OPERATIONS RESEARCH
DOI 10.1287/opre.1070.0445ec
pp. ec1–ec7

e-companion

ONLY AVAILABLE IN ELECTRONIC FORM

Electronic Companion—"Relaxations of Weakly Coupled Stochastic Dynamic Programs" by Daniel Adelman and Adam J. Mersereau, *Operations Research* DOI 10.1287/opre.1070.0445.

Electronic Companion to "Relaxations of Weakly Coupled Stochastic Dynamic Programs"

Daniel Adelman, Adam J. Mersereau

This document accompanies "Relaxations of Weakly Coupled Stochastic Dynamic Programs" by Adelman and Mersereau. References herein point to that document, and we use notation specified there.

Appendix A: Proof of Proposition 1

Observe that (6) solves (5) if the following difference can be shown to equal zero:

$$\begin{aligned}
&\max_{\mathbf{a}\in A(\mathbf{s})}\left\{\sum_{i=1}^{I} r_i(s_i,a_i)+\boldsymbol{\lambda}^\top\left[\mathbf{b}-\sum_{i=1}^{I}\mathbf{D}_i(s_i,a_i)\right]+\beta\sum_{\mathbf{s}'\in S}p(\mathbf{s}'|\mathbf{s},\mathbf{a})\left[\frac{\boldsymbol{\lambda}^\top\mathbf{b}}{1-\beta}+\sum_{i=1}^{I}V_i^{\boldsymbol{\lambda}}(s_i')\right]\right\}\\
&\qquad -\frac{\boldsymbol{\lambda}^\top\mathbf{b}}{1-\beta}-\sum_{i=1}^{I}V_i^{\boldsymbol{\lambda}}(s_i)\\
=\;& -\frac{\boldsymbol{\lambda}^\top}{1-\beta}\left[\mathbf{b}-\mathbf{b}(1-\beta)-\beta\mathbf{b}\right]\\
&\qquad +\max_{\mathbf{a}\in A(\mathbf{s})}\left\{\sum_{i=1}^{I}\left(r_i(s_i,a_i)-\boldsymbol{\lambda}^\top\mathbf{D}_i(s_i,a_i)\right)+\beta\sum_{\mathbf{s}'\in S}p(\mathbf{s}'|\mathbf{s},\mathbf{a})\sum_{i=1}^{I}V_i^{\boldsymbol{\lambda}}(s_i')\right\}-\sum_{i=1}^{I}V_i^{\boldsymbol{\lambda}}(s_i)
\end{aligned}$$

The term in square brackets vanishes. We introduce the notation S_{-i} to represent the set $\times_{j\neq i}S_j$ and $\mathbf{s}_{-i}$ to represent the vector $\mathbf{s}$ exclusive of component i. Simplifying, the expression becomes

$$\begin{aligned}
&\max_{\mathbf{a}\in A(\mathbf{s})}\left\{\sum_{i=1}^{I}\left(r_i(s_i,a_i)-\boldsymbol{\lambda}^\top\mathbf{D}_i(s_i,a_i)\right)+\beta\sum_{i=1}^{I}\sum_{s_i'\in S_i}\sum_{\mathbf{s}'_{-i}\in S_{-i}}p(\mathbf{s}'|\mathbf{s},\mathbf{a})V_i^{\boldsymbol{\lambda}}(s_i')\right\}-\sum_{i=1}^{I}V_i^{\boldsymbol{\lambda}}(s_i)\\
=\;&\max_{\mathbf{a}\in A(\mathbf{s})}\left\{\sum_{i=1}^{I}\left(r_i(s_i,a_i)-\boldsymbol{\lambda}^\top\mathbf{D}_i(s_i,a_i)\right)+\beta\sum_{i=1}^{I}\sum_{s_i'\in S_i}V_i^{\boldsymbol{\lambda}}(s_i')\sum_{\mathbf{s}'_{-i}\in S_{-i}}p(\mathbf{s}'|\mathbf{s},\mathbf{a})\right\}-\sum_{i=1}^{I}V_i^{\boldsymbol{\lambda}}(s_i)\\
=\;&\max_{\mathbf{a}\in A(\mathbf{s})}\left\{\sum_{i=1}^{I}\left(r_i(s_i,a_i)-\boldsymbol{\lambda}^\top\mathbf{D}_i(s_i,a_i)\right)\right.\\
&\qquad \left.+\beta\sum_{i=1}^{I}\sum_{s_i'\in S_i}V_i^{\boldsymbol{\lambda}}(s_i')p_i(s_i'|s_i,a_i)\sum_{\mathbf{s}'_{-i}\in S_{-i}}\prod_{k\neq i}p_k(s_k'|s_k,a_k)\right\}-\sum_{i=1}^{I}V_i^{\boldsymbol{\lambda}}(s_i)\\
=\;&\max_{\mathbf{a}\in A(\mathbf{s})}\left\{\sum_{i=1}^{I}\left(r_i(s_i,a_i)-\boldsymbol{\lambda}^\top\mathbf{D}_i(s_i,a_i)\right)+\beta\sum_{i=1}^{I}\sum_{s_i'\in S_i}p_i(s_i'|s_i,a_i)V_i^{\boldsymbol{\lambda}}(s_i')\right\}-\sum_{i=1}^{I}V_i^{\boldsymbol{\lambda}}(s_i)
\end{aligned}$$

$$= \sum_{i=1}^{I} \left[\max_{a_i \in A_i(s_i)} \left\{ \left(r_i(s_i,a_i) - \boldsymbol{\lambda}^{\mathsf{T}} \mathbf{D}_i(s_i,a_i) \right) + \beta \sum_{s_i' \in S_i} p_i(s_i'|s_i,a_i) V_i^{\boldsymbol{\lambda}}(s_i) \right\} \right] - \sum_{i=1}^{I} V_i^{\boldsymbol{\lambda}}(s_i), \qquad (19)$$

for all $\mathbf{s} \in S$.

Equation (7) is the optimality equation for a discounted dynamic program with finite state space and bounded rewards, thus its solution exists by Puterman (1994), Theorem 6.2.5. A solution to (7) will also clearly make (19) equal to zero. Thus, our proposed value function (6) solves optimality equations (5) and is optimal by Puterman (1994), Thm. 6.2.2. □

Appendix B: A Counterexample to the State-Wise Bound

Assume two subproblems ($I = 2$), each of which is defined on two states. Let the subproblem state spaces each be $\{1,2\}$ so that the overall problem state space is $\mathbf{s} = (s_1, s_2) \in \{(1,1),(1,2),(2,1),(2,2)\}$. We assume the discount factor $\beta = 0$ so that dynamics are immaterial, and the initial state distribution is $\alpha_{11} = \Pr\{\mathbf{s^0} = (1,1)\} = 0.1$, $\alpha_{12} = \Pr\{\mathbf{s^0} = (1,2)\} = 0.7$, $\alpha_{21} = \Pr\{\mathbf{s^0} = (2,1)\} = 0.1$, $\alpha_{22} = \Pr\{\mathbf{s^0} = (2,2)\} = 0.1$.

Two controls, "active" ($a_i = 1$) and "passive" ($a_i = 0$) are available for subproblem i, and the controller must abide by the constraint $\sum_i D_i(s_i,a_i) \le 1$. The controller seeks to maximize $\sum_i r_i(s_i,a_i)$. Constraint coefficients and rewards are as follows:

- $r_i(s_i,a_i) = 0$ for all i, s_i, a_i, except $r_1(1,1) = 1$ and $r_2(2,1) = 1$.
- $D_i(s_i,a_i) = 0$ for all i, s_i, a_i, except $D_1(1,1) = 1.0$, $D_1(2,0) = 0.4$, $D_1(2,1) = 0.6$, $D_2(1,1) = 0.6$, and $D_2(2,1) = 0.6$.

The following table summarizes the solutions of the two relaxations.

$H^{LP}(\boldsymbol{\alpha}) = 1$	$H^{\lambda^*}(\boldsymbol{\alpha}) = 1.24$ $(\lambda^* = 1)$
$V_1^{LP}(1) = 1$	$V_1^{\lambda^*}(1) = 0$
$V_1^{LP}(2) = 1$	$V_1^{\lambda^*}(2) = -0.4$
$V_2^{LP}(1) = 0$	$V_2^{\lambda^*}(1) = 0$
$V_2^{LP}(2) = 0$	$V_2^{\lambda^*}(2) = 0.4$

For state $(2,1)$, this implies $J^{LP}(2,1;\boldsymbol{\alpha}) = 1$, while $J^{\lambda^*}(2,1;\boldsymbol{\alpha}) = 0.6$.

Appendix C: Proofs of Theorems 2 and 3

Proof of Theorem 2. We proceed by proving (i) ⇒ (iii) ⇒ (ii) ⇒ (i).

First, we show (i) ⇒ (iii). Set $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$ equal to optimal multipliers in the Lagrangian problem, and let $\mathbf{d}(\cdot) \in \bar{A}(\cdot)$ be an optimal policy for the original problem. Fix any $\mathbf{s} \in S$, then

$$\begin{aligned} J^{\boldsymbol{\lambda}^*}(\mathbf{s}) &= J^{\boldsymbol{\lambda}^*,\mathbf{d}}(\mathbf{s}) + J^{\boldsymbol{\lambda}^*}(\mathbf{s}) - J^{\boldsymbol{\lambda}^*,\mathbf{d}}(\mathbf{s}) \\ &= \left[J^{\mathbf{d}}(\mathbf{s}) + W^{\boldsymbol{\lambda}^*,\mathbf{d}}(\mathbf{s}) \right] + J^{\boldsymbol{\lambda}^*}(\mathbf{s}) - J^{\boldsymbol{\lambda}^*,\mathbf{d}}(\mathbf{s}) \\ &= J(\mathbf{s}) + W^{\boldsymbol{\lambda}^*,\mathbf{d}}(\mathbf{s}) + J^{\boldsymbol{\lambda}^*}(\mathbf{s}) - J^{\boldsymbol{\lambda}^*,\mathbf{d}}(\mathbf{s}). \end{aligned}$$

By Lemma 1, condition (i) is equivalent to $J^{\boldsymbol{\lambda}^*}(\mathbf{s}) = J(\mathbf{s})$ for all $\mathbf{s} \in S$. This implies

$$0 = W^{\boldsymbol{\lambda}^*,\mathbf{d}}(\mathbf{s}) + \left(J^{\boldsymbol{\lambda}^*}(\mathbf{s}) - J^{\boldsymbol{\lambda}^*,\mathbf{d}}(\mathbf{s}) \right)$$

Since $W^{\boldsymbol{\lambda}^*,\mathbf{d}}(\mathbf{s}) \ge 0$ and $J^{\boldsymbol{\lambda}^*}(\mathbf{s})$ upper bounds $J^{\boldsymbol{\lambda}^*,\mathbf{d}}(\mathbf{s})$, both terms in the above equation must equal zero. The first term equal to zero implies that $\boldsymbol{\lambda}^{*\mathsf{T}} [\mathbf{b} - \mathbf{D}(\mathbf{s},\mathbf{d}(\mathbf{s}))] = 0$ for all $\mathbf{s} \in S$, while the second term equal to zero implies $J^{\boldsymbol{\lambda}^*}(\mathbf{s}) = J^{\boldsymbol{\lambda}^*,\mathbf{d}}(\mathbf{s})$ (which incidentally implies statement (ii)). Thus,

$$\max_{\mathbf{a} \in A(\mathbf{s})} r(\mathbf{s},\mathbf{a}) + \boldsymbol{\lambda}^{*\mathsf{T}} [\mathbf{b} - \mathbf{D}(\mathbf{s},\mathbf{a})] + \beta \mathsf{E}\left[J^{\boldsymbol{\lambda}^*}(\mathbf{s}')|\mathbf{s},\mathbf{a} \right] = r(\mathbf{s},\mathbf{d}(\mathbf{s})) + \boldsymbol{\lambda}^{*\mathsf{T}} [\mathbf{b} - \mathbf{D}(\mathbf{s},\mathbf{d}(\mathbf{s}))] + \beta \mathsf{E}\left[J^{\boldsymbol{\lambda}^*}(\mathbf{s}')|\mathbf{s},\mathbf{a} \right]$$

Substituting $J(\mathbf{s}) = J^{\boldsymbol{\lambda}^*}(\mathbf{s})$ yields

$$\max_{\mathbf{a}\in A(\mathbf{s})} r(\mathbf{s},\mathbf{a}) + \boldsymbol{\lambda}^{*\mathsf{T}}\left[\mathbf{b} - \mathbf{D}(\mathbf{s},\mathbf{a})\right] + \beta \mathsf{E}\left[J(\mathbf{s}')|\mathbf{s},\mathbf{a}\right] = r(\mathbf{s},\mathbf{d}(\mathbf{s})) + \boldsymbol{\lambda}^{*\mathsf{T}}\left[\mathbf{b} - \mathbf{D}(\mathbf{s},\mathbf{d}(\mathbf{s}))\right] + \beta \mathsf{E}\left[J(\mathbf{s}')|\mathbf{s},\mathbf{a}\right],$$

which is equivalent to equation (12).

To show (iii) $\Rightarrow$ (ii), take the $\mathbf{d}$ in condition (iii) and suppose there existed a state $\mathbf{s} \in S$ and an action $\hat{\mathbf{a}} \in \bar{A}(\mathbf{s})$ with $r(\mathbf{s},\hat{\mathbf{a}}) + \beta \mathsf{E}\left[J(\mathbf{s}')|\mathbf{s},\hat{\mathbf{a}}\right] > r(\mathbf{s},\mathbf{d}(\mathbf{s})) + \beta \mathsf{E}\left[J(\mathbf{s}')|\mathbf{s},\mathbf{d}(\mathbf{s})\right]$. We have $\boldsymbol{\lambda}^{\mathsf{T}}\left[\mathbf{b} - \mathbf{D}(\mathbf{s},\hat{\mathbf{a}})\right] \geq 0$ because $\hat{\mathbf{a}} \in \bar{A}(\mathbf{s})$, and $\boldsymbol{\lambda}^{\mathsf{T}}\left[\mathbf{b} - \mathbf{D}(\mathbf{s},\mathbf{d}(\mathbf{s}))\right] = 0$ by assumption, so $\hat{\mathbf{a}}$ satisfies

$$r(\mathbf{s},\hat{\mathbf{a}}) + \boldsymbol{\lambda}^{\mathsf{T}}\left[\mathbf{b} - \mathbf{D}(\mathbf{s},\hat{\mathbf{a}})\right] + \beta \mathsf{E}\left[J(\mathbf{s}')|\mathbf{s},\hat{\mathbf{a}}\right] > r(\mathbf{s},\mathbf{d}(\mathbf{s})) + \boldsymbol{\lambda}^{\mathsf{T}}\left[\mathbf{b} - \mathbf{D}(\mathbf{s},\mathbf{d}(\mathbf{s}))\right] + \beta \mathsf{E}\left[J(\mathbf{s}')|\mathbf{s},\mathbf{d}(\mathbf{s})\right]$$

This contradicts equation (12), so it must be true that

$$\mathbf{d}(\mathbf{s}) \in \operatorname{argmax}_{\mathbf{a}\in\bar{A}(\mathbf{s})} r(\mathbf{s},\mathbf{a}) + \beta \mathsf{E}\left[J(\mathbf{s}')|\mathbf{s},\mathbf{a}\right].$$

$\mathbf{d}$ is then an optimal policy for the exact problem, so $J^{\mathbf{d}}(\mathbf{s}) = J(\mathbf{s})$ for all $\mathbf{s} \in S$, and equation (12) is then equivalent to

$$\mathbf{d}(\mathbf{s}) \in \operatorname{argmax}_{\mathbf{a}\in A(\mathbf{s})} r(\mathbf{s},\mathbf{a}) + \boldsymbol{\lambda}^{\mathsf{T}}\left[\mathbf{b} - \mathbf{D}(\mathbf{s},\mathbf{a})\right] + \beta \mathsf{E}\left[J^{\mathbf{d}}(\mathbf{s}')|\mathbf{s},\mathbf{a}\right], \qquad \text{for all } \mathbf{s} \in S.$$

Thus $J^{\mathbf{d}}(\cdot)$ solves the optimality equations for the Lagrangian problem with multipliers $\boldsymbol{\lambda}$, so $J^{\boldsymbol{\lambda}}(\mathbf{s}) = J^{\boldsymbol{\lambda},\mathbf{d}}(\mathbf{s})$ for all $\mathbf{s} \in S$, implying $H^{\boldsymbol{\lambda}}(\boldsymbol{\alpha}) = H^{\boldsymbol{\lambda},\mathbf{d}}(\boldsymbol{\alpha})$.

Finally, we show (ii) $\Rightarrow$ (i).

$$\begin{aligned} H^{\boldsymbol{\lambda}^*}(\boldsymbol{\alpha}) &\leq H^{\boldsymbol{\lambda}}(\alpha) \\ &\leq H^{\boldsymbol{\lambda},\mathbf{d}}(\boldsymbol{\alpha}) \\ &= \sum_{\mathbf{s}\in S} \alpha(\mathbf{s})\left(J^{\mathbf{d}}(\mathbf{s}) + W^{\boldsymbol{\lambda},\mathbf{d}}(\mathbf{s})\right) \\ &= \sum_{\mathbf{s}\in S} \alpha(\mathbf{s}) J^{\mathbf{d}}(\mathbf{s}) \\ &\leq H(\boldsymbol{\alpha}) \end{aligned}$$

It is known that $H^{\boldsymbol{\lambda}^*}(\boldsymbol{\alpha}) \geq H(\boldsymbol{\alpha})$, so this tells us $H^{\boldsymbol{\lambda}^*}(\boldsymbol{\alpha}) = H(\boldsymbol{\alpha})$. $\square$

Proof of Theorem 3. We can write the Lagrangian problem for $\boldsymbol{\lambda} = \boldsymbol{\lambda}_{LP}$ as follows, by expressing the original aggregated Lagrangian dynamic program as a linear program:

$$\begin{aligned} H^{\boldsymbol{\lambda}_{LP}}(\boldsymbol{\alpha}) = \min_{V^L} \; & \sum_{\mathbf{s}\in S} \alpha(\mathbf{s}) V^L(\mathbf{s}) \\ \text{s.t.} \; & V^L(\mathbf{s}) \geq r(\mathbf{s},\mathbf{a}) + \boldsymbol{\lambda}_{LP}^{\mathsf{T}}\left[\mathbf{b} - \mathbf{D}(\mathbf{s},\mathbf{a})\right] + \beta \mathsf{E}\left[V^L(\mathbf{s}')|\mathbf{s},\mathbf{a}\right], \text{ for all } \mathbf{s} \in S, \mathbf{a} \in A(\mathbf{s}) \end{aligned} \tag{20}$$

By the theorem condition,

$$\begin{aligned} \max_{\mathbf{a}\in A(\mathbf{s})} r(\mathbf{s},\mathbf{a}) + \boldsymbol{\lambda}_{LP}^{\mathsf{T}}\left[\mathbf{b} - \mathbf{D}(\mathbf{s},\mathbf{a})\right] + \beta \mathsf{E}\left[V^{LP}(\mathbf{s}')|\mathbf{s},\mathbf{a}\right] & \\ = r(\mathbf{s},\mathbf{d}(\mathbf{s})) + \boldsymbol{\lambda}_{LP}^{\mathsf{T}}\left[\mathbf{b} - \mathbf{D}(\mathbf{s},\mathbf{d}(\mathbf{s})\right] + \beta \mathsf{E}\left[V^{LP}(\mathbf{s}')|\mathbf{s},\mathbf{d}(\mathbf{s})\right] & \\ = r(\mathbf{s},\mathbf{d}(\mathbf{s})) + \beta \mathsf{E}\left[V^{LP}(\mathbf{s}')|\mathbf{s},\mathbf{d}(\mathbf{s})\right] & \\ \leq V^{LP}(\mathbf{s}), \text{ for all } \mathbf{s} \in S, & \end{aligned}$$

where the final inequality follows from the feasibility of $V^{LP}(\boldsymbol{\alpha})$ in (LP). Thus, $V^{LP}(\cdot)$ is feasible in the linear program (20), and gives objective value $\sum_{\mathbf{s}\in S} \alpha(\mathbf{s}) V^{LP}(\mathbf{s})$ in the same problem. Thus, $H^{\boldsymbol{\lambda}^*}(\boldsymbol{\alpha}) \leq H^{\boldsymbol{\lambda}_{LP}}(\boldsymbol{\alpha}) \leq H^{LP}(\boldsymbol{\alpha})$. Given that we have $H^{\boldsymbol{\lambda}^*}(\boldsymbol{\alpha}) \geq H^{LP}(\boldsymbol{\alpha})$ in general, we conclude that $H^{\boldsymbol{\lambda}^*}(\boldsymbol{\alpha}) = H^{LP}(\boldsymbol{\alpha})$ here. $\square$

Appendix D: The Restless Bandit Problem

The restless bandit problem, introduced by Whittle (1988) and recently considered by Bertsimas and Niño-Mora (2000), entails controlling a set of I subproblems, or "arms," each of which can be operated at each time stage in either an "active" or "passive" mode. The number of "active" controls at each stage is limited to b. Conditional on whether it is operated in "active" or "passive" mode, each arm generates rewards and transitions to a new state in a Markovian fashion independent of the other arms. The restless bandit problem fits into the framework of weakly coupled dynamic programs with a single linking constraint

$$\sum_i \mathbb{1}\{a_i = \text{"active"}\} \leq b, \tag{21}$$

where we use the symbol $\mathbb{1}\{\cdot\}$ to indicate the binary indicator function. Constraint (21) differs from the classic restless bandit problem of Whittle (1988), which enforces equality. We use the version with "$\leq$" to connect with the rest of our paper, noting that both the Lagrangian and LP-based relaxations can be extended to the equality case. The multiarmed bandit problem is the special case of the restless bandit problem in which "passive" arms are assumed to self-transition and $b=1$. Bertsimas and Niño-Mora (2000) define a family of linear programming relaxations (distinct from the LP-based relaxation we consider here) of the restless bandit problem, and Hawkins (2003) shows that the Lagrangian relaxation is equivalent to Bertsimas and Niño-Mora's "first-order" relaxation.

We have generally observed the Lagrangian and LP-based relaxations to give simiar bounds for restless bandit problems, and we can prove the equivalence of the bounds for the special case of the restless bandit problem with $b=1$ (which includes the multiarmed bandit problem). Let $r_i(s_i,1)$ (respectively, $r_i(s_i,0)$) indicate "active" (respectively, "passive") rewards produced when subproblem i is in state s_i, and let $p_i(s_i'|s_i,1)$ (respectively $p_i(s_i'|s_i,0)$) represent the subsequent state transition probabilities.

THEOREM 7. *For a $b=1$ restless bandit problem with $\lambda^* > 0$, $H^{\lambda^*}(\boldsymbol{\alpha}) = H^{LP}(\boldsymbol{\alpha})$*

Proof. We already have $H^{LP}(\boldsymbol{\alpha}) \leq H^{\lambda^*}(\boldsymbol{\alpha})$ from Corollary 1, so it remains to show $H^{LP}(\boldsymbol{\alpha}) \geq H^{\lambda^*}(\boldsymbol{\alpha})$. For the special case of the restless bandit problem with $b=1$, we can express $H^{LP}(\boldsymbol{\alpha})$ and $H^{\lambda^*}(\boldsymbol{\alpha})$ as

$$\begin{aligned}
(DLP): H^{LP}(\boldsymbol{\alpha}) = \max_{\mathbf{x}} \; & \sum_{\mathbf{s}\in S}\left[\sum_{i=1}^{I} x_{\mathbf{s},i}\left(r_i(s_i,1)+\sum_{j\neq i} r_j(s_j,0)\right) + x_{\mathbf{s},0}\sum_{j=1}^{I} r_j(s_j,0)\right] \\
\text{s.t.} \; & \sum_{\mathbf{s}'\in S: s_i'=s_i}\left(x_{\mathbf{s}',0}+\sum_{j=1}^{I} x_{\mathbf{s}',j}\right) - \beta\sum_{\mathbf{s}'\in S} p_i(s_i|s_i',1)x_{\mathbf{s}',i} \\
& -\beta\sum_{\mathbf{s}'\in S} p_i(s_i|s_i',0)\left(x_{\mathbf{s}',0}+\sum_{j\neq i} x_{\mathbf{s}',j}\right) = \alpha_i(s_i), \quad s_i\in S_i, i\in\{1,\ldots,I\} \\
& x_{\mathbf{s},i}\geq 0, \quad \mathbf{s}\in S, i\in\{1,\ldots,I\} \\
& x_{\mathbf{s},0}\geq 0,
\end{aligned}$$

$$\begin{aligned}
(DL): H^{\lambda^*}(\boldsymbol{\alpha}) = \max_{\mathbf{x}} \; & \sum_{i=1}^{I}\sum_{s_i\in S_i}\left(r_i(s_i,1)x_{s_i1}+r_i(s_i,0)x_{s_i0}\right) \\
\text{s.t.} \; & \sum_{i=1}^{I}\sum_{s_i\in S_i} x_{s_i1} \leq \frac{1}{1-\beta} \\
& x_{s_i0}+x_{s_i1} - \beta\sum_{s_i'\in S_i}\left(p_i(s_i|s_i',1)x_{s_i'1}+p_i(s_i|s_i',0)x_{s_i'0}\right) \\
& \qquad = \alpha_i(s_i), \quad s_i\in S_i, i\in\{1,\ldots,I\} \\
& x_{s_i0}, x_{s_i1}\geq 0, \quad s_i\in S_i
\end{aligned}$$

The problems (DLP) and (DL) are the same as in Sections 2.3 and 2.4.

Take an optimal solution $\{\bar{x}_{s_i1}, \bar{x}_{s_i0}, s_i \in S_i\}$ to (DL), then a solution $\{x_{\mathbf{s},i}, \mathbf{s} \in S, i \in \{1,\dots,I\}\}$ that satisfies

$$\begin{aligned}
\bar{x}_{s_i1} &= \sum_{\mathbf{s}' \in S : s'_i = s_i} x_{\mathbf{s}',i}, \qquad s_i \in S_i \\
\bar{x}_{s_i0} &= \sum_{\mathbf{s}' \in S : s'_i = s_i} \left(x_{\mathbf{s}',0} + \sum_{j \neq i} x_{\mathbf{s}',j} \right), \qquad s_i \in S_i \\
& x_{\mathbf{s},i} \geq 0, \quad \mathbf{s} \in S, i \in \{1,\dots,I\} \\
& x_{\mathbf{s},0} \geq 0
\end{aligned}$$

will be feasible in (DLP) and give optimal objective value equal to $H^{\lambda^*}(\boldsymbol{\alpha})$. Thus, showing that this system is feasible is sufficient for the desired result.

Feasibility of this system is equivalent to boundedness of the program

$$\begin{aligned}
(DF): \max_{\mathbf{u}} \quad & \sum_{i=1}^{I} \sum_{s_i \in S_i} (u_{s_i1} \bar{x}_{s_i1} + u_{s_i0} \bar{x}_{s_i0}) \\
\text{s.t.} \quad & u_{s_i1} + \sum_{j \neq i} u_{s_j0} \leq 0, \qquad \mathbf{s} \in S, i \in \{1,\dots,I\} \\
& \sum_{j=1}^{I} u_{s_j0} \leq 0, \qquad \mathbf{s} \in S
\end{aligned}$$

Because all $\bar{x}_{s_i0}$ and $\bar{x}_{s_i1}$ are greater than or equal to zero, and the constraints of (DF) are the same for each state, boundedness of (DF) is equivalent to boundedness of the following program

$$\begin{aligned}
(DF'): \max_{\mathbf{u}} \quad & \sum_{i=1}^{I} \left(u_{i1} \sum_{s_i \in S_i} \bar{x}_{s_i1} + u_{i0} \sum_{s_i \in S_i} \bar{x}_{s_i0} \right) \\
\text{s.t.} \quad & u_{i1} + \sum_{j \neq i} u_{j0} \leq 0, \qquad i \in \{1,\dots,I\} \\
& \sum_{j=1}^{I} u_{j0} \leq 0
\end{aligned}$$

This is equivalent to feasibility of

$$\begin{aligned}
w_i &= \sum_{s_i \in S_i} \bar{x}_{s_i1}, \qquad i \in \{1,\dots,I\} \\
z + \sum_{j \neq i} w_j &= \sum_{s_i \in S_i} \bar{x}_{s_i0}, \qquad i \in \{1,\dots,I\} \\
w_i &\geq 0, \quad i \in \{1,\dots,I\} \\
z &\geq 0.
\end{aligned}$$

The solution $z = 0$ and $w_i = \sum_{s_i \in S_i} \bar{x}_{x_i1} \geq 0$ is feasible because $z + \sum_{j \neq i} w_j = \sum_{j \neq i} \sum_{s_j \in S_j} \bar{x}_{s_j1} = \frac{1}{1-\beta} - \sum_{s_i \in S_i} \bar{x}_{s_i1} = \sum_{s_i \in S_i} \bar{x}_{s_i0}$, where the second inequality follows from complementary slackness (given that $\lambda^* > 0$) applied to the first constraint of (DL), and the third inequality follows by adding the second set of constraints in (DL) over S_i for any i. $\square$

The restless bandit problem (with arbitrary b) represents a class of problems for which we can generate columns particularly easily in the column generation algorithm presented in Section 5

for computing the LP-based bound. In the case of the restless bandit problem, problem (18) is equivalent to

$$
\begin{aligned}
\max_{\mathbf{s},\mathcal{J}} \quad & \sum_{i\in\mathcal{J}} f_i(s_i, \text{“active”}) + \sum_{i\notin\mathcal{J}} f_i(s_i, \text{“passive”}) \\
\text{s.t.} \quad & \mathbf{s}\in S, \mathcal{J}\subseteq\{1,\ldots,I\}, |\mathcal{J}|\le b \\
&= \max_{\mathcal{J}:|\mathcal{J}|\le b}\left\{\sum_{i\in\mathcal{J}}\max_{s_i\in S_i} f_i(s_i, \text{“active”}) + \sum_{i\notin\mathcal{J}}\max_{s_i\in S_i} f_i(s_i, \text{“passive”})\right\} \\
&= \max_{\mathcal{J}:|\mathcal{J}|\le b}\left\{\sum_{i\in\mathcal{J}}\max_{s_i\in S_i} f_i(s_i, \text{“active”}) + \sum_{i\notin\mathcal{J}}\max_{s_i\in S_i} f_i(s_i, \text{“passive”})\right. \\
&\qquad\qquad \left. + \sum_{i\in\mathcal{J}}\max_{s_i\in S_i} f_i(s_i, \text{“passive”}) - \sum_{i\in\mathcal{J}}\max_{s_i\in S_i} f_i(s_i, \text{“passive”})\right\} \\
&= \max_{\mathcal{J}:|\mathcal{J}|\le b}\left\{\sum_{i\in\{1,\ldots,I\}}\max_{s_i\in S_i} f_i(s_i, \text{“passive”})\right. \\
&\qquad\qquad \left. + \sum_{i\in\mathcal{J}}\left[\max_{s_i\in S_i} f_i(s_i, \text{“active”}) - \max_{s_i\in S_i} f_i(s_i, \text{“passive”})\right]\right\} \\
&= \sum_{i\in\{1,\ldots,I\}}\max_{s_i\in S_i} f_i(s_i, \text{“passive”}) \\
&\qquad\qquad + \max_{\mathcal{J}:|\mathcal{J}|\le b}\left\{\sum_{i\in\mathcal{J}}\left[\max_{s_i\in S_i} f_i(s_i, \text{“active”}) - \max_{s_i\in S_i} f_i(s_i, \text{“passive”})\right]\right\}.
\end{aligned}
$$

This implies that in the restless bandit case, problem (18) can be solved simply by ranking the subproblems by the quantity $[\max_{s_i} f_i(s_i, \text{“active”}) - \max_{s_i} f_i(s_i, \text{“passive”})]$, assigning the “active” action to subproblems in a greedy fashion, then choosing subproblem states s_i maximizing the reward given the assigned action. Thus, column generation in the restless bandit case can be accomplished extremely efficiently.

Appendix E: Bertsekas' Duality Gap Result

In this appendix, we adapt to our setting the duality gap result in Proposition 5.26 of Bertsekas (1982).

Fix subproblem i and subproblem state s_i. We first define a function that returns a scalar measuring the lack of convexity in $r_i(s_i,\cdot)$ viewed as a function of the action $a_i \in A_i(s_i)$. For simplicity we will assume that the finite set $A_i(s_i)$ has dimension 1. First, define

$$
\tilde{r}_i(s_i,\tilde{a}_i) = \inf\left\{\gamma r_i(s_i,a_i^1) + (1-\gamma) r_i(s_i,a_i^2) : \tilde{a}_1 = \gamma a_i^1 + (1-\gamma)a_i^2; a_i^1, a_i^2 \in A_i(s_i); 0\le\gamma\le 1\right\}
$$

for all $\tilde{a}_i \in \text{conv}(A_i(s_i))$, where $\text{conv}(A_i(s_i))$ is the convex hull of $A_i(s_i)$. Similarly, define

$$
\tilde{\mathbf{D}}_i(s_i,\tilde{a}_i) = \inf\left\{\gamma \mathbf{D}_i(s_i,a_i^1) + (1-\gamma)\mathbf{D}_i(s_i,a_i^2) : \tilde{a}_1 = \gamma a_i^1 + (1-\gamma)a_i^2; a_i^1, a_i^2 \in A_i(s_i); 0\le\gamma\le 1\right\}. \tag{22}
$$

for all $\tilde{a}_i \in \text{conv}(A_i(s_i))$, where the infimum is taken separately for each of the N components of the function $\mathbf{D}_i$. These can be viewed as “convexified” versions of the underlying functions defined on the convex hull of their domain.

Lastly, define

$$
\hat{r}_i(s_i,\tilde{a}_i) = \inf\left\{r_i(s_i,a_i) : \mathbf{D}_i(s_i,a_i) \le \tilde{\mathbf{D}}_i(s_i,\tilde{a}_i); a_i \in A_i(s_i)\right\}
$$

for all $\tilde{a}_i \in \text{conv}(A_i(s_i))$. Then a measure of non-convexity is

$$\rho_i(r_i(s_i,\cdot)) = \sup\left\{\hat{r}_i(s_i,\tilde{a}_i) - \tilde{r}_i(s_i,\tilde{a}_i) : \tilde{a}_i \in \text{conv}(A_i(s_i))\right\},$$

for all subproblems i and subproblem states s_i. Because we have, for all $\tilde{a}_i \in \text{conv}(A_i(s_i))$,

$$\hat{r}_i(s_i,\tilde{a}_i) \leq \sup\{r_i(s_i,a_i) : a_i \in A_i(s_i)\} \qquad \text{and} \qquad \tilde{r}_i(s_i,\tilde{a}_i) \geq \inf\{r_i(s_i,a_i) : a_i \in A_i(s_i)\},$$

it follows that

$$\rho_i(r_i(s_i,\cdot)) \leq \sup\{r_i(s_i,a_i) : a_i \in A_i(s_i)\} - \inf\{r_i(s_i,a_i) : a_i \in A_i(s_i)\}.$$

So if $r_i(s_i,\cdot)$ is taken from a bounded set, then $\rho_i(r_i(s_i,\cdot))$ is bounded.

Now consider the optimization problem (P($\mathbf{s}$)) and its Lagrangian dual (D($\mathbf{s}$)), for a fixed state $\mathbf{s}$. Consider the following assumptions

ASSUMPTION 1. *For every state* $\mathbf{s}$, $\bar{A}(\mathbf{s}) \neq \emptyset$.

ASSUMPTION 2. *For each subproblem* i *and subproblem state* s_i,

$$\{(a_i, \mathbf{D}_i(s_i,a_i), r_i(s_i,a_i)) : a_i \in A_i(s_i)\}$$

is compact.

ASSUMPTION 3. *For each subproblem* i, *given any* $\tilde{a}_i \in conv(A_i(s_i))$, *there exists an* $a_i \in A_i(s_i)$ *such that* $\mathbf{D}_i(s_i,a_i) \leq \tilde{\mathbf{D}}^i(s_i,\tilde{a}_i)$.

This last assumption can be shown to hold when the infimum in (22) is attained. In our case, all three assumptions hold, the latter two because $A_i(s_i)$ is a finite set.

THEOREM 8 **(Bertsekas (1982), Proposition 5.26)**. *Fix state* $\mathbf{s} \in S$. *Under the three assumptions above, there holds*

$$\inf D(\mathbf{s}) - \sup P(\mathbf{s}) \leq (N+1)\mathcal{E}(\mathbf{s}),$$

where $\mathcal{E}(\mathbf{s}) = \max\{\rho_i(r_i(s_i,\cdot)) : i \in \{1,\ldots,I\}\}$.

References

Bertsekas, D. P. 1982. *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, New York.

Bertsimas, D., J. Niño-Mora. 2000. Restless bandits, linear programmming relaxations, and a primal-dual index heuristic. *Oper. Res.* **48**(1) 80–90.

Hawkins, J. 2003. A Lagrangian decomposition approach to weakly coupled dynamic optimization problems and its applications. Ph.D. thesis, Operations Research Center, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Puterman, M. 1994. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley, New York.

Whittle, P. 1988. Restless bandits: Activity allocation in a changing world. *A Celebration of Applied Probability. J. Appl. Probab.* **25A** 287–298.