} g_{ka}^m z_{ka}^m = \sum_{m=1}^{M} \sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} g_{ka}^m \beta^{t-1} x_{ka}^m(t) = \sum_{t=1}^{T} \sum_{m=1}^{M} \sum_{k \in \mathcal{S}^m} \sum_{a \in \mathcal{A}} \beta^{t-1} g_{ka}^m x_{ka}^m(t),$$

which is exactly the objective of $(\mathbf{x}, \mathbf{A})$ in problem (3) with a horizon of T, and is equal to $Z_T^*(\mathbf{s})$. Since $\mathbf{z}$ is a feasible solution for problem (12) and $Z_{ALR}^*(\mathbf{s})$ is the optimal reward, it must be that $Z_{ALR}^*(\mathbf{s}) \geq Z_T^*(\mathbf{s})$, which concludes the proof of part (a). □

EC.1.7.2. Proof of Part (b) By Theorem 2, we know that $Z_{ALO}^*(\mathbf{s}) = Z_{ALR}^*(\mathbf{s})$ for all $\mathbf{s} \in \mathcal{S}$, and by part (c) of Theorem 3, we know that $Z_T^*(\mathbf{s}) \leq Z_{ALR}^*(\mathbf{s})$. Therefore, it follows that $Z_T^*(\mathbf{s}) \leq Z_{ALO}^*(\mathbf{s})$ for every $\mathbf{s} \in \mathcal{S}$. □

EC.1.7.3. Proof of Part (c) Part (b) asserts that $Z_T^*(\mathbf{s}) \leq Z_{ALO}^*(\mathbf{s})$, while Proposition 7 (Corollary 1 of Adelman and Mersereau 2008) asserts that $Z_{ALO}^*(\mathbf{s}) \leq Z_{CLR}^*(\mathbf{s})$. Combining these two inequalities, we obtain the desired result. □

EC.1.8. Proof of Theorem 4

To show that $Z_T^*(\mathbf{s}) = Z_{ALR(T)}^*(\mathbf{s})$, write the dual of the ALR(T) problem (13):

$$\underset{\mathbf{z}}{\text{maximize}} \quad \sum_{t=1}^{T+1}\sum_{m=1}^{M}\sum_{k\in\mathcal{S}^m}\sum_{a\in\mathcal{A}} \beta^{t-1}\cdot g_{ka}^m \cdot z_{ka}^m(t) \tag{EC.6a}$$

$$\text{subject to} \quad \sum_{k\in\mathcal{S}^m}\sum_{a\in\mathcal{A}} p_{kja}^m z_{ka}^m(t) = \sum_{a\in\mathcal{A}} z_{ja}^m(t+1), \quad \forall\, m\in\{1,\ldots,M\},\ j\in\mathcal{S}^m,\ t\in\{1,\ldots,T-1\}, \tag{EC.6b}$$

$$\sum_{k\in\mathcal{S}^m}\sum_{a\in\mathcal{A}} p_{kja}^m z_{ka}^m(T) + \beta\sum_{k\in\mathcal{S}^m}\sum_{a\in\mathcal{A}} p_{kja}^m z_{ka}^m(T+1) = \sum_{a\in\mathcal{A}} z_{ja}^m(T+1), \quad \forall\, m\in\{1,\ldots,M\}, j\in\mathcal{S}^m, \tag{EC.6c}$$

$$\sum_{a\in\mathcal{A}} z_{ka}^m(1) = \alpha_k^m(\mathbf{s}), \quad \forall\, m\in\{1,\ldots,M\},\ k\in\mathcal{S}^m, \tag{EC.6d}$$

$$\sum_{k\in\mathcal{S}^m} z_{ka}^m(t) = \sum_{k\in\mathcal{S}^{m+1}} z_{ka}^{m+1}(t), \quad \forall\, m\in\{1,\ldots,M-1\},\ t\in\{1,\ldots,T+1\}, \tag{EC.6e}$$

$$z_{ka}^m(t)\geq 0, \quad \forall\, m\in\{1,\ldots,M\},\ k\in\mathcal{S}^m,\ a\in\mathcal{A}. \tag{EC.6f}$$

This problem is identical to the finite fluid problem (3) (the z variables in problem (EC.6) have the same meaning as the x variables in problem (3)); the difference is that constraint (EC.6e) in problem (EC.6) is expressed using additional variables (the $A_a(t)$ variables) and through a different constraint (constraint (3d)). Therefore, it holds that $Z_T^*(\mathbf{s}) = Z_{ALR(T)}^*(\mathbf{s})$.

To show that $Z_T^*(\mathbf{s}) = Z_{ALO(T)}^*(\mathbf{s})$, let us take the dual of the finite fluid problem (3):

$$\underset{\mathbf{V},\boldsymbol{\phi}}{\text{minimize}} \quad \sum_{m=1}^{M}\sum_{k\in\mathcal{S}^m} \alpha_k^m(\mathbf{s}) V_k^m(1) \tag{EC.7a}$$

$$\text{subject to} \quad J_k^m(t) \geq g_{ka}^m + \phi_a^m(t) + \beta\sum_{j\in\mathcal{S}^m} p_{kja}^m J_j^m(t+1), \quad \forall\, m\in\{1,\ldots,M\},\ k\in\mathcal{S}^m,\ t\in\{1,\ldots,T\}, \tag{EC.7b}$$

$$J_k^m(T+1) \geq g_{ka}^m + \phi_a^m(T+1) + \beta\sum_{j\in\mathcal{S}^m} p_{kja}^m J_j^m(T+1), \quad \forall\, m\in\{1,\ldots,M\},\ k\in\mathcal{S}^m, \tag{EC.7c}$$

$$\sum_{m=1}^{M} \phi_a^m(t)\geq 0, \quad \forall\, m\in\{1,\ldots,M\},\ t\in\{1,\ldots,T+1\}. \tag{EC.7d}$$

To show that $Z_T^*(\mathbf{s}) \geq Z_{ALO(T)}^*(\mathbf{s})$, let $(\mathbf{V},\boldsymbol{\phi})$ be an optimal solution of the finite fluid dual (EC.7), and define a solution $\mathbf{J}$ for ALO(T) problem (14) as

$$J_k^m(t) = V_k^m(t)$$

for each $m \in \{1,\dots,M\}$, $k \in \mathcal{S}^m$ and $t \in \{1,\dots,T+1\}$. Now observe that for any $\mathbf{s} \in \mathcal{S}$ and $t \in \{1,\dots,T\}$, we have

$$\begin{aligned}\sum_{m=1}^{M} J^m_{s^m}(t) &= \sum_{m=1}^{M} V^m_k(t) \\ &\geq \sum_{m=1}^{M}\left(g^m_{s^m a} + \phi^m_a(t) + \beta \sum_{j\in\mathcal{S}^m} p^m_{s^m ja} J^m_j(t+1)\right) \\ &= \sum_{m=1}^{M} g^m_{s^m a} + \sum_{m=1}^{M}\phi^m_a(t) + \beta\sum_{m=1}^{M}\sum_{j\in\mathcal{S}^m} p^m_{s^m ja}J^m_j(t+1) \\ &\geq \sum_{m=1}^{M} g^m_{s^m a} + \beta\sum_{m=1}^{M}\sum_{j\in\mathcal{S}^m} p^m_{s^m ja}J^m_j(t+1)\end{aligned}$$

where the first step follows by our construction of $\mathbf{J}$; the second step, through constraint (EC.7b) that $(\mathbf{V},\boldsymbol{\phi})$ satisfies; the third step by distributing the sum; and the last step by constraint (EC.7d). A similar argument can be used to establish that $\mathbf{J}$ satisfies

$$\sum_{m=1}^{M} J^m_{s^m}(T+1) \geq \sum_{m=1}^{M} g^m_{s^m a} + \beta\sum_{m=1}^{M}\sum_{j\in\mathcal{S}^m} p^m_{s^m ja}J^m_j(T+1).$$

The solution $\mathbf{J}$ is therefore feasible for ALO(T) problem (14); moreover, its objective value is

$$\begin{aligned}\sum_{m=1}^{M}\sum_{k\in\mathcal{S}^m}\alpha^m_k(\mathbf{s})J^m_k(1) &= \sum_{m=1}^{M}\sum_{k\in\mathcal{S}^m}\alpha^m_k(\mathbf{s})V^m_k(1) \\ &= Z^*_T(\mathbf{s}).\end{aligned}$$

We therefore must have that $Z^*_T(\mathbf{s}) \geq Z^*_{ALO(T)}(\mathbf{s})$.

To show that $Z^*_T(\mathbf{s}) \leq Z^*_{ALO(T)}(\mathbf{s})$, let $\mathbf{J}$ be an optimal solution of the ALO(T) problem (14). Define a solution for the finite fluid dual (EC.7) as follows:

$$\begin{aligned}&V^m_k(t) = J^m_k(t), \quad \forall\, m \in \{1,\dots,M\},\ k\in\mathcal{S}^m,\\ &\phi^m_a(t) = \min_{k\in\mathcal{S}^m}\left(J^m_k(T+1) - g^m_{ka} - \beta\sum_{j\in\mathcal{S}^m} p^m_{kja}J^m_j(t+1)\right), \quad \forall m\in\{1,\dots,M\},\ a\in\mathcal{A},\ t\in\{1,\dots,T\},\\ &\phi^m_a(T+1) = \min_{k\in\mathcal{S}^m}\left(J^m_k(T+1) - g^m_{ka} - \beta\sum_{j\in\mathcal{S}^m} p^m_{kja}J^m_j(T+1)\right), \quad \forall m\in\{1,\dots,M\},\ a\in\mathcal{A}.\end{aligned}$$

Observe that by definition of $\phi^m_a(t)$ for $t < T+1$, we have that

$$\phi^m_a(t) \leq J^m_k(t) - g^m_{ka} - \beta\sum_{j\in\mathcal{S}^m} p^m_{kja}J^m_j(t+1),$$

for any $m\in\{1,\dots,M\}$ and $k\in\mathcal{S}^m$, which we can re-arrange to obtain

$$J^m_k(t) \geq g^m_{ka} + \phi^m_a(t) + \beta\sum_{j\in\mathcal{S}^m} p^m_{kja}J^m_j(t+1),$$

or equivalently

$$V_k^m(t) \geq g_{ka}^m + \phi_a^m(t) + \beta \sum_{j \in \mathcal{S}^m} p_{kja}^m V_j^m(t+1).$$

Similarly, for $T+1$, we have

$$\phi_a^m(T+1) \leq J_k^m(T+1) - g_{ka}^m - \beta \sum_{j \in \mathcal{S}^m} p_{kja}^m J_j^m(T+1),$$

which can be re-arranged to get

$$J_k^m(T+1) \geq g_{ka}^m + \phi_a^m(T+1) + \beta \sum_{j \in \mathcal{S}^m} p_{kja}^m J_j^m(T+1),$$

or equivalently

$$V_k^m(T+1) \geq g_{ka}^m + \phi_a^m(T+1) + \beta \sum_{j \in \mathcal{S}^m} p_{kja}^m V_j^m(T+1);$$

this verifies constraints (EC.7b) and (EC.7c).

To verify the last constraint (EC.7d), we have for any given $a \in \mathcal{A}$ and $t \in \{1, \ldots, T\}$ that

$$\begin{aligned}\sum_{m=1}^{M} \phi_a^m(t) &= \sum_{m=1}^{M} \min_{k \in \mathcal{S}^m} \left\{ J_k^m(t) - g_{ka}^m - \beta \sum_{j \in \mathcal{S}^m} p_{kja}^m J_j^m(t+1) \right\} \\ &= \min_{\bar{\mathbf{s}} \in \mathcal{S}} \left\{ \sum_{m=1}^{M} \left(J_{\bar{s}^m}^m(t) - g_{\bar{s}^m a}^m - \beta \sum_{j \in \mathcal{S}^m} p_{\bar{s}^m ja}^m J_j^m \right) \right\},\end{aligned}$$

where, as in the proof of part (a) of Theorem 2 (see Section EC.1.6), the order of summation and minimization can be interchanged because the minimizations are independent of each other. Since $\mathbf{J}$ is feasible for problem (14), it must be that

$$\sum_{m=1}^{M} J_{\bar{s}^m}^m(t) - \sum_{m=1}^{M} g_{\bar{s}^m a}^m - \beta \sum_{m=1}^{M} \sum_{j \in \mathcal{S}^m} p_{\bar{s}^m ja}^m J_j^m(t+1) \geq 0,$$

for all $\mathbf{s} \in \mathcal{S}$; it therefore follows that

$$\min_{\bar{\mathbf{s}} \in \mathcal{S}} \left\{ \sum_{m=1}^{M} J_{\bar{s}^m}^m(t) - \sum_{m=1}^{M} g_{\bar{s}^m a}^m - \beta \sum_{m=1}^{M} \sum_{j \in \mathcal{S}^m} p_{\bar{s}^m ja}^m J_j^m(t+1) \right\} \geq 0,$$

which establishes that

$$\sum_{m=1}^{M} \phi_a^m(t) \geq 0.$$

Similar steps can be used to establish that

$$\sum_{m=1}^{M} \phi_a^m(T+1) \geq 0$$

for $a \in \mathcal{A}$. Thus, $(\mathbf{V}, \boldsymbol{\phi})$ is a feasible solution for the finite fluid dual (EC.7); it is also easy to see that its objective value is exactly $Z_{ALO(T)}^*(\mathbf{s})$. This establishes that $Z_T^*(\mathbf{s}) \leq Z_{ALO(T)}^*(\mathbf{s})$.

Together with $Z_T^*(\mathbf{s}) \geq Z_{ALO(T)}^*(\mathbf{s})$, this establishes that $Z_T^*(\mathbf{s}) = Z_{ALO(T)}^*(\mathbf{s})$. This concludes the proof. □

EC.1.9. Proof of Proposition 9

The proof follows by showing that $Z^*_{BNM}(\mathbf{s}) \leq Z^*_{ALR}(\mathbf{s})$ and $Z^*_{BNM}(\mathbf{s}) \geq Z^*_{ALR}(\mathbf{s})$. To do this, we follow our approach in the other proofs of using one problem's optimal solution to construct a feasible solution for the other problem that has the same objective value as the first problem's optimal solution. The most difficult part of the proof is deriving a suitable feasible solution for one problem from the optimal solution of the other; verifying that the solutions are feasible is straightforward, but somewhat laborious. In what follows, we will present several identities that are useful in verifying feasibility and then provide the correct feasible solutions, but in the interest of space, we will omit the verification of the solutions.

Let us derive a few properties of problems (12) and (15) that will be useful to us. First of all, notice that for any $\mathbf{z}$ feasible for problem (12), for a fixed $m \in \{1, \ldots, M\}$, by summing constraint (12b) over all states $j \in \mathcal{S}^m$, we have

$$\sum_{j \in \mathcal{S}^m} \sum_{a'=1}^{M} z^m_{ja'} - \beta \sum_{j \in \mathcal{S}^m} \sum_{k \in \mathcal{S}^m} \sum_{a=1}^{M} p^m_{kja} z^m_{ka} = \sum_{j \in \mathcal{S}^m} \alpha^m_j(\mathbf{s}),$$

which is equal to

$$\sum_{j \in \mathcal{S}^m} \sum_{a'=1}^{M} z^m_{ja'} - \beta \sum_{k \in \mathcal{S}^m} \sum_{a=1}^{M} \left(\sum_{j \in \mathcal{S}^m} p^m_{kja} \right) \cdot z^m_{ka} = \sum_{j \in \mathcal{S}^m} \alpha^m_j(\mathbf{s}).$$

Since $\sum_{j \in \mathcal{S}^m} p^m_{kja} = 1$ and $\sum_{j \in \mathcal{S}^m} \alpha^m_j(\mathbf{s}) = 1$, this reduces to

$$\sum_{j \in \mathcal{S}^m} \sum_{a'=1}^{M} z^m_{ja'} - \beta \sum_{k \in \mathcal{S}^m} \sum_{a=1}^{M} z^m_{ka} = 1.$$

Observe that the two sums are identical; by dividing through by $(1-\beta)$, we obtain the following identity:

$$\sum_{k \in \mathcal{S}^m} \sum_{a=1}^{M} z^m_{ka} = \frac{1}{1-\beta}, \qquad \forall\, m \in \{1, \ldots, M\}. \tag{EC.8}$$

Observe that by following the same steps, we can derive a similar identity for the $\mathbf{w}$ variables of problem (15) using constraint (15c):

$$\sum_{k \in \mathcal{S}^m} \sum_{a' \in \{0,1\}} w^m_{ka'} = \frac{1}{1-\beta}, \qquad \forall\, m \in \{1, \ldots, M\}. \tag{EC.9}$$

Lastly, notice that by combining constraint (15b) and identity (EC.9) for a fixed $m \in \{1, \ldots, M\}$, we get that

$$\sum_{m'=1}^{M} \sum_{k \in \mathcal{S}^{m'}} w^{m'}_{k1} = \sum_{k \in \mathcal{S}^m} \sum_{a \in \{0,1\}} w^m_{ka}.$$

Notice that both sums contain the sum $\sum_{k\in\mathcal{S}^m} w_{k1}^m$; by subtracting it out, we obtain the following identity:

$$\sum_{m'\neq m}\sum_{k\in\mathcal{S}^{m'}} w_{k1}^{m'} = \sum_{k\in\mathcal{S}^m} w_{k0}^m, \qquad \forall\, m\in\{1,\ldots,M\}. \tag{EC.10}$$

Having defined these identities, we are ready to move on with the proof.

To prove that $Z_{ALR}^*(\mathbf{s}) \geq Z_{BNM}^*(\mathbf{s})$, suppose that $\mathbf{w}$ is an optimal solution to problem (15). We will proceed in two different cases.

Case 1 of $Z_{ALR}^*(\mathbf{s}) \geq Z_{BNM}^*(\mathbf{s})$. Suppose that $\sum_{j\in\mathcal{S}^m} w_{j0}^m > 0$ for every $m\in\{1,\ldots,M\}$. Define a solution to problem (12) as

$$\begin{aligned}
z_{km}^m &= w_{k1}^m, && \forall\, m\in\{1,\ldots,M\},\ k\in\mathcal{S}^m,\\
z_{ka}^m &= \frac{w_{k0}^m}{\sum_{j\in\mathcal{S}^m} w_{j0}^m}\cdot\sum_{j\in\mathcal{S}^a} w_{j1}^a && \forall\, m\in\{1,\ldots,M\},\ a\neq m,\ k\in\mathcal{S}^m.
\end{aligned}$$

It can be shown that $\mathbf{z}$ is feasible for problem (12) and that its objective in problem (12) is exactly $Z_{BNM}^*(\mathbf{s})$.

Case 2 of $Z_{ALR}^*(\mathbf{s}) \geq Z_{BNM}^*(\mathbf{s})$. In the second case, we suppose that there exists an $\tilde{m}$ such that $\sum_{j\in\mathcal{S}^m} w_{j0}^{\tilde{m}} = 0$. In this case, the expected discounted frequency with which we do not active bandit $\tilde{m}$ when it is in any of its states is zero; intuitively, this means that we are always activating bandit $\tilde{m}$.

Armed with this intuition and the above identity, let us define a solution $\mathbf{z}$ for problem (12) as

$$\begin{aligned}
z_{ka}^m &= 0, && \forall\, m\in\{1,\ldots,M\},\ k\in\mathcal{S}^m,\ a\neq\tilde{m},\\
z_{k\tilde{m}}^m &= w_{k0}^m, && \forall\, m\neq\tilde{m},\ k\in\mathcal{S}^{\tilde{m}},\\
z_{k\tilde{m}}^{\tilde{m}} &= w_{k1}^{\tilde{m}}, && \forall k\in\mathcal{S}^{\tilde{m}}.
\end{aligned}$$

As in Case 1, it can then be shown that $\mathbf{z}$ is feasible for problem (12) and that its objective in problem (12) is exactly $Z_{BNM}^*(\mathbf{s})$.

This proves that $Z_{BNM}^*(\mathbf{s}) \leq Z_{ALR}^*(\mathbf{s})$; we now turn our attention to $Z_{BNM}^*(\mathbf{s}) \geq Z_{ALR}^*(\mathbf{s})$. To prove this inequality, suppose that we have a solution $\mathbf{z}$ to problem (12). Define the solution $\mathbf{w}$ to problem (15) as

$$\begin{aligned}
w_{k0}^m &= \sum_{a\neq m} z_{ka}^m, && \forall\, m\in\{1,\ldots,M\},\ k\in\mathcal{S}^m,\\
w_{k1}^m &= z_{km}^m, && \forall\, m\in\{1,\ldots,M\},\ k\in\mathcal{S}^m.
\end{aligned}$$

It can be shown that $\mathbf{w}$ is feasible for problem (15) and that its objective in problem (15) is exactly $Z^*_{ALR}(\mathbf{s})$.

From here, since $Z^*_{ALR}(\mathbf{s}) \leq Z^*_{BNM}(\mathbf{s})$ and $Z^*_{ALR}(\mathbf{s}) \geq Z^*_{BNM}(\mathbf{s})$, it follows that $Z^*_{ALR}(\mathbf{s}) = Z^*_{BNM}(\mathbf{s})$, as required. □

EC.2. Counterexample to show that $Z^*(\mathbf{s}) \leq J^*(\mathbf{s})$ does not always hold

To develop the counterexample, we proceed in two steps. First, we develop a useful bound that relates the optimal value of a truncated fluid formulation and the infinite fluid formulation (problem (2)). Then, we describe a specific problem instance where the numerical values of the optimal DP value function and the optimal value of the finite fluid formulation, together with the bound we just developed, allow us to conclude that $Z^*(\mathbf{s}) \leq J^*(\mathbf{s})$ does not hold.

EC.2.1. Bound

Clearly, we are not able to solve the optimization problem (2), since it is a problem with a countably infinite number of constraints and variables. However, if we only include the decision variables and constraints corresponding to $t = 1$ to $t = T$, and truncate the objective function to the finite sum $\sum_{t=1}^T \sum_{m=1}^M \sum_{k=1}^n \sum_{a=1}^M \beta^{t-1} g^m_{ka} x^m_{ka}(t)$, we obtain a finite formulation that we are able to solve. (This is in fact problem (3) without the $T+1$ objective term that accounts for the system's evolution from $t = T+1$ on.) Let $Z^*_{T,\text{trunc}}(\mathbf{s})$ be the optimal value of this truncated problem. The finite-length optimal solution $(\mathbf{x}, \mathbf{A})$ that attains this objective value can be easily extended for times $t > T$ (e.g., by setting $A_1(t) = 1$ for $t > T$, $x^m_{ka}(t)$ and $A_a(t)$ are uniquely determined for $t > T$) to yield a completed, infinite-length solution. Using this infinite-length solution, we see that

$$Z^*_{T,\text{trunc}}(\mathbf{s}) = \sum_{t=1}^{T} \sum_{m=1}^{M} \sum_{k=1}^{n} \sum_{a=1}^{M} \beta^{t-1} g^m_{ka} x^m_{ka}(t) \leq \sum_{t=1}^{\infty} \sum_{m=1}^{M} \sum_{k=1}^{n} \sum_{a=1}^{M} \beta^{t-1} g^m_{ka} x^m_{ka}(t) \leq Z^*(\mathbf{s}),$$

where the first step follows by the definition of $Z^*_T(\mathbf{s})$ as the optimal value of the truncated formulation, the second by the fact that the value of the sum can only increase when it is extended from a finite one to an infinite one (all g^m_{ka} and $x^m_{ka}(t)$ values are nonnegative), and the third by the fact that the completed, infinite-length solution is a feasible solution to the infinite fluid formulation (2).

EC.2.2. Instance

Consider a decomposable MDP with $M = 3$, $\mathcal{S}^1 = \mathcal{S}^2 = \mathcal{S}^3 = \{1, 2, 3\}$, with the following reward data:

$$\mathbf{g}^1_1 = \begin{bmatrix} 3.3323 \\ 0.2765 \\ 2.2476 \end{bmatrix}, \ \mathbf{g}^2_2 = \begin{bmatrix} 2.6230 \\ 8.9540 \\ 4.5576 \end{bmatrix}, \mathbf{g}^3_3 = \begin{bmatrix} 2.8975 \\ 7.8568 \\ 0.8381 \end{bmatrix}, \tag{EC.11}$$

$$\mathbf{g}^m_a = \mathbf{0}, \ \forall m \in \{1, 2, 3\}, \ a \in \{1, 2, 3\}, \ a \neq m, \tag{EC.12}$$

and the following probability transition data:

$$\mathbf{p}_1^1 = \begin{bmatrix} 0.0540 & 0.3790 & 0.5670 \\ 0.0190 & 0.0390 & 0.9420 \\ 0.3120 & 0.5330 & 0.1550 \end{bmatrix}, \quad \mathbf{p}_1^2 = \begin{bmatrix} 0.8560 & 0 & 0.1440 \\ 0.1590 & 0.3150 & 0.5260 \\ 0.0320 & 0.8050 & 0.1630 \end{bmatrix}, \quad \mathbf{p}_1^3 = \begin{bmatrix} 0.0740 & 0.9020 & 0.0240 \\ 0.3440 & 0.0200 & 0.6360 \\ 0.5820 & 0.3750 & 0.0430 \end{bmatrix} \tag{EC.13}$$

$$\mathbf{p}_a^m = \mathbf{I}, \ \forall m \in \{1,2,3\}, \ a \in \{1,2,3\}, \ a \neq m. \tag{EC.14}$$

As we will see in Section 5, this problem actually corresponds to a regular multiarmed bandit problem with three arms.

Suppose that the discount factor is 0.9 and the initial state $\mathbf{s}$ is set to $(3,3,1)$. Solving the truncated version of problem (2) with $T = 100$ for initial state $\mathbf{s}$, we obtain an objective value of $Z^*_{T,\text{trunc}}(\mathbf{s}) = 57.7134$, while the optimal DP value function is $J^*(\mathbf{s}) = 57.3812$. By the bound above, we have

$$J^*(\mathbf{s}) = 57.3812 < 57.7134 = Z^*_{T,\text{trunc}}(\mathbf{s}) \leq Z^*(\mathbf{s}),$$

i.e., that $J^*(\mathbf{s}) < Z^*(\mathbf{s})$. This allows us to conclude that $J^*(\mathbf{s}) \geq Z^*(\mathbf{s})$ does not always hold.

EC.3. Derivation of alternate Lagrangian relaxation

In this section, we derive the ALR formulation. The steps that we follow here are essentially the same as those used in Adelman and Mersereau (2008) to derive the Lagrangian relaxation formulation, applied to the specific weakly-coupled MDP that is at the heart of the ALR. For completeness, we show the main steps of the derivation here.

The optimal value function for the true MDP of interest satisfies the following Bellman equation:

$$J^*(\mathbf{s}) = \max_{\substack{(a^1,\ldots,a^M) \in \mathcal{A} \times \cdots \times \mathcal{A}: \\ \mathbb{I}\{a^m = a\} - \mathbb{I}\{a^{m+1} = a\} = 0, \\ \forall m \in \{1,\ldots,M-1\}, \ a \in \mathcal{A}}} \left(\sum_{m=1}^M g^m_{s^m a^m} + \beta \cdot \sum_{\bar{\mathbf{s}} \in \mathcal{S}} \left(\prod_{m=1}^M p^m_{s^m \bar{s}^m a^m} \right) J^*(\bar{\mathbf{s}}) \right). \tag{EC.15}$$

In the Lagrangian relaxation approach (Hawkins 2003, Adelman and Mersereau 2008), we dualize the action consistency constraint on the action vectors $(a^1, \ldots, a^M)$. For each constraint in the maximization, we introduce a Lagrange multiplier $\lambda^m_a \in \mathbb{R}$, and penalize the violation of the corresponding (m,a) constraint in the Bellman iteration. We obtain a new value function $J^{\boldsymbol{\lambda}}(\mathbf{s})$ which satisfies the following modified Bellman equation:

$$\begin{aligned} J^{\boldsymbol{\lambda}}(\mathbf{s}) = \max_{(a^1,\ldots,a^M) \in \mathcal{A} \times \cdots \times \mathcal{A}} & \left(\sum_{m=1}^M g^m_{s^m a^m} + \beta \cdot \sum_{\bar{\mathbf{s}} \in \mathcal{S}} \left(\prod_{m=1}^M p^m_{s^m \bar{s}^m a^m} \right) J^{\boldsymbol{\lambda}}(\bar{\mathbf{s}}) \right. \\ & \left. - \sum_{m=1}^{M-1} \sum_{a \in \mathcal{A}} \lambda^m_a (\mathbb{I}\{a^m = a\} - \mathbb{I}\{a^{m+1} = a\}) \right) \qquad \text{(EC.16)} \\ = \max_{(a^1,\ldots,a^M) \in \mathcal{A} \times \cdots \times \mathcal{A}} & \left(\sum_{m=1}^M g^m_{s^m a^m} + \beta \cdot \sum_{\bar{\mathbf{s}} \in \mathcal{S}} \left(\prod_{m=1}^M p^m_{s^m \bar{s}^m a^m} \right) J^{\boldsymbol{\lambda}}(\bar{\mathbf{s}}) \right. \end{aligned}$$

$$- \sum_{m=1}^{M-1} (\lambda^m_{a^m} - \lambda^m_{a^{m+1}}) \Bigg) \tag{EC.17}$$

We will show that the solution to the above Bellman equation is of the form

$$J^{\boldsymbol{\lambda}}(\mathbf{s}) = \sum_{m=1}^{M} V^{m,\boldsymbol{\lambda}}(s^m), \tag{EC.18}$$

where each $V^{m,\boldsymbol{\lambda}}$ is a component-wise value function satisfying

$$V^{m,\boldsymbol{\lambda}}(s^m) = \max_{a \in \mathcal{A}} \left(g^m_{s^m a} + \beta \sum_{\bar{s}^m \in \mathcal{S}^m} p^m_{s^m \bar{s}^m a^m} V^{m,\boldsymbol{\lambda}}(\bar{s}^m) - \mathbb{I}\{m < M\} \cdot \lambda^m_a + \mathbb{I}\{m > 1\} \cdot \lambda^{m-1}_a \right) \tag{EC.19}$$

To see this, we will show that the above form satisfies equation (EC.18). We have

$$\begin{aligned}
&\max_{(a^1,\dots,a^M) \in \mathcal{A} \times \dots \times \mathcal{A}} \left(\sum_{m=1}^{M} g^m_{s^m a^m} + \beta \cdot \sum_{\bar{\mathbf{s}} \in \mathcal{S}} \left(\prod_{m=1}^{M} p^m_{s^m \bar{s}^m a^m} \right) \left(\sum_{m=1}^{M} V^{m,\boldsymbol{\lambda}}(\bar{s}^m) \right) - \sum_{m=1}^{M-1} (\lambda^m_{a^m} - \lambda^m_{a^{m+1}}) \right) \\
&= \max_{(a^1,\dots,a^M) \in \mathcal{A} \times \dots \times \mathcal{A}} \left(\sum_{m=1}^{M} g^m_{s^m a^m} + \beta \cdot \sum_{m=1}^{M} \sum_{\bar{s}^m \in \mathcal{S}^m} p^m_{s^m \bar{s}^m a^m} V^{m,\boldsymbol{\lambda}}(\bar{s}^m) - \sum_{m=1}^{M-1} (\lambda^m_{a^m} - \lambda^m_{a^{m+1}}) \right) \\
&= \sum_{m=1}^{M} \max_{a^m \in \mathcal{A}} \left(g^m_{s^m a^m} + \beta \sum_{\bar{s}^m \in \mathcal{S}^m} p^m_{s^m \bar{s}^m a^m} V^{m,\boldsymbol{\lambda}}(\bar{s}^m) - \mathbb{I}\{m < M\} \cdot \lambda^m_a + \mathbb{I}\{m > 1\} \cdot \lambda^{m-1}_a \right) \\
&= \sum_{m=1}^{M} V^{m,\boldsymbol{\lambda}}(s^m),
\end{aligned}$$

where the first equality follows by the linearity of expectation (the sum $\sum_{\bar{\mathbf{s}} \in \mathcal{S}} \left(\prod_{m=1}^{M} p^m_{s^m \bar{s}^m a^m} \right) \left(\sum_{m=1}^{M} V^{m,\boldsymbol{\lambda}}(\bar{s}^m) \right)$ can be viewed as the expectation of a sum of random variables that correspond to each component's value function at a random next state); the second by the fact that the maximizations over each of the a^m variables are independent of each other; and the third by definition of the $V^{m,\boldsymbol{\lambda}}_k$'s.

To now derive the ALR formulation, let $\mathbf{s}$ be the initial state. The value of $\mathbf{s}$ is $J^{\boldsymbol{\lambda}}(\mathbf{s}) = \sum_{m=1}^{M} V^{m,\boldsymbol{\lambda}}(s^m)$. Each component value function $V^{m,\boldsymbol{\lambda}}$, described by the Bellman equation in equation (EC.19), can be evaluated at s^m by solving the following linear optimization problem, where $\boldsymbol{\lambda}$ is fixed (i.e., not a decision variable):

$$V^{m,\boldsymbol{\lambda}}(s^m) \quad = \quad \underset{\mathbf{V}^m}{\text{minimize}} \quad \sum_{k \in \mathcal{S}^m} \alpha^m_k(\mathbf{s}) V^m_k \tag{EC.20a}$$

$$\text{subject to} \quad V^m_k \geq g^m_{ka} - \mathbb{I}\{m < M\} \cdot \lambda^m_a + \mathbb{I}\{m > 1\} \cdot \lambda^{m-1}_a + \beta \cdot \sum_{j \in \mathcal{S}^m} p^m_{kja} V^m_j, \quad \forall\ k \in \mathcal{S}^m,\ a \in \mathcal{A}. \tag{EC.20b}$$

The value $J^{\boldsymbol{\lambda}}(\mathbf{s})$ can then be expressed as the optimal value of the following optimization problem, which combines the above component-wise optimization problems into one problem:

$$
\begin{aligned}
J^{\boldsymbol{\lambda}}(\mathbf{s}) \quad = \quad \underset{\mathbf{V}}{\text{minimize}} \quad & \sum_{m=1}^{M} \sum_{k \in \mathcal{S}^m} \alpha_k^m(\mathbf{s}) V_k^m && \text{(EC.21a)} \\
\text{subject to} \quad & V_k^m \geq g_{ka}^m - \mathbb{I}\{m < M\} \cdot \lambda_a^m + \mathbb{I}\{m > 1\} \cdot \lambda_a^{m-1} + \beta \cdot \sum_{j \in \mathcal{S}^m} p_{kja}^m V_j^m, \\
& \quad \forall\ m \in \{1, \ldots, M\},\ k \in \mathcal{S}^m,\ a \in \mathcal{A}. && \text{(EC.21b)}
\end{aligned}
$$

As in Adelman and Mersereau (2008), it can be shown that the optimal value of the above problem, which is equal to $J^{\boldsymbol{\lambda}}(\mathbf{s})$, is an upper bound on $J^*(\mathbf{s})$. We now seek to find the tightest such upper bound, that is, $\min_{\boldsymbol{\lambda}} J^{\boldsymbol{\lambda}}(\mathbf{s})$. This can be accomplished by optimizing over $\boldsymbol{\lambda}$ as an additional decision variable in the above optimization problem:

$$
\begin{aligned}
\underset{\mathbf{V}, \boldsymbol{\lambda}}{\text{minimize}} \quad & \sum_{m=1}^{M} \sum_{k \in \mathcal{S}^m} \alpha_k^m(\mathbf{s}) V_k^m && \text{(EC.22a)} \\
\text{subject to} \quad & V_k^m \geq g_{ka}^m - \mathbb{I}\{m < M\} \cdot \lambda_a^m + \mathbb{I}\{m > 1\} \cdot \lambda_a^{m-1} + \beta \cdot \sum_{j \in \mathcal{S}^m} p_{kja}^m V_j^m, \\
& \quad \forall\ m \in \{1, \ldots, M\},\ k \in \mathcal{S}^m,\ a \in \mathcal{A}. && \text{(EC.22b)}
\end{aligned}
$$

The above formulation is exactly the ALR formulation (11).