Structural Results (updated 05/04/2016)

This document includes following structural results and corresponding proofs for the Dynamic Priority Assignment research.

- type independence of optimal action
- optimality of threshold-type policy
- optimality of monotone policy
- structural properties of the value function

Type Independence of Optimal Action

When we compute optimal action by backward induction to maximize total expected reward, we don't need to enumerate all the possible action combinations for all call types, which is as many as 2^m when there are m call types. Instead, **we can directly compare total expected reward from taking $a^i = 0$ and total expected reward from taking $a^i = 1$, and choose the one that gives larger reward, while fixing actions for all other types as an arbitrary value**.
To be specific, the value at time epoch t and state $s = (s^A, s^B)$ with action $a = (a^1, \ldots, a^m)$ is a sum of 3 different terms:

$$\begin{aligned} V_t(s,a) = R(s,a) \\ &+ P(s^A+1, s^B|s,a)V_{t+1}(s^A+1, s^B) \\ &+ P(s^A, s^B+1|s,a)V_{t+1}(s^A, s^B+1) \\ &+ P(s^A-1, s^B|s,a)V_{t+1}(s^A-1, s^B) \\ &+ P(s^A, s^B-1|s,a)V_{t+1}(s^A, s^B-1) \\ &+ P(s^A, s^B|s,a)V_{t+1}(s^A, s^B) \end{aligned}$$

Here we have two kinds of terms involving action: reward and transition probability.

1. difference in reward function

$$\begin{aligned} R(s, a^1, \ldots, a^{k-1}, 0, a^{k+1}, \ldots, a^m) - R(s, a^1, \ldots, a^{k-1}, 0, a^{k+1}, \ldots, a^m) \\ = \lambda^{k2}(\alpha^{k2}(U_{HA}f^H(s^A) - U_{HB}f^H(s^B)) + (1-\alpha^{k2})(U_{LA}f^L(s^A) - U_{LB}f^L(s^B)))/\Lambda \end{aligned}$$

2. difference in transition probability

$$P(j|s,a^1,\dots,a^{k-1},0,a^{k+1},\dots,a^m) - P(j|s,a^1,\dots,a^{k-1},1,a^{k+1},\dots,a^m)$$
$$= \begin{cases} \lambda^{k2}/\Lambda & \text{if } j=(s^A+1,s^B) \\ -\lambda^{k2}/\Lambda & \text{if } j=(s^A,s^B+1) \\ 0 & \text{if } j=(s^A-1,s^B) \\ 0 & \text{if } j=(s^A,s^B-1) \\ I(s^A=N^A)\lambda^{k2}/\Lambda - I(s^B=N^B)\lambda^{k2}/\Lambda & \text{if } j=(s^A,s^B) \\ 0 & \text{otherwise.} \end{cases}$$

Hence we have

$$V_t(s,a^1,\dots,a^{k-1},0,a^{k+1},\dots,a^m) - V_t(s,a^1,\dots,a^{k-1},1,a^{k+1},\dots,a^m)$$
$$= \lambda^{k2}(\alpha^{k2}(U_{HA}f^H(s^A) - U_{HB}f^H(s^B)) + (1-\alpha^{k2})(U_{LA}f^L(s^A) - U_{LB}f^L(s^B)))/\Lambda$$
$$+ \lambda^{k2}(V_{t+1}(s^A,s^B+1) - V_{t+1}(s^A+1,s^B))/\Lambda$$

Therefore, the difference between taking $a^k=0$ and $a^k=1$ is independent of actions taken for other types $i \in \{1,\dots,k-1,k+1,\dots,m\}$.
Therefore in order to determine optimal action for given state and time, we can fix actions for all other types by arbitrary choices and make a decision only for one call type. Therefore we only need m times of comparison, which significantly reduces computational effort.

Justification We are not making the combinatorial decision because there are always less than 1 call that arrives per each time epoch (per each decision point) due to poisson arrival and uniformization. Therefore, when we look into a specific call type to make a decision, actions chosen for other call types is not considered, because they have no effect to future status change, as we are guaranteed to have no more than 1 call before the next decision point.

Optimality of Threshold-Type Policy

We want to derive the condition that we have an optimal threshold policy in terms of accuracy of a call type(α^i).
To rephrase, given the time epoch t and state s, we can claim that

- *If $\alpha_i > \alpha_j$, we are always more likely to send an ALS server to type i than type j*
- *We can find a threshold accuracy value $\bar{\alpha}_t(s)$ such that we send the ALS server for all call types with $\alpha^i > \bar{\alpha}_t(s)$ and we send the BLS server for all call types with $\alpha^i < \bar{\alpha}_t(s)$.*

under a specific condition that will derived below.

Let a^{-i} denote $(a^1, \dots, a^{i-1}, a^{i+1}, \dots, a^m)$, *i.e.* set of actions for all types but type i. By type independence argument, $a^i = 0$ is optimal if and only if

$$0 < V_t(s, 0, a^{-i}) - V_t(s, 1, a^{-i}).$$

Then we can substitute the cost-to-go function with its definition, arrange terms, and divide both sides by λ^{j2}/Λ which is assumed to be strictly positive. Then we get

$$\begin{aligned} 0 &< \alpha^j (U_{HA} f(s^A) - U_{HB} f(s^B)) + (1 - \alpha^j)(U_{LA} f(s^A) - U_{LB} f(s^B)) \\ &+ V_{t+1}(s^A + 1, s^B) - V_{t+1}(s^A, s^B + 1). \end{aligned}$$

Now let's suppose that

$$U_{HA} f^H(s^A) - U_{HB} f^H(s^B) - U_{LA} f^L(s^A) + U_{LB} f^L(s^B) > 0. \tag{1}$$

which is true when (1) $N^A - s^A \geq N^B - s^B$ or (2) $U_{H0} >> U_{H1}$. Then we can arrange terms again to get that $a_t^j = 0$ if and only if

$$\begin{aligned} \alpha^j &\geq \frac{U_{LA} f^L(s^A) - V_{LB} f^L(s^B) - V_{t+1}(s^A + 1, s^B) + V_{t+1}(s^A, s^B + 1)}{U_{HA} f^H(s^A) - U_{HB} f^H(s^B) - U_{LA} f^L(s^A) + U_{LB} f^L(s^B)} \\ &= \bar{\alpha}_t(s). \end{aligned}$$

This is exactly the threshold-type policy we seek.

Justification We are not making the combinatorial decision because there are always less than 1 call that arrives per each time epoch (per each decision point) due to poisson arrival and uniformization. Therefore, the frequency of a call type arrival has no effect to future status change, as we are guaranteed to have no more than 1 call before the next decision point.

Optimality of Control-limit Policy

We derive a condition that there exist an optimal control-limit policy in terms of system state.

To rephrase, given the time epoch t and call type i, we claim that for each time epoch t,

- *If* $a_t^i(s^A, s^B) = 0$*, then* $a_t(j, s^B) = 0$ *for any* $j < s^A$*.*
- *We can find a threshold number* $\bar{s}_t^i$ *such that we send the ALS server to call type* i *if and only if number of busy ALS servers in the system is less than* $\bar{s}_A^i$*.*

Analogously for BLS server, we claim that

- *If* $a_t^i(s^A, s^B) = 1$*, then* $a_t(s^A, j) = 1$ *for any* $j < s^B$*.*
- *We can find a threshold number* $\bar{s}_B^i$ *such that we send the BLS server to call type* i *if and only if number of busy BLS servers in the system is less than* $\bar{s}_t^i$*.*

This is equivalent to the claim that there exist an optimal policy that is monotone nondecreasing in s^A and monotone nonincreasing in s^B. It is intuitive as well as is convenient to know the conditions under which we have an optimal control-limit policy in terms of number of busy ALS and BLS servers.

Begin with the number of busy servers, s^A. If $a_t^i(s^A, s^B) = 0$ assures that $a_t^i(s^A - 1, s^B) = 0$, then by induction, we can claim that we have an optimal control-limit policy in terms of number of busy ALS servers.
We've seen that $a_t^i(s^A, s^B) = 0$ is equivalent to

$$\begin{aligned} V_{t+1}(s^A, s^B + 1) - V_{t+1}(s^A + 1, s^B) < \alpha^{i2} f^H(s^A) U_{HA} + (1 - \alpha^{i2}) f^L(s^A) U_{LA} \\ - \alpha^{i2} f^H(s^B) U_{HB} + (1 - \alpha^{i2}) f^L(s^B) U_{LB}. \end{aligned} \tag{2}$$

And by optimality of threshold policy, if above inequality is true, then the following inequality should be true:

$$\begin{aligned} V_{t+1}(s^A - 1, s^B + 1) - V_{t+1}(s^A, s^B) < \alpha^{i2} f^H(s^A - 1) U_{HA} + (1 - \alpha^{i2}) f^L(s^A - 1) U_{LA} \\ - \alpha^{i2} f^H(s^B) U_{HB} + (1 - \alpha^{i2}) f^L(s^B) U_{LB}. \end{aligned} \tag{3}$$

which is equivalent to $a_t^i(s^A - 1, s^B) = 0$.
We seek for a sufficient condition that makes (2) implies (3). Since $f^H(\cdot)$ and $f^L(\dot{})$ are monotonic nonincreasing, RHS of (2) is small than RHS of (3). Therefore, if we can argue that LHS of (2) is greater than LHS of (3)

$$V_{t+1}(s^A + 1, s^B) - V_{t+1}(s^A, s^B + 1) - V_{t+1}(s^A, s^B) + V_{t+1}(s^A - 1, s^B + 1) \leq 0 \tag{4}$$

then we can claim that (1) implies (2), and this could be a sufficient condition for the optimality of threshold policy.

Now we suggest two sufficient conditions that validates (4) altogether:

$$V_{t+1}(s^A + 1, s^B) - V_{t+1}(s^A, s^B) - V_{t+1}(s^A, s^B) + V_{t+1}(s^A - 1, s^B) \leq 0 \tag{5}$$

$$V_{t+1}(s^A, s^B) - V_{t+1}(s^A, s^B + 1) - V_{t+1}(s^A - 1, s^B) + V_{t+1}(s^A - 1, s^B + 1) \leq 0 \tag{6}$$

that are equivalently

$V_t(s)$ is concave in s^A; the amount of increase in total expected reward from having one more ALS server available is greater when we have more ALS servers available in the system.

$V_t(s)$ is supermodular on a lattice $S = S^A \times S^B$; the amount of increase in total expected reward from having one more ALS server available is greater when we have more BLS servers available in the system. (Plus, we can say analogously for BLS; gain in reward by having one more BLS server available is greater when we have more ALS servers available in the system)

Both conditions are proven as structural properties of $V_t(s^A, s^B)$, as long as we have concave $f^H(\cdot)$ and $f^L(\cdot)$. □

Structural Properties of Value Functions

1. $V_t(s)$ is nonincreasing in s^A.

Before the proof, here's a list of prelimilary results:

(a) $R_t(s^A, s^B, a)$ is **nonincreasing** in s^A for all s^B, a and t.
Proof For any $s^A < N^A$, we have

$$\begin{aligned}
0 \le & R_t(s^A+1, s^B, a) - R_t(s^A, s^B, a) \\
&= \left(f^H(s^A+1) - f^H(s^A)\right) \cdot \sum_i \left(\lambda_t^{i1} U_{HA} + \lambda_t^{i2}(1-a^i)\alpha^i U_{HA}\right) \\
&\quad + \left(f^L(s^A+1) - f^L(s^A)\right) \cdot \sum_i (1-a^i)\alpha^i U_{LA}
\end{aligned}$$

since terms inside the sum are positive and $f^H(\cdot)$, $f^L(\cdot)$ are monotone nonincreasing.

(b) $\sum_{j^A=k}^{\infty} P_t(j^A, j^B | s^A, s^B, a)$ is nondecreasing in s^A for all k, j^B, s^B, a and t.

Proof

$$\sum_{j^A=k}^{\infty} P_t(j^A, j^B | s^A+1, s^B, a) - \sum_{j^A=k}^{\infty} P_t(j^A, j^B | s^A, s^B, a)$$

$$= \begin{cases}
\sum_i(\lambda_t^{i1} + \lambda_t^{i2}(1-a^i)/\Lambda & \text{if } (k, j^B) = (s^A+2, s^B) \\
\sum_i(\lambda_t^{i3} + \lambda_t^{i2} a^i)/\Lambda & \text{if } (k, j^B) = (s^A+1, s^B+1) \\
1 - ((s^A+1)\tau^A + s^B\tau^B + \sum_i \sum_p \lambda_t^{iP})/\Lambda & \\
+I(s^A+1 = N^A)\sum_i(\lambda_t^{i1} + \lambda_t^{i2}(1-a^i)/\Lambda & \\
+I(s^B = N^B)\sum_i(\lambda_t^{i3} + \lambda_t^{i2} a^i)/\Lambda & \text{if } (k, j^B) = (s^A+1, s^B) \\
(s^A+1)\tau^A/\Lambda & \text{if } (k, j^B) = (s^A, s^B) \\
s^B\tau^B/\Lambda & \text{if } (k, j^B) = (s^A+1, s^B-1) \\
0 & \text{otherwise.}
\end{cases}$$

(c) $R_T(s^A, s^B)$ is nonincreasing in s^A: true as we assume zero salvage value.

Now, we will prove the monotonicity of $V_t(s^A, s^B)$ by mathematical induction. By (c) above, the result holds for $t = T$. Assume now that $U_{n+1}(s^A, s^B)$ is nonincreasing for $t = n+1, \ldots, T$ in s^A. Then by definition we have

$$\begin{aligned}
V_n(s^A, s^B) &= \max_a \left\{ R_n(s^A, s^B, a) + \sum_{j^B} \sum_{j}^{A} P_n(j^A, j^B | s^A, s^B, a) V_{n+1}(j^A, j^B) \right\} \\
&= R_n(s^A, s^B, a^*) + \sum_{j^B} \sum_{j^A} P_n(j^A, j^B | s^A, s^B, a^*) V_{n+1}(j^A, j^B) \\
&\le R_n(s^A - 1, s^B, a^*) + + \sum_{j^B} \sum_{j^A} P_n(j^A, j^B | s^A - 1, s^B, a^*) V n + 1(j^A, j^B) \\
&\le V_n(s^A - 1, s^B).
\end{aligned}$$

The inequality of the third line is true by (a),(b) and assumption on nonincreasing $V_{n+1}(s)$. Thus $V_n(s^A, s^B)$ is nonincreasing in s^A, the induction hypothesis is satisfied, and the result follows.

2. $V_t(s)$ is **concave(convex)** in s^A and s^B if $f^H(\cdot)$ and $f^L(\cdot)$ are concave(convex).

We show for the concavity only, and the proof for the convexity is just analogous. Hence let's assume that $f^H(\cdot)$ and $f^L(\cdot)$ are concave from here.

First we need our reward function to be concave in s^A for all a:

$$
\begin{aligned}
0 &\geq R_t(s^A+1, s^B|a) - R_t(s^A, s^B|a) - R_t(s^A, s^B|a) - R_t(s^A-1, s^B|a) \\
&= \big(f^H(s^A+1) + f^H(s^A-1) - 2f^H(s^A)\big) \cdot \sum_i \big(\lambda_t^{i1} U_{HA} + \lambda_t^{i2}(1-a^i)\alpha^i U_{HA}\big)/\Lambda \\
&+ \big(f^L(s^A+1) + f^L(s^A-1) - 2f^L(s^A)\big) \cdot \sum_i \lambda_t^{i2}(1-a^i)(1-\alpha^i)U_{LA}/\Lambda
\end{aligned}
$$

Two terms in the summation are all nonnegative, so a direct sufficient condition for this inequality to be true is concavity of $f^H(\cdot)$ and $f^L(\cdot)$.

Next, we need a stochastic version of convexity for the remaining terms: $\sum_j P_t(j|s,a)V_t(j)$. We can show that

$$
\sum_{j^B \in S^B} \sum_{j^A \in S^A} P_t(j^A, j^B | s^A, s^B, a) V_t(j^A, j^B)
$$

is convex in s^A for all a, given that $V_t(j^A, j^B)$ is nonincreasing and convex in s^A. Here we assume that we already have convex reward function. We will use mathematical induction. At $t = T$, the argument is clearly true as we assume zero salvage value. Assume true for $t+1$. Then for t, after arranging terms, we get

$$
\begin{aligned}
&\sum_{j^B \in S^B} \sum_{j^A \in S^A} \Big(P_t(j^A, j^B|s^A+1, s^B, a)V_t(j^A, j^B) - P_t(j^A, j^B|s^A, s^B, a)V_t(j^A, j^B)\Big) \\
&- \sum_{j^B \in S^B} \sum_{j^A \in S^A} \Big(P_t(j^A, j^B|s^A, s^B, a)V_t(j^A, j^B) - P_t(j^A, j^B|s^A-1, s^B, a)V_t(j^A, j^B)\Big)
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_i \big(\lambda_t^{i1} + \lambda_t^{i2}(1-a^i)\big)\big(V_{t+1}(s^A+2, s^B) - 2V_{t+1}(s^A+1, s^B) + V_{t+1}(s^A, s^B)\big)/\Lambda \\
&+ \sum_i \big(\lambda_t^{i3} + \lambda_t^{i2}a^i\big)\big(V_{t+1}(s^A+1, s^B+1) - 2V_{t+1}(s^A, s^B+1) + V_{t+1}(s^A-1, s^B+1)\big)/\Lambda \\
&+ \big(\Lambda - s^A\tau^A - s^B\tau^B - \sum_i \sum_p \lambda^{ip}\big) \\
&\cdot \big(V_{t+1}(s^A+1, s^B) - 2V_{t+1}(s^A, s^B) + V_{t+1}(s^A-1, s^B)\big)/\Lambda \\
&+ s^A\tau^A\big(V_{t+1}(s^A, s^B) - 2V_{t+1}(s^A-1, s^B) + V_{t+1}(s^A-2, s^B)\big)/\Lambda \\
&+ s^B\tau^B\big(V_{t+1}(s^A+1, s^B-1) - 2V_{t+1}(s^A, s^B-1) + V_{t+1}(s^A-1, s^B-1)\big)/\Lambda \\
&+ \tau^A\big(-V_{t+1}(s^A+1, s^B) + V_{t+1}(s^A-1, s^B) + V_{t+1}(s^A, s^B) - V_{t+1}(s^A-2, s^B)\big)/\Lambda
\end{aligned}
$$

$$
\begin{aligned}
=&\sum_i \big(\lambda_t^{i1}+\lambda_t^{i2}(1-a^i)\big)\big(V_{t+1}(s^A+2,s^B)-2V_{t+1}(s^A+1,s^B)+V_{t+1}(s^A,s^B)\big)/\Lambda\\
+&\sum_i \big(\lambda_t^{i3}+\lambda_t^{i2}a^i\big)\big(V_{t+1}(s^A+1,s^B+1)-2V_{t+1}(s^A,s^B+1)+V_{t+1}(s^A-1,s^B+1)\big)/\Lambda\\
+&\big(\Lambda-s^A\tau^A-s^B\tau^B-\sum_i\sum_p\lambda^{ip}\big)\\
&\cdot\big(V_{t+1}(s^A+1,s^B)-2V_{t+1}(s^A,s^B)+V_{t+1}(s^A-1,s^B)\big)/\Lambda\\
+&(s^A-1)\tau^A\big(V_{t+1}(s^A,s^B)-2V_{t+1}(s^A-1,s^B)+V_{t+1}(s^A-2,s^B)\big)/\Lambda\\
+&s^B\tau^B\big(V_{t+1}(s^A+1,s^B-1)-2V_{t+1}(s^A,s^B-1)+V_{t+1}(s^A-1,s^B-1)\big)/\Lambda\\
+&\tau^A\big(-V_{t+1}(s^A+1,0)+2V_{t+1}(s^A,s^B)-V_{t+1}(s^A-1,s^B)\big)/\Lambda
\end{aligned}
$$

$$
\begin{aligned}
=&\sum_i \big(\lambda_t^{i1}+\lambda_t^{i2}(1-a^i)\big)\big(V_{t+1}(s^A+2,s^B)-2V_{t+1}(s^A+1,s^B)+V_{t+1}(s^A,s^B)\big)/\Lambda\\
+&\sum_i \big(\lambda_t^{i3}+\lambda_t^{i2}a^i\big)\big(V_{t+1}(s^A+1,s^B+1)-2V_{t+1}(s^A,s^B+1)+V_{t+1}(s^A-1,s^B+1)\big)/\Lambda\\
+&\big(\Lambda-(s^A+1)\tau^A-s^B\tau^B-\sum_i\sum_p\lambda^{ip}\big)\\
&\cdot\big(V_{t+1}(s^A+1,s^B)-2V_{t+1}(s^A,s^B)+V_{t+1}(s^A-1,s^B)\big)/\Lambda\\
+&(s^A-1)\tau^A\big(V_{t+1}(s^A,s^B)-2V_{t+1}(s^A-1,s^B)+V_{t+1}(s^A-2,s^B)\big)/\Lambda\\
+&s^B\tau^B\big(V_{t+1}(s^A+1,s^B-1)-2V_{t+1}(s^A,s^B-1)+V_{t+1}(s^A-1,s^B-1)\big)/\Lambda\\
\geq&0,
\end{aligned}
$$

because we initially assumed that $V_{t+1}(s^A,s^B)$ is concave in s^A.

3. $V_t(s)$ is modular on $S^A\times S^B$.

   First, we need the reward function be modular on $S^A\times S^B$.

$$
\begin{aligned}
&R_t(s^A+1,s^B+1)-R_t(s^A,s^B+1)-R_t(s^A+1,s^B)+R_t(s^A,s^B)\\
=&\big(f^H(s^A+1)-f^H(s^A)-f^H(s^A+1)+f^H(s^A)\big)\cdot\sum_i\big(\lambda_t^{i1}+\lambda_t^{i2}(1-a^i)\alpha^i\big)U_{HA}/\Lambda\\
+&\big(f^L(s^A+1)-f^L(s^A)-f^L(s^A+1)+f^L(s^A)\big)\cdot\sum_i\lambda_t^{i2}(1-a^i)(1-\alpha^i)U_{LA}/\Lambda\\
+&\big(f^H(s^B+1)-f^H(s^B)-f^H(s^B+1)+f^H(s^B)\big)\cdot\sum_i\lambda_t^{i2}a^i\alpha^iU_{HB}/\Lambda\\
+&\big(f^L(s^B+1)-f^L(s^B)-f^L(s^B+1)+f^L(s^B)\big)\cdot\sum_i\big(\lambda_t^{i3}+\lambda_t^{i2}a^i\alpha^i\big)U_{LB}/\Lambda\\
=&0.
\end{aligned}
$$

   Next we show a stochastic version of modularity for the remaining terms

$$\sum_j P_t(j^A,j^B|s^A,s^B,a)V_t(j^A,j^B).$$

The proof is very similar to above concavity case; we rely on mathematical induction framework, assuming that $V_{t+1}(s^A, s^B)$ is modular and aiming to prove that $V_t(s^A, s^B)$ is modular.

$$\sum_j \big(P_t(j^A, j^B|s^A+1, s^B+1, a) - P_t(j^A, j^B|s^A, s^B+1, a)\big)V_t(j^A, j^B)$$

$$-\sum_j \big(P_t(j^A, j^B|s^A+1, s^B, a) - P_t(j^A, j^B|s^A, s^B, a)\big)V_t(j^A, j^B)$$

$$
\begin{aligned}
=&\sum_i \big(\lambda_t^{i1} + \lambda_t^{i2}(1-a^i)\big)/\Lambda \\
&\cdot\big(V_{t+1}(s^A+1, s^B+1) - V_{t+1}(s^A, s^B+1) - V_{t+1}(s^A+1, s^B) + V_{t+1}(s^A, s^B)\big) \\
&+\sum_i \big(\lambda_t^{i3} + \lambda_t^{i2}a^i\big)/\Lambda \\
&\cdot\big(V_{t+1}(s^A+1, s^B+2) - V_{t+1}(s^A, s^B+2) - V_{t+1}(s^A+1, s^B+1) + V_{t+1}(s^A, s^B+1)\big) \\
&+\big(\Lambda - s^A\tau^A - s^B\tau^B - \sum_i\sum_p \lambda^{ip}\big)/\Lambda \\
&\cdot\big(V_{t+1}(s^A+1, s^B+1) - V_{t+1}(s^A, s^B+1) - V_{t+1}(s^A+1, s^B) + V_{t+1}(s^A, s^B)\big) \\
&+s^A\tau^A\big(V_{t+1}(s^A, s^B+1) - V_{t+1}(s^A-1, s^B+1) - V_{t+1}(s^A, s^B) + V_{t+1}(s^A-1, s^B)\big)/\Lambda \\
&+s^B\tau^B\big(V_{t+1}(s^A+1, s^B) - V_{t+1}(s^A, s^B) - V_{t+1}(s^A+1, s^B-1) + V_{t+1}(s^A, s^B-1)\big)/\Lambda \\
&+\tau^A\big(-V_{t+1}(s^A+1, s^B+1) + V_{t+1}(s^A+1, s^B) + V_{t+1}(s^A, s^B+1) - V_{t+1}(s^A, s^B)\big)/\Lambda \\
&+\tau^B\big(-V_{t+1}(s^A+1, s^B+1) + V_{t+1}(s^A+1, s^B) + V_{t+1}(s^A, s^B+1) - V_{t+1}(s^A, s^B)\big)/\Lambda
\end{aligned}
$$

$$
\begin{aligned}
=&\sum_i \big(\lambda_t^{i1} + \lambda_t^{i2}(1-a^i)\big)/\Lambda \\
&\cdot\big(V_{t+1}(s^A+1, s^B+1) - V_{t+1}(s^A, s^B+1) - V_{t+1}(s^A+1, s^B) + V_{t+1}(s^A, s^B)\big) \\
&+\sum_i \big(\lambda_t^{i3} + \lambda_t^{i2}a^i\big)/\Lambda \\
&\cdot\big(V_{t+1}(s^A+1, s^B+2) - V_{t+1}(s^A, s^B+2) - V_{t+1}(s^A+1, s^B+1) + V_{t+1}(s^A, s^B+1)\big) \\
&+\big(\Lambda - (s^A+1)\tau^A - (s^B+1)\tau^B - \sum_i\sum_p \lambda^{ip}\big)/\Lambda \\
&\cdot\big(V_{t+1}(s^A+1, s^B+1) - V_{t+1}(s^A, s^B+1) - V_{t+1}(s^A+1, s^B) + V_{t+1}(s^A, s^B)\big) \\
&+s^A\tau^A\big(V_{t+1}(s^A, s^B+1) - V_{t+1}(s^A-1, s^B+1) - V_{t+1}(s^A, s^B) + V_{t+1}(s^A-1, s^B)\big)/\Lambda \\
&+s^B\tau^B\big(V_{t+1}(s^A+1, s^B) - V_{t+1}(s^A, s^B) - V_{t+1}(s^A+1, s^B-1) + V_{t+1}(s^A, s^B-1)\big)/\Lambda \\
=&0,
\end{aligned}
$$

because we assumed that $V_{t+1}(s^A, s^B)$ is modular on $S^A \times S^B$.