# Optimal Control of Dynamic Bipartite Matching Models

Arnaud Cadas
arnaud.cadas@inria.fr
Inria
DI ENS, CNRS, PSL Research University
Paris, France

Ana Bušić
ana.busic@inria.fr
Inria
DI ENS, CNRS, PSL Research University
Paris, France

Josu Doncel
josu.doncel@ehu.eus
University of the Basque Country, UPV-EHU
Leioa, Spain

## ABSTRACT

A dynamic bipartite matching model is given by a bipartite matching graph which determines the possible matchings between the various types of supply and demand items. Both supply and demand items arrive to the system according to a stochastic process. Matched pairs leave the system and the others wait in the queues, which induces a holding cost. We model this problem as a Markov Decision Process and study the discounted cost and the average cost case. We first consider a model with two types of supply and two types of demand items with an N-shaped matching graph. For linear cost function, we prove that an optimal matching policy gives priority to the end edges of the matching graph and is of threshold type for the diagonal edge. In addition, for the average cost problem, we compute the optimal threshold value. According to our numerical experiments, threshold-type policies perform also very well for more general bipartite graphs.

**ACM Reference Format:**
Arnaud Cadas, Ana Bušić, and Josu Doncel. 2019. Optimal Control of Dynamic Bipartite Matching Models. In *12th EAI International Conference on Performance Evaluation Methodologies and Tools (VALUETOOLS 2019), March 12–15, 2019, Palma, Spain.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3306309.3306317

## 1 INTRODUCTION

The theory of matching started with Peterson and König and was under a lot of interests in graph theory with problems like maximum matchings. It was extended to online matching setting [10, 14, 16] where one population is static and the other arrives according to a stochastic process. In the recent years, fully dynamic matching models have been considered where both populations are random. The importance of matching models was shown through applications in many fields: health [3, 11], ridesharing [4], power grid [20], or pattern recognition [19].

We study matching models from a queueing theory perspective, where a supply item and a demand item arrive to the system at each time step and can be matched or stay in buffers. [15, Theorem 1] proves that in a matching model where items arrive one by one, there exists no arrival distribution which verifies the necessary stability conditions for bipartite matching graphs. This result justifies why we assume arrivals by pairs as in [5, 6]. We consider that there is a holding cost that is a function of the buffer sizes. Our objective is to find the optimal matching policy in the discounted cost problem and in the average cost problem for general bipartite matching graphs. For this purpose, we model this problem as a Markov Decision Process.

The search for good policies and the performance analysis of various matching models have received great interest in the recent literature. For example, the FCFS infinite bipartite matching model was introduced in [9] and further studied in [1, 2] that established the reversibility of the dynamics and the product form of stationary distribution. In [5] the bipartite matching model was extended to other matching policies. In that paper, the authors established the necessary stability conditions and studied the stability region of various policies, including priorities and MaxWeigth. It is shown that MaxWeight has maximal stability region. For ride-sharing systems, state-dependent dispatch policies were identified in [4] which achieved exponential decay of the demand-dropping probability in heavy traffic regime. In [12], the authors presented the imbalance process and derived a lower bound on the holding costs.

Optimality results are scarce, and have been derived for some matching models in the asymptotic regimes. An extension of the greedy primal-dual algorithm was developed in [17] and was proved to be asymptotically optimal for the long-term average matching reward. However, they considered rewards on the edges, which differs from our model with holding costs. In [6], the authors considered the asymptotic heavy-traffic setting, and identified a policy that is approximately optimal with bounded regret using a workload relaxation approach.

In this paper we consider the non-asymptotic setting, i.e. we allow for any arrival rates under the stability conditions established in [5]. We first consider a matching model with two supply and two demand classes. For this system, we show that the optimal matching policy is of threshold type for the diagonal edge and with priority to the end edges of the matching graph. We also compute the optimal threshold in the case of the average cost.

For more general bipartite matching graphs, the optimal matching policy identified in the case $N$ can be generalized. We give a heuristic for general bipartite graphs where threshold-type policies performs also very well according to our preliminary numerical experiments.

## 2 MODEL DESCRIPTION

We consider a (bipartite) matching graph $(\mathcal{D} \cup \mathcal{S}, \mathcal{E})$ where $\mathcal{D} = \{d_1, d_2, \dots, d_{n_D}\}$ and $\mathcal{S} = \{s_1, s_2, \dots, s_{n_S}\}$ are, respectively, the set of demand nodes (or queues) and the set of supply nodes. $\mathcal{E} \subset \mathcal{D} \times \mathcal{S}$ is the set of allowed matching pairs. In Figure 1 it is depicted an example of a matching graph with three demand nodes and three


*VALUETOOLS 2019, March 12–15, 2019, Palma, Spain*

ACM ISBN 978-1-4503-6596-3/19/03…$15.00
https://doi.org/10.1145/3306309.3306317

supply nodes. In each time slot $n$, a demand item and a supply item arrive to the system according to the i.i.d. arrival process $A(n)$. We assume independence between demand and supply arrivals. The demand item arrives to the queue $d_i$ with probability $\alpha_i$ and the supply item arrives to the queue $s_j$ with probability $\beta_j$, i.e:

$$\forall i \in \{1, \cdots, n_{\mathcal{D}}\}, \forall j \in \{1, \cdots, n_{\mathcal{S}}\} \quad \mathbb{P}(A(n) = e_{d_i} + e_{s_j}) = \alpha_i \beta_j$$

with $\sum_{i=1}^{n_{\mathcal{D}}} \alpha_i = 1$, $\sum_{j=1}^{n_{\mathcal{S}}} \beta_j = 1$ and where $e_k \in \mathbb{N}^{n_{\mathcal{D}}+n_{\mathcal{S}}}$ is the vector of all zeros except in the $k$-th coordinate where it is equal to one, $k \in \mathcal{D} \cup \mathcal{S}$. We assume that the $\alpha_i$ and $\beta_j$ are chosen such that the arrival distribution satisfies the necessary and sufficient conditions for stabilizability of the MDP model: Ncond given in [6], i.e $\forall D \subsetneq \mathcal{D}$, $\forall S \subsetneq \mathcal{S}$:

$$\sum_{d_i \in D} \alpha_i < \sum_{s_j \in \mathcal{S}(D)} \beta_j \text{ and } \sum_{s_j \in S} \beta_j < \sum_{d_i \in \mathcal{D}(S)} \alpha_i$$

where $\mathcal{D}(j) = \{i \in \mathcal{D} : (i, j) \in \mathcal{E}\}$ is the set of demand classes that can be matched with a class $j$ supply and $\mathcal{S}(i) = \{j \in \mathcal{S} : (i, j) \in \mathcal{E}\}$ is the set of supply classes that can be matched with a class $i$ demand. The extension to subsets $S \subset \mathcal{S}$ and $D \subset \mathcal{D}$ is $\mathcal{D}(S) = \bigcup_{j \in S} \mathcal{D}(j)$ and $\mathcal{S}(D) = \bigcup_{i \in D} \mathcal{S}(i)$.

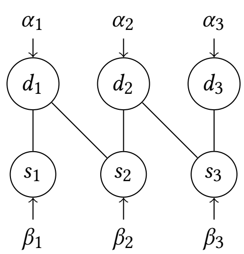


**Figure 1: A matching graph with three supply classes and three demand classes.**

We denote by $x_k(n)$ the queue length of node $k$ just after the arrivals at time slot $n$, where $k \in \mathcal{D} \cup \mathcal{S}$. Let $X(n) = (x_k(n))_{k \in \mathcal{D} \cup \mathcal{S}}$ be the vector of the queue length of all the nodes just after the arrivals. We must have $\sum_{k \in \mathcal{D}} x_k(n) = \sum_{k \in \mathcal{S}} x_k(n)$ for all $n$. Matchings at time $n$ are carried out after the arrivals at time $n$. Hence, $X(n)$ evolves over time according to the following expression:

$$X(n + 1) = X(n) - u(X(n)) + A(n + 1), \tag{1}$$

where $u$ is a deterministic Markovian decision rule which maps the current state $X(n)$ to the vector of the items that are matched at time $n$. Thus, $X$ is a Markov Decision Process where the control is denoted by $u$. It is sufficient to consider only deterministic Markovian decision rules and not all history-dependent randomized decision rules as proved in [18, Theorem 5.5.3] and [18, Proposition 6.2.1]. When the state of the system is $X(n) = x$, $u(x)$ must belong to the set of admissible matchings which is defined as:

$$U_x = \Big\{(u_i)_{i \in \mathcal{D} \cup \mathcal{S}} \in \mathbb{N}^{n_{\mathcal{D}}+n_{\mathcal{S}}} : \forall i \in \mathcal{D} \cup \mathcal{S} \; u_i \leq x_i, \\ \forall S \subset \mathcal{S} \sum_{j \in S} u_j \leq \sum_{i \in \mathcal{D}(S)} u_i, \; \forall D \subset \mathcal{D} \sum_{i \in D} u_i \leq \sum_{j \in \mathcal{S}(D)} u_j \Big\}$$

$U_x$ is defined for all $x \in \mathcal{X}$ where $\mathcal{X} = \{(x_i)_{i \in \mathcal{D} \cup \mathcal{S}} \in \mathbb{N}^{n_{\mathcal{D}}+n_{\mathcal{S}}} : \sum_{i \in \mathcal{D}} x_i = \sum_{j \in \mathcal{S}} x_j\}$ is the set of all the possible states of the system. We consider a linear cost function on the buffer size of the nodes: $c(X(n)) = \sum_{k \in \mathcal{D} \cup \mathcal{S}} c_k x_k(n)$. Our analysis presented in the following sections holds for more general cost functions as long as they satisfy the assumptions of Theorem 2.1 and Theorem 2.2. We chose a linear cost function because it satisfies these assumptions and allow us to give an analytical form for the optimal threshold. The buffer size of the nodes is infinite, thus we are in the unbounded costs setting.

A matching policy $\pi$ is a sequence of deterministic Markovian decision rules, i.e. $\pi = (u(X(n)))_{n \geq 1}$. The goal is to obtain the optimal matching policy for two optimization problems:

- The average cost problem: $g^* = \inf_\pi g^\pi$ with $g^\pi(x) = \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \mathbb{E}_x^\pi [c(X(n))]$
- The discounted cost problem: $v_\theta^* = \inf_\pi v_\theta^\pi$ with $v_\theta^\pi(x) = \lim_{N \to \infty} \sum_{n=0}^{N-1} \theta^t \mathbb{E}_x^\pi [c(X(n))]$

where $\theta \in [0, 1)$ is the discount factor. Both problems admit an optimal stationary policy, i.e. the decision rule depend only on the state of the system and not on the time [18]. The notation $\mathbb{E}_x^\pi$ indicates that the expectation is over the arrival process, given that $X(0) = x$ and using the matching policy $\pi$ to determine the matched items $u(X(n))$ for all $n$.

As $A(n)$ are i.i.d., to ease the notation from now on, we denote by $A$ a random variable with the same distribution as $A(1)$. For a given function $v$, $X(n) = x$, $u \in U_x$, we define for all $0 \leq \theta \leq 1$:

$$L_u^\theta v(x) = c(x) + \theta \mathbb{E}[v(x - u + A)]$$
$$L^\theta v(x) = c(x) + \min_{u \in U_x} \theta \mathbb{E}[v(x - u + A)]$$

and in particular, we define $T_u = L_u^1$ and $T = L^1$. A solution of the discounted cost problem can be obtained as a solution of the Bellman fixed point equation $v = L^\theta v$. In the average cost problem, the Bellman equation is given by $g^* + v = Tv$.

We say that a value function $v$ or a decision rule $u$ is structured if it satisfies a special property, such as being increasing, decreasing or convex. Throughout the article, by increasing we mean nondecreasing and we will use strictly increasing for increasing. A policy is called structured when it only uses structured decision rules.

The framework of this work is that of property preservation when we apply the Dynamic Programming operator. First, we identify a set of structured value functions $V^\sigma$ and a set of structured deterministic Markovian decision rules $D^\sigma$ such that if the value function belongs to $V^\sigma$ an optimal decision rule belongs to $D^\sigma$. Then, we show that the properties of $V^\sigma$ are preserved by the Dynamic Programming operator and that they hold in the limit. Theorem 2.1 [13, Theorem 1] lets us conclude that there exists an optimal policy which can be chosen in the set of structured stationary matching policies $\Pi^\sigma = \{\pi = (u(X(n)))_{n \geq 1} : u \in D^\sigma\}$.

Theorem 2.1. *[13, Theorem 1] Assume that the following properties hold: there exists positive function w on the state space $\mathcal{X}$ such that*

$$\sup_{(x,u)} \frac{c(x, u)}{w(x)} < +\infty, \tag{2}$$

$$\sup_{(x,u)} \frac{1}{w(x)} \sum_y \mathbb{P}(y|x, u) w(y) < +\infty, \tag{3}$$

*and for every $\mu$, $0 \le \mu < 1$, there exists $\eta$, $0 \le \eta < 1$ and some integer $J$ such that for every $J$-tuple of Markov deterministic decision rules $\pi = (u_1, \dots, u_J)$ and every $x$*

$$\mu^J \sum_y P_\pi(y|x)w(y) < \eta w(x), \tag{4}$$

*where $P_\pi$ denotes the $J$-step transition matrix under policy $\pi$. Let $0 \le \theta < 1$. Let $V_w$ the set of functions in the state space which have a finite $w$-weighted supremum norm, i.e., $\sup_x |v(x)/w(x)| < +\infty$. Assume that*

- ($*$) *for each $v \in V_w$, there exists a deterministic Markov decision rule $u$ such that $L^\theta v = L_u^\theta v$.*

*Let $V^\sigma$ and $D^\sigma$ be such that*

- (a) *$v \in V^\sigma$ implies that $L^\theta v \in V^\sigma$;*
- (b) *$v \in V^\sigma$ implies that there exists a decision $u' \in D^\sigma$ such that $u' \in \arg\min_u L_u^\theta v$;*
- (c) *$V^\sigma$ is a closed subset of the set of value functions under pointwise convergence.*

*Then, there exists an optimal stationary policy $\pi^* = (u^*(X(n)))_{n\ge1}$ that belongs to $\Pi^\sigma$ with $u^* \in \arg\min_u L_u^\theta v$.*

This result is an adapted version of [18, Theorem 6.11.3]. The former removes the need to verify that $V^\sigma \subset V_w$ (assumption made in the latter) and its statement separates the structural requirements (*(a)*, *(b)* and *(c)*) from the technical requirements related to the unboundedness of the cost function ((2), (3), (4) and *($*$)*).

In the case of the average cost problem, we will use the results of the discounted cost problem. We consider the average cost problem as a limit when $\theta$ tends to one and we show that the properties still hold for this limit. In order to prove the optimality in the average cost case, we will use [18, Theorem 8.11.1]:

Theorem 2.2. *[18, Theorem 8.11.1] Suppose that the following properties hold:*

- (A1) $\exists\, C \in \mathbb{R}, \forall x \in \mathcal{X},\ -\infty < C \le c(x) < +\infty$,
- (A2) $\forall x \in \mathcal{X}, \forall\, 0 \le \theta < 1,\ v_\theta^*(x) < +\infty$
- (A3) $\exists\, H \in \mathbb{R}, \forall x \in \mathcal{X}, \forall\, 0 \le \theta < 1,\ -\infty < H \le v_\theta^*(x) - v_\theta^*(0)$
- (A4) *There exists a nonnegative function $M(x)$ such that*
  - (a) $\forall x \in \mathcal{X},\ M(x) < +\infty$
  - (b) $\forall x \in \mathcal{X}, \forall\, 0 \le \theta < 1,\ v_\theta^*(x) - v_\theta^*(0) \le M(x)$
  - (c) *There exists $u \in U_0$ for which $\sum_y \mathbb{P}(y|0,u)M(y) < +\infty$*

*Let $H$ and $M$ be defined in Assumptions (A3) and (A4). We define a subset $V_H^\sigma$ of $V^\sigma$ which contains all the value functions $v \in V^\sigma$ such that $H \le v(x) - v(0) \le M(x)$ for all $x \in \mathcal{X}$. Then, if*

- (a) *for any sequence $(\theta_n)_{n\ge0}$, $0 \le \theta_n < 1$, for which $\lim_{n\to+\infty} \theta_n = 1$,*

  $$\lim_{n\to+\infty} [v_{\theta_n}^* - v_{\theta_n}^*(0)e] \in V_H^\sigma \quad \text{with } e(x) = 1 \text{ for all } x \in \mathcal{X}$$

  *and*
- (b) *$v \in V_H^\sigma$ implies that there exists a decision $u' \in D^\sigma$ such that $u' \in \arg\min_u T_u v$;*

*Then $D^\sigma \cap \arg\min_u T_u v \ne \emptyset$ and $u^* \in D^\sigma \cap \arg\min_u T_u v$ implies that the stationary matching policy which uses $u^*$ is lim sup average optimal.*

# 3 THE CASE $N$

We consider the system that is formed by two supply nodes and two demand nodes with a $N$-shaped matching graph (see Figure 2). Let $\ell_i$ be the edge between $d_i$ and $s_i$, i=1,2, and $\ell_3$ the edge between $d_1$ and $s_2$. Let us also define $\ell_4$ as the imaginary edge between $d_2$ and $s_1$ (imaginary because $\ell_4 \notin \mathcal{E}$) that we introduce to ease the notations. To ensure stability, we assume that $\alpha > \beta$.

The set of all the possible states of the system is

$$\mathcal{X} = \{x = (x_{d_1}, x_{d_2}, x_{s_1}, x_{s_2}) \in \mathbb{N}^4 : x_{d_1} + x_{d_2} = x_{s_1} + x_{s_2}\}$$

and the set of possible matchings, when the state of the system is $x \in \mathcal{X}$, is:

$$U_x = \{u = (u_{d_1}, u_{d_2}, u_{s_1}, u_{s_2}) \in \mathbb{N}^4 : (a)\ \forall i \in \{d_1, d_2, s_1, s_2\}\ u_i \le x_i,\ (b_1)\ u_{s_1} \le u_{d_1},\ (b_2)\ u_{d_2} \le u_{s_2},\ (c)\ u_{d_1} + u_{d_2} = u_{s_1} + u_{s_2}\}$$

We will show that the optimal policy for this case has a specific structure. For this purpose, we first present the properties of the value function. Then, we show how these properties characterize the optimal decision rule and how they are preserved by the dynamic programming operator. Finally, we prove the desired results in Theorem 3.11 and Theorem 3.12.

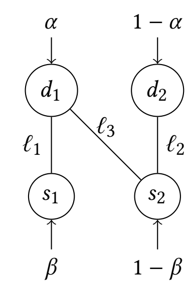


**Figure 2: The $N$-shaped matching graph.**

## 3.1 Value Function Properties

Let $e_i$ be the vector of all zeros except in the $i$-th coordinate, $i \in \{d_1, d_2, s_1, s_2\}$. Let $e_{\ell_1} = e_{d_1} + e_{s_1} = (1,0,1,0)$, $e_{\ell_2} = e_{d_2} + e_{s_2} = (0,1,0,1)$, $e_{\ell_3} = e_{d_1} + e_{s_2} = (1,0,0,1)$ and $e_{\ell_4} = e_{d_2} + e_{s_1} = (0,1,1,0)$. We start by defining increasing properties in $\ell_1$, $\ell_2$ and $\ell_4$:

*Definition 3.1 (Increasing property).* Let $i \in \{1,2,4\}$. We say that a function $v$ is increasing in $\ell_i$ or $v \in \mathcal{I}_{\ell_i}$ if

$$\forall x \in \mathcal{X}, \quad v(x + e_{\ell_i}) \ge v(x).$$

Remark 1. *The increasing property in $\ell_4$ can be interpreted as the fact that we prefer to match $\ell_1$ and $\ell_2$ rather than to match $\ell_3$. Indeed, $v(x + e_{\ell_4}) = v(x + e_{\ell_1} + e_{\ell_2} - e_{\ell_3}) \ge v(x)$.*

We also define the convexity in $\ell_3$ and $\ell_4$ as follows:

*Definition 3.2 (Convexity property).* A function $v$ is convex in $\ell_3$ or $v \in \mathcal{C}_{\ell_3}$ if $v(x + e_{\ell_3}) - v(x)$ is increasing in $\ell_3$, i.e.,

$$\forall x \in \mathcal{X}, x_{d_1} \ge x_{s_1} - 1 \quad v(x + 2e_{\ell_3}) - v(x + e_{\ell_3}) \ge v(x + e_{\ell_3}) - v(x).$$

Likewise, $v$ is convex in $\ell_4$ or $v \in C_{\ell_4}$ if $v(x+e_{\ell_4}) - v(x)$ is nondecreasing in $\ell_4$, i.e.,

$$\forall x \in \mathcal{X}, x_{s_1} \geq x_{d_1} - 1 \quad v(x+2e_{\ell_4}) - v(x+e_{\ell_4}) \geq v(x+e_{\ell_4}) - v(x).$$

*Definition 3.3 (Boundary property).* A function $v \in \mathcal{B}$ if

$$\forall j \in \{1,2,3,4\}, x = e_{\ell_j} \quad v(x) - v(x+e_{\ell_4}) \leq v(x+e_{\ell_3}) - v(x).$$

As we will show in Proposition 3.7, the properties $\mathcal{I}_{\ell_1}, \mathcal{I}_{\ell_2}, \mathcal{I}_{\ell_4}$ and $C_{\ell_3}$ characterize the optimal decision rule. On the other hand, $C_{\ell_4}$ and $\mathcal{B}$ are required to show that $C_{\ell_3}$ is preserved by the operator $L^\theta$.

We aim to characterize the optimal matching policy using Theorem 2.1 and Theorem 2.2. Thus, in the remainder of the article, we will consider the following set of structured value functions $V^\sigma = \mathcal{I}_{\ell_1} \cap \mathcal{I}_{\ell_2} \cap \mathcal{I}_{\ell_4} \cap C_{\ell_3} \cap C_{\ell_4} \cap \mathcal{B}$.

## 3.2 Optimal decision rule

In this section, we show that, for any $v \in V^\sigma$, there is a control of threshold-type in $\ell_3$ with priority to $\ell_1$ and $\ell_2$ that minimizes the $L_u^\theta v$.

*Definition 3.4 (Threshold-type decision rule).* A decision rule $u_x$ is said to be of threshold type in $\ell_3$ with priority to $\ell_1$ and $\ell_2$ if:

1. it matches all of $\ell_1$ and $\ell_2$.
2. it matches $\ell_3$ only if the remaining items (in $d_1$ and $s_2$) are above a specific threshold, denoted by $t$ (with $t \in \mathbb{N} \cup \infty$).

This means that:

- $(u_x)_{s_1} = \min(x_{d_1}, x_{s_1})$
- $(u_x)_{d_2} = \min(x_{d_2}, x_{s_2})$
- $(u_x)_{d_1} = \min(x_{d_1}, x_{s_1}) + k_t(x)$
- $(u_x)_{s_2} = \min(x_{d_2}, x_{s_2}) + k_t(x)$

where $k_t(x) = \begin{cases} 0 & if\ x_{d_1} - x_{s_1} \leq t \\ x_{d_1} - x_{s_1} - t & otherwise \end{cases}$.

We define $D^\sigma$ as the set of decision rules that are of threshold type in $\ell_3$ with priority to $\ell_1$ and $\ell_2$ for any $t \in \mathbb{N} \cup \infty$.

Remark 2. *If $t = \infty$, the decision rule will never match $\ell_3$. If $x_{d_1} - x_{s_1} \leq t < \infty$, the decision rule will match $\ell_3$ until the remaining items in $d_1$ and $s_2$ are equal to the threshold $t$.*

In the next proposition, we establish that there exists an optimal decision rule with priority to $\ell_1$ and $\ell_2$.

Proposition 3.5. *Let $v \in \mathcal{I}_{\ell_1} \cap \mathcal{I}_{\ell_2} \cap \mathcal{I}_{\ell_4}$, let $0 \leq \theta \leq 1$. For any $x \in \mathcal{X}$, there exists $u^* \in U_x$ such that $u^* \in \arg\min_{u \in U_x} L_u^\theta v(x)$, $(u^*)_{s_1} = \min(x_{d_1}, x_{s_1})$ and $(u^*)_{d_2} = \min(x_{d_2}, x_{s_2})$. In particular, this result holds for the average operator: $T_u$.*

Proof. Let $v \in \mathcal{I}_{\ell_1} \cap \mathcal{I}_{\ell_2} \cap \mathcal{I}_{\ell_4}$, $0 \leq \theta \leq 1$, $x \in \mathcal{X}$, $u \in U_x$ and $p = u_{d_1} - u_{s_1}$ (the number of matchings in $\ell_3$ of $u$). The maximum number of matchings in $\ell_1$ is denoted by $m_{\ell_1} = \min(x_{d_1}, x_{s_1})$ and in $\ell_2$ by $m_{\ell_2} = \min(x_{d_2}, x_{s_2})$.

Let $p_0 = \min(p, x_{s_1} - u_{s_1}, x_{d_2} - u_{d_2})$ be the number of possible matchings that can be transformed from $\ell_3$ to $\ell_1$ and $\ell_2$ matchings. We define a policy $u_0$ that removes the $p_0$ matchings in $\ell_3$ and matches $p_0$ times $\ell_1$ and $\ell_2$, that is, $u_0 = u + p_0(e_{\ell_1} + e_{\ell_2} - e_{\ell_3})$. We verify that this policy is admissible, i.e $u_0 \in U_x$: *(c)* is true because $u \in U_x$. *(a)* is true because $(u_0)_{s_1} = (u)_{s_1} + p_0 \leq (u)_{s_1} + x_{s_1} - (u)_{s_1} = x_{s_1}$ and $(u_0)_{d_2} = (u)_{d_2} + p_0 \leq (u)_{d_2} + x_{d_2} - (u)_{d_2} = x_{d_2}$. *($b_1$)* and *($b_2$)* are true because $(u_0)_{s_1} = (u)_{s_1} + p_0 \leq (u)_{s_1} + p = (u)_{d_1}$ and $(u_0)_{d_2} = (u)_{d_2} + p_0 \leq (u)_{d_2} + p = (u)_{s_2}$. Then, we can use the fact that $v \in \mathcal{I}_{\ell_4}$ to show that $L_{u_0}^\theta v(x) \leq L_u^\theta v(x)$.

Moreover, we define $u'$ that matches all the the possible $\ell_1$ and $\ell_2$ of $x - u_0$, that is, of the remaining items when we apply $u_0$: $u' = u_0 + e_{\ell_1}(m_{\ell_1} - (u_0)_{s_1}) + e_{\ell_2}(m_{\ell_2} - (u_0)_{d_2})$. We also verify that this policy is admissible, i.e $u' \in U_x$: *(c)*, *($b_1$)* and *($b_2$)* are true because $u_0 \in U_x$. If $p_0 = p$, then

$$\begin{aligned}(u')_{d_1} &= (u_0)_{d_1} + m_{\ell_1} - (u_0)_{s_1} = m_{\ell_1} + u_{d_1} - u_{s_1} - p_0 \\ &= m_{\ell_1} + p - p_0 \leq x_{d_1} \\ (u')_{s_2} &= (u_0)_{s_2} + m_{\ell_2} - (u_0)_{d_2} = m_{\ell_2} + u_{s_2} - u_{d_2} - p_0 \\ &= m_{\ell_2} + p - p_0 \leq x_{s_2}\end{aligned}$$

If $p_0 = x_{s_1} - u_{s_1}$, then

$$\begin{aligned}(u')_{d_1} &= m_{\ell_1} + u_{d_1} - u_{s_1} - p_0 = m_{\ell_1} + u_{d_1} - x_{s_1} \leq u_{d_1} \\ (u')_{s_2} &= m_{\ell_2} + u_{d_1} - u_{s_1} - p_0 = m_{\ell_2} + u_{d_1} - x_{s_1} \leq x_{d_2} + u_{d_1} - x_{s_1} \\ &= x_{s_2} + u_{d_1} - x_{d_1} \leq x_{s_2}\end{aligned}$$

If $p_0 = x_{d_2} - u_{d_2}$, then

$$\begin{aligned}(u')_{d_1} &= m_{\ell_1} + u_{s_2} - u_{d_2} - p_0 = m_{\ell_1} + u_{s_2} - x_{d_2} \leq x_{s_1} + u_{s_2} - x_{d_2} \\ &= x_{d_1} + u_{s_2} - x_{s_2} \leq x_{d_1} \\ (u')_{s_2} &= m_{\ell_2} + u_{s_2} - u_{d_2} - p_0 = m_{\ell_2} + u_{s_2} - x_{d_2} \leq u_{s_2}\end{aligned}$$

In every case, *(a)* is true. Hence, since $v \in \mathcal{I}_{\ell_1} \cap \mathcal{I}_{\ell_2}$, it results that $L_{u'}^\theta v(x) \leq L_{u_0}^\theta v(x)$.

As a result, we have $L_{u'}^\theta v(x) \leq L_{u_0}^\theta v(x)$, $(u')_{s_1} = (u_0)_{s_1} + m_{\ell_1} - (u_0)_{s_1} = \min(x_{s_1}, x_{d_1})$ and $(u')_{d_2} = (u_0)_{d_2} + m_{\ell_2} - (u_0)_{d_2} = \min(x_{s_2}, x_{d_2})$. This was done for any $u \in U_x$ and because $U_x$ is finite for every $x \in \mathcal{X}$, we can choose $u^*$ such that it belongs to $\arg\min_{u \in U_x} L_u^\theta v(x)$ giving the final result. □

From this result, it follows that there exists an optimal decision rule that matches all possible $\ell_1$ and $\ell_2$. In addition, due to Proposition 3.5 and *(c)* from the definition of $U_x$, there exists $u' \in \arg\min_{u \in U_x} L_u^\theta v(x)$ such that we have $u'_{d_1} = \min(x_{s_1}, x_{d_1}) + k$ and $u'_{s_2} = \min(x_{s_2}, x_{d_2}) + k$. Our goal now is to find the optimal number of matchings in $\ell_3$ (i.e the optimal $k$). We introduce first some notation:

*Definition 3.6.* Let $0 \leq \theta \leq 1$, $x \in \mathcal{X}$. We define:

$$K_x = \begin{cases} \{0\} & if\ x_{d_1} \leq x_{s_1} \\ \{0, \cdots, \min(x_{d_1} - x_{s_1}, x_{s_2} - x_{d_2})\} & otherwise \end{cases}$$

the set of possible matching in $\ell_3$ after having matched all possible $\ell_1$ and $\ell_2$.

Remark 3. *The state of the system after having matched all possible $\ell_1$ and $\ell_2$ is of the form $(0, l, l, 0)$ if $x_{d_1} \leq x_{s_1}$ and of the form $(l, 0, 0, l)$ otherwise (because of the definition of $\mathcal{X}$ and $U_x$).*

Finally, we prove that a decision rule of threshold type in $\ell_3$ with priority to $\ell_1$ and $\ell_2$ is optimal. This is done by choosing the right $t$ for different cases such that $k_t(x)$ is the optimal number of matchings in $\ell_3$ for a given $x$.

Proposition 3.7. *Let $v \in \mathcal{I}_{\ell_1} \cap \mathcal{I}_{\ell_2} \cap \mathcal{I}_{\ell_4} \cap \mathcal{C}_{\ell_3}$. Let $0 \le \theta \le 1$. There exists $u^* \in D^\sigma$ (see Definition 3.4) such that $u^* \in \arg\min_{u \in U_x} L_u^\theta v$. In particular, this result holds for the average operator: $T_u$.*

Proof. Let $x \in \mathcal{X}$ and $u \in U_x$. We note $m_{\ell_1} = \min(x_{s_1}, x_{d_1})$ and $m_{\ell_2} = \min(x_{s_2}, x_{d_2})$. We supposed that $v \in \mathcal{I}_{\ell_1} \cap \mathcal{I}_{\ell_2} \cap \mathcal{I}_{\ell_4}$, so we can use Proposition 3.5 : $\exists u' \in U_x$ such that $L_{u'}^\theta v(x) \le L_u^\theta v(x)$ and $u' = (m_{\ell_1} + k, m_{\ell_2}, m_{\ell_1}, m_{\ell_2} + k)$ with $k \in K_x$. We now have to prove that there exists $t \in \mathbb{N} \cap \infty$ such that

$$L_{u^*}^\theta v(x) \le L_{u'}^\theta v(x), \quad \forall k \in K_x \tag{5}$$

where $u^* \in D^\sigma$ (see Definition 3.4). If $x_{s_1} \ge x_{d_1}$, then $K_x = 0$ and we have $u^* = u'$ which satisfies (5). Otherwise, $x_{s_1} < x_{d_1}$ and $K_x \neq \{0\}$. Therefore, the state of the system after having matched $u'$ (or $u^*$), i.e $x - u'$ (or $x - u^*$), is of the form $(l, 0, 0, l)$. So when we compare $L_{u^*}^\theta v(x)$ with $L_{u'}^\theta v(x)$, this comes down to comparing $\mathbb{E}[v(A + j^* e_{\ell_3})]$ with $\mathbb{E}[v(A + j' e_{\ell_3})]$ ($j^*, j' \in K_x$).

First of all, suppose that $\forall j \in \mathbb{N},\ \mathbb{E}[v(A + (j+1)e_{\ell_3})] - \mathbb{E}[v(A + je_{\ell_3})] \le 0$. We choose $t = \infty$, so $k_t(x) = 0$ and $u^* = (m_{\ell_1}, m_{\ell_2}, m_{\ell_1}, m_{\ell_2})$. By assumption, we have $L_{u^*}^\theta v(x) \le L_{u^*+e_{\ell_3}}^\theta v(x) \le \cdots \le L_{u^*+ke_{\ell_3}}^\theta v(x)$ for all $k \in K_x$ and because $u^* + ke_{\ell_3} = u'$, we have proven (5).

Then, suppose that $\mathbb{E}[v(A+e_{\ell_3})] - \mathbb{E}[v(A)] \ge 0$. We choose $t = 0$, so $k_t(x) = x_{d_1} - x_{s_1}$ and $u^* = (m_{\ell_1} + x_{d_1} - x_{s_1}, m_{\ell_2}, m_{\ell_1}, m_{\ell_2} + x_{d_1} - x_{s_1})$. By assumption and because $v$ is convex in $\ell_3$, we have $L_{u^*}^\theta v(x) \le L_{u^*-e_{\ell_3}}^\theta v(x) \le \cdots \le L_{u^*-ke_{\ell_3}}^\theta v(x)$ for all $k \in K_x$ and because $u^* - (x_{d_1} - x_{s_1} - k)e_{\ell_3} = u'$ (with $x_{d_1} - x_{s_1} - k \in K_x$ for all $k \in K_x$), we have proven (5).

Finally, suppose that $\exists j \in \mathbb{N}^*,\ \mathbb{E}[v(A + (j+1)e_{\ell_3})] - \mathbb{E}[v(A + je_{\ell_3})] \ge 0$. Let $\underline{j} = \min\{j \in \mathbb{N}^* :\ \mathbb{E}[v(A + (j+1)e_{\ell_3})] - \mathbb{E}[v(A + je_{\ell_3})] \ge 0\}$. By definition of $\underline{j}$ and by convexity of $v$ in $\ell_3$, we have

$$\mathbb{E}[v(A+(\underline{j}-l)e_{\ell_3})] - \mathbb{E}[v(A+(\underline{j}-l-1)e_{\ell_3})] \le 0 \quad \forall l \in [\![0\,;\,\underline{j}-1]\!] \tag{6}$$

and

$$\mathbb{E}[v(A + (\underline{j} + 1 + l)e_{\ell_3})] - \mathbb{E}[v(A + (\underline{j} + l)e_{\ell_3})] \ge 0 \quad \forall l \in \mathbb{N} \tag{7}$$

We choose $t = \underline{j}$. If $x_{d_1} - x_{s_1} \le \underline{j}$, then we have $k_t(x) = 0$ and $L_{u^*}^\theta v(x) \le L_{u'}^\theta v(x)$ for all $k \in K_x$ by (6) ($0 \le k \le x_{d_1} - x_{s_1} \le \underline{j}$). Otherwise $x_{d_1} - x_{s_1} > \underline{j}$, then $k_t(x) = x_{d_1} - x_{s_1} - \underline{j}$ and $L_{u^*}^\theta v(x) = c(x) + \mathbb{E}[v(A + \underline{j}e_{\ell_3})]$. Therefore, for all $k \in K_x$, $L_{u^*}^\theta v(x) \le L_{u'}^\theta v(x)$ by (6) if $k \ge \underline{j}$ or by (7) if $k \le \underline{j}$, which proves (5).

□

## 3.3 Value Function Property Preservation

In this section, we show that the properties of the value function defined in Section 3.1 are preserved by the dynamic programming operator. In other words, we show that if $v \in V^\sigma$, then $L^\theta v \in V^\sigma$. We recall that $V^\sigma = \mathcal{I}_{\ell_1} \cap \mathcal{I}_{\ell_2} \cap \mathcal{I}_{\ell_4} \cap \mathcal{C}_{\ell_3} \cap \mathcal{C}_{\ell_4} \cap \mathcal{B}$. To prove the desired result, the cost function must satisfy the same properties as the value function.

Assumption 1. *The cost function $c$ is a nonnegative function which belongs to $V^\sigma$.*

In the remainder of the article, we will suppose that Assumption 1 holds.

We first show that the monotonicity on $\ell_1$, $\ell_2$ and $\ell_4$ is also preserved by the dynamic programming operator.

Lemma 3.8. *If a function $v \in \mathcal{I}_{\ell_1} \cap \mathcal{I}_{\ell_2} \cap \mathcal{I}_{\ell_4}$, then $L^\theta v \in \mathcal{I}_{\ell_1} \cap \mathcal{I}_{\ell_2} \cap \mathcal{I}_{\ell_4}$.*

Proof. Let $\underline{x} \in \mathcal{X}$. We define $\overline{x} = \underline{x} + e_{\ell_1}$. Since $v$ is increasing with $\ell_1$, we have that $v(\overline{x}) \ge v(\underline{x})$. We aim to show that $L^\theta v(\overline{x}) \ge L^\theta v(\underline{x})$.

Let $u_{\overline{x}} \in \arg\min_{u \in U_{\overline{x}}} L^\theta v(\overline{x})$. Since $(\overline{x})_{s_1} \ge 1$ and $(\overline{x})_{d_1} \ge 1$, using Proposition 3.5, we can choose $u_{\overline{x}}$ such that $(u_{\overline{x}})_{s_1} = \min(\overline{x}_{d_1}, \overline{x}_{s_1}) \ge 1$ and $(u_{\overline{x}})_{d_1} \ge (u_{\overline{x}})_{s_1} \ge 1$ since $u_{\overline{x}} \in U_{\overline{x}}$. Therefore, we can define $u_{\underline{x}} = u_{\overline{x}} - e_{\ell_1}$. $u_{\underline{x}} \in U_{\underline{x}}$ because $u_{\overline{x}} \in U_{\overline{x}}$, $(u_{\overline{x}})_{d_1} - 1 \le \overline{x}_{d_1} - 1 = \underline{x}_{d_1}$ and $(u_{\overline{x}})_{s_1} - 1 \le \overline{x}_{s_1} - 1 = \underline{x}_{s_1}$. Besides, $\overline{x} - u_{\overline{x}} = \underline{x} - u_{\underline{x}}$ and $c(\overline{x}) \ge c(\underline{x})$ since $c \in \mathcal{I}_{\ell_1}$ from Assumption 1. Hence,

$$\begin{aligned} L_{u_{\underline{x}}}^\theta v(\underline{x}) &= c(\underline{x}) + \theta\mathbb{E}[v(\underline{x} - u_{\underline{x}} + A)] \\ &= c(\underline{x}) - c(\overline{x}) + c(\overline{x}) + \theta\mathbb{E}[v(\overline{x} - u_{\overline{x}} + A)] \\ &= c(\underline{x}) - c(\overline{x}) + L^\theta v(\overline{x}) \\ &\le L^\theta v(\overline{x}). \end{aligned}$$

And, since $u_{\underline{x}} \in U_{\underline{x}}$, then by definition $L^\theta v(\underline{x}) \le L_{u_{\underline{x}}}^\theta v(\underline{x})$ and, as a result, $L^\theta v(\underline{x}) \le L^\theta v(\overline{x})$. The same arguments with $\overline{x} = \underline{x} + e_{\ell_2}$ can be made to show that $L^\theta v(\underline{x}) \le L^\theta v(\overline{x})$.

The proof is similar for $\mathcal{I}_{\ell_4}$ but also requires to handle the case when no matching can be made in $\ell_3$. Let $\underline{x} \in \mathcal{X}$. We denote $\overline{x} = \underline{x} + e_{\ell_1} + e_{\ell_2} - e_{\ell_3}$. Since $v \in \mathcal{I}_{\ell_4}$, we know that $v(\underline{x}) \le v(\overline{x})$. $c(\underline{x}) \le c(\overline{x})$ holds because of Assumption 1. We aim to show that $L^\theta v(\underline{x}) \le L^\theta v(\overline{x})$.

Using Proposition 3.5, let $u_{\overline{x}} \in \arg\min_{u \in U_{\overline{x}}} L_v^\theta v(\overline{x})$ such that $(u_{\overline{x}})_{d_2} = \min(\overline{x}_{d_2}, \overline{x}_{s_2})$ and $(u_{\overline{x}})_{s_1} = \min(\overline{x}_{d_1}, \overline{x}_{s_1})$. We define $u_{\underline{x}} = u_{\overline{x}} - e_{\ell_1} - e_{\ell_2} + e_{\ell_3}$. Suppose that $\underline{x}_{d_1} \ge 1$ and $\underline{x}_{s_2} \ge 1$, we have that $\underline{x} - u_{\underline{x}} = \overline{x} - u_{\overline{x}}$ and $u_{\underline{x}} \in U_{\underline{x}}$ because $u_{\overline{x}} \in U_{\overline{x}}$, $0 \le (u_{\overline{x}})_{s_1} - 1 \le \overline{x}_{s_1} - 1 = \underline{x}_{s_1}$ and $0 \le (u_{\overline{x}})_{d_2} - 1 \le \overline{x}_{d_2} - 1 = \underline{x}_{d_2}$. Thus,

$$\begin{aligned} L_{u_{\underline{x}}}^\theta v(\underline{x}) &= c(\underline{x}) + \theta\mathbb{E}[v(\underline{x} - u_{\underline{x}} + A)] \\ &= c(\underline{x}) - c(\overline{x}) + c(\overline{x}) + \theta\mathbb{E}[v(\overline{x} - u_{\overline{x}} + A)] \\ &= c(\underline{x}) - c(\overline{x}) + L^\theta v(\overline{x}) \\ &\le L^\theta v(\overline{x}), \end{aligned}$$

and $L^\theta v(\underline{x}) \le L_{u_{\underline{x}}}^\theta v(\underline{x})$ as $u_{\underline{x}} \in U_{\underline{x}}$. Suppose now that $\underline{x}_{d_1} = 0$ or $\underline{x}_{s_2} = 0$. In that case, we can not do more matchings in state $\overline{x}$ than we can do in state $\underline{x}$: $u_{\overline{x}} \in U_{\underline{x}}$. Thus,

$$\begin{aligned} L_{u_{\overline{x}}}^\theta v(\underline{x}) &= c(\underline{x}) + \theta\mathbb{E}[v(\underline{x} - u_{\overline{x}} + A)] \\ &\le c(\underline{x}) + \theta\mathbb{E}[v(\overline{x} - u_{\overline{x}} + A)] \text{ since } v \in \mathcal{I}_{\ell_4} \\ &= c(\underline{x}) - c(\overline{x}) + c(\overline{x}) + \theta\mathbb{E}[v(\overline{x} - u_{\overline{x}} + A)] \\ &= c(\underline{x}) - c(\overline{x}) + L^\theta v(\overline{x}) \\ &\le L^\theta v(\overline{x}), \end{aligned}$$

and $L^\theta v(\underline{x}) \le L_{u_{\overline{x}}}^\theta v(\underline{x})$ as $u_{\overline{x}} \in U_{\underline{x}}$. In both cases we get the desired result $L^\theta v(\underline{x}) \le L^\theta v(\overline{x})$.

□

The proof that the dynamic programming operator preserves the convexity in $\ell_3$ was more difficult than anticipated. We had to introduce the boundary property $\mathcal{B}$ to show that $L^\theta v \in C_{\ell_3}$ for a specific case.

Lemma 3.9. *If $v \in \mathcal{I}_{\ell_1} \cap \mathcal{I}_{\ell_2} \cap \mathcal{I}_{\ell_4} \cap C_{\ell_3} \cap \mathcal{B}$, then $L^\theta v \in C_{\ell_3}$.*

Proof. Let $\underline{x} \in \mathcal{X}, \underline{x}_{d_1} \geq \underline{x}_{s_1} - 1, \overline{x} = \underline{x} + e_{\ell_3}$ and $\overline{\overline{x}} = \overline{x} + e_{\ell_3}$. Since $v$ is convex in $\ell_3$, we have $v(\overline{x}) - v(\underline{x}) \leq v(\overline{\overline{x}}) - v(\overline{x})$ (this inequality also holds for the cost function $c$ because of Assumption 1). We aim to show that $L^\theta v(\overline{x}) - L^\theta v(\underline{x}) \leq L^\theta v(\overline{\overline{x}}) - L^\theta v(\overline{x})$. For $y \in \left\{\underline{x}, \overline{x}, \overline{\overline{x}}\right\}$, let $u_y \in \arg\min_u L_u^\theta v(y)$. From Proposition 3.7, we can choose $u_y$ such that $u_y = \min(y_{d_1}, y_{s_1}) e_{\ell_1} + \min(y_{d_2}, y_{s_2}) e_{\ell_2} + k_t(y) e_{\ell_3}$

Let us also define $m = \underline{x} - u_{\underline{x}} + A$. Suppose that $\underline{x}_{d_1} \geq \underline{x}_{s_1}$, we can distinguish 3 cases: (a) $k_t(\overline{x}) > 0$, (b) $k_t(\overline{x}) = 0$ and $k_t(\overline{\overline{x}}) > 0$ and (c) $k_t(\overline{x}) = 0$ and $k_t(\overline{\overline{x}}) = 0$:

(a) If $k_t(\overline{x}) > 0$. Then,

$$\begin{aligned} L^\theta v(\overline{x}) - L^\theta v(\underline{x}) &= c(\overline{x}) - c(\underline{x}) + \theta \mathbb{E}[v(m) - v(m)] \\ &\leq c(\overline{\overline{x}}) - c(\overline{x}) + \theta \mathbb{E}[v(m) - v(m)] \\ &= L^\theta v(\overline{\overline{x}}) - L^\theta v(\overline{x}) \end{aligned}$$

because $c \in C_{\ell_3}$.

(b) If $k_t(\overline{x}) = 0$ and $k_t(\overline{\overline{x}}) > 0$. Then,

$$\begin{aligned} L^\theta v(\overline{x}) - L^\theta v(\underline{x}) &= c(\overline{x}) - c(\underline{x}) + \theta \mathbb{E}[v(m + e_{\ell_3}) - v(m)] \\ &= c(\overline{x}) - c(\underline{x}) + \theta \mathbb{E}[v(m + e_{\ell_3}) - v(m + e_{\ell_3} - e_{\ell_3})] \\ &= c(\overline{x}) - c(\underline{x}) + L^\theta v(\overline{x}) - L^\theta_{u_{\overline{x}} + e_{\ell_3}} v(\overline{x}) \\ &\leq c(\overline{\overline{x}}) - c(\overline{x}) \\ &= c(\overline{\overline{x}}) - c(\overline{x}) + \theta \mathbb{E}[v(m + e_{\ell_3}) - v(m + e_{\ell_3})] \\ &= L^\theta v(\overline{\overline{x}}) - L^\theta v(\overline{x}) \end{aligned}$$

because $c \in C_{\ell_3}$ and because $k_t(\underline{x}) = k_t(\overline{x}) = 0$ and $1 \in K_{\overline{x}}$.

(c) If $k_t(\overline{x}) = 0$ and $k_t(\overline{\overline{x}}) = 0$. Then,

$$\begin{aligned} L^\theta v(\overline{x}) - L^\theta v(\underline{x}) &= c(\overline{x}) - c(\underline{x}) + \theta \mathbb{E}[v(m + e_{\ell_3}) - v(m)] \\ &\leq c(\overline{\overline{x}}) - c(\overline{x}) + \theta \mathbb{E}[v(m + 2e_{\ell_3}) - v(m + e_{\ell_3})] \\ &= L^\theta v(\overline{\overline{x}}) - L^\theta v(\overline{x}) \end{aligned}$$

because $c \in C_{\ell_3}$ and $v \in C_{\ell_3}$.

Suppose now that $\underline{x}_{d_1} = \underline{x}_{s_1} - 1$, we can distinguish 2 cases: $k_t(\overline{\overline{x}}) > 0$ and $k_t(\overline{\overline{x}}) = 0$:

- If $k_t(\overline{\overline{x}}) > 0$. Then,

$$\begin{aligned} L^\theta v(\overline{x}) - L^\theta v(\underline{x}) &= c(\overline{x}) - c(\underline{x}) + \theta \mathbb{E}[v(m - e_{\ell_4}) - v(m)] \\ &\leq c(\overline{x}) - c(\underline{x}) + \theta \mathbb{E}[v(m) - v(m)] \\ &\leq c(\overline{\overline{x}}) - c(\overline{x}) + \theta \mathbb{E}[v(m - e_{\ell_4}) - v(m - e_{\ell_4})] \\ &= L^\theta v(\overline{\overline{x}}) - L^\theta v(\overline{x}) \end{aligned}$$

because $v \in \mathcal{I}_{\ell_4}$ and $c \in C_{\ell_3}$.

- If $k_t(\overline{\overline{x}}) = 0$. Then,

$$\begin{aligned} L^\theta v(\overline{x}) - L^\theta v(\underline{x}) &= c(\overline{x}) - c(\underline{x}) + \theta \mathbb{E}[v(m - e_{\ell_4}) - v(m)] \\ &\leq c(\overline{\overline{x}}) - c(\overline{x}) + \theta \mathbb{E}[v(m - e_{\ell_4} + e_{\ell_3}) - v(m - e_{\ell_4})] \\ &= L^\theta v(\overline{\overline{x}}) - L^\theta v(\overline{x}) \end{aligned}$$

because $v \in \mathcal{B}$ and $c \in C_{\ell_3}$.

□

Then, to show that the dynamic programming operator preserves $\mathcal{B}$, we had to introduce an other property: $C_{\ell_4}$ (convexity in $\ell_4$). The preservation of these two last properties are combined in the next lemma (the proof can be found in the detailed version of the present paper [8, Appendix A]).

Lemma 3.10. *If $v \in \mathcal{I}_{\ell_1} \cap \mathcal{I}_{\ell_2} \cap \mathcal{I}_{\ell_4} \cap C_{\ell_3} \cap C_{\ell_4} \cap \mathcal{B}$, then $L^\theta v \in C_{\ell_4} \cap \mathcal{B}$.*

In this section, we have shown that the structural properties of the value function presented in Section 3.1 are preserved by the dynamic programming operator. That is, if $v \in V^\sigma$, then $L^\theta v$ also belongs to this set. Using this result, we give the structure of the optimal policy in the next section.

### 3.4 Structure of the Optimal Policy

We now present that, using the result of Theorem 2.1, there exists an optimal matching policy which is formed of a sequence of decision rules that belong to $D^\sigma$ (with a fixed threshold). In this section, we also assume that the cost function is linear. All the previous results did not require the linearity of the cost function. However, it is useful to assume it from now on in order to respect the technical assumptions of Theorem 2.1 due to unbounded costs and to compute the optimal threshold in the average cost case.

Theorem 3.11. *The optimal control for the discounted cost problem is of threshold type in $\ell_3$ with priority to $\ell_1$ and $\ell_2$.*

Proof. We apply Theorem 2.1 where $V^\sigma$ is the set of functions satisfying Definition 3.1 to Definition 3.3 and $D^\sigma$ the set defined in Definition 3.4.

We first show that the technical details given in (2)-(4) are verified. We choose $w(x) = \sum_i x_i + 1$. In our case, the cost is a linear function of $x$ therefore, $c(x, u)/w(x) \leq \max_{i \in \mathcal{D} \cup \mathcal{S}} c_i$. This shows (2). In addition,

$$\begin{aligned} \frac{1}{w(x)} \sum_y \mathbb{P}(y|x, u) w(y) &= \mathbb{E}\left[\frac{w(x - u + a)}{w(x)} \middle| A = a\right] \\ &\leq E\left[\frac{w(x + a)}{w(x)} \middle| A = a\right] = \frac{\sum_i x_i + 3}{\sum_i x_i + 1} \leq 3 \end{aligned}$$

since $w(x)$ is increasing and two items arrive to the system in each step following a process which is independent of the state of the system. This shows (3). Finally, we can repeat the previous argument to show that for all $J$-step matching policy $\pi$

$$\sum_y P_\pi(y|x) w(y) \leq \sum_y P_{\pi_0}(y|x) w(y) = w(x) + 2J.$$

where $\pi_0 = (0, \cdots, 0)$ is the policy which does not match any items. Therefore, (4) is satisfied if there exist $J$ integer and $\eta < 1$ such that

$$\mu^J(w(x) + 2J) \leq \eta w(x) \iff \eta > \frac{\mu^J(w(x) + 2J)}{w(x)}.$$

Since it is decreasing with $J$ and when $J \to \infty$ it tends to zero, there exists a $J$ integer such that $\eta$ is less than one and, therefore, (4) is also verified.

Since for each state of the system, the set of admissible matching policies is finite, we can use [18, Theorem 6.2.10] to show that $(*)$ (see Theorem 2.1) holds.

We now focus on the structural conditions of the theorem. From Lemma 3.8, Lemma 3.9 and Lemma 3.10 of Section 3.3, it follows $(a)$ since they show that if $v \in V^\sigma$, then $L^\theta v \in V^\sigma$. The result of Proposition 3.7 shows (b) because the policy that belong to $D^\sigma$ minimize $L_u^\theta v$ if $v \in V^\sigma$. Finally, since limits preserve inequalities, the point-wise convergence of functions of $V^\sigma$ belong to this set, which shows (c). □

The following theorem shows that the previous result is also proven for the average cost problem.

Theorem 3.12. *The optimal control for the average cost problem is of threshold type in $\ell_3$ with priority to $\ell_1$ and $\ell_2$.*

Proof. We want to apply Theorem 2.2 using the same value function set $V^\sigma$ and the same decision rule set $D^\sigma$ as in the proof of the previous proposition. Let us show first that the Assumptions *(A1)* to *(A4)* hold.

Assumption *(A1)* holds using $C = 0$ because of Assumption 1. Following the proof of Proposition 3.13, we can define a stationary policy $u_t^\infty \in \Pi^{T_{\ell_3}}$ for which the derived Markov chain is positive recurrent and $g^{u_t^\infty} < \infty$. Moreover, the set $\{x \in \mathcal{X} : c(x) < g^{u_t^\infty}\}$ is nonempty because $g^{u_t^\infty} > 0$ almost surely and $c(0) = 0$. It is also finite because $g^{u_t^\infty} < \infty$, $c \in V^\sigma$ and $t$ is finite. Therefore, we can use [18, Theorem 8.10.9] and Assumptions *(A2)* to *(A4)* hold.

Let $(\theta_n)_{n \in \mathbb{N}}$ be a sequence such that $0 \leq \theta_n < 1$ for all $n \in \mathbb{N}$ and $\lim_{n \to +\infty} \theta_n = 1$. Let $n \in \mathbb{N}$. We know that $v^*_{\theta_n} \in V^\sigma$ (see the proof of Theorem 3.11). The inequalities in Definition 3.1 to Definition 3.2 still hold if we add a constant to $v$, thus $v^*_{\theta_n} - v^*_{\theta_n}(0)e \in V^\sigma$. Using Assumption *(A3)* and Assumption *(A4)*, we have $H \leq v^*_{\theta_n} - v^*_{\theta_n}(0)e \leq M$, so $v^*_{\theta_n} - v^*_{\theta_n}(0)e \in V_H^\sigma$. This last result holds for each $n \in \mathbb{N}$ and since limits preserve inequalities $V_H^\sigma$ is a closed set, $\lim_{n \to +\infty} [v^*_{\theta_n} - v^*_{\theta_n}(0)e] \in V_H^\sigma$ which shows (a). The result of Proposition 3.7 shows (b) because the policy that belong to $D^\sigma$ minimize $L_u^1 v = T_u v$ if $v \in V_H^\sigma \subset V^\sigma$. □

## 3.5 Computing the Optimal Threshold

We consider the matching policy of threshold type in $\ell_3$ with priority to $\ell_1$ and $\ell_2$ in the average cost case.

Proposition 3.13. *Let $\rho = \frac{\beta(1-\alpha)}{\alpha(1-\beta)} \in (0,1)$, $R = \frac{c_{s_1} + c_{d_2}}{c_{d_1} + c_{s_2}}$ and $\Pi^{T_{\ell_3}}$ be the set of matching policy of threshold type in $\ell_3$ with priority to $\ell_1$ and $\ell_2$. Assume that the cost function is a linear function. The optimal threshold $t^*$, which minimizes the average cost on $\Pi^{T_{\ell_3}}$, is*

$$t^* = \begin{cases} \lceil k \rceil & if\ f(\lceil k \rceil) \leq f(\lfloor k \rfloor) \\ \lfloor k \rfloor & otherwise \end{cases}$$

*where $k = \frac{\log \frac{\rho - 1}{(R+1)\log \rho}}{\log \rho} - 1$ and $f(x) = (c_{d_1} + c_{s_2})x + (c_{d_1} + c_{d_2} + c_{s_1} + c_{s_2})\frac{\rho^{x+1}}{1-\rho} - (c_{d_1} + c_{s_2})\frac{\rho}{1-\rho} + ((c_{d_1} + c_{s_1})\alpha\beta + (c_{d_2} + c_{s_2})(1 - \alpha)(1 - \beta) + (c_{d_2} + c_{s_1})(1 - \alpha)\beta + (c_{d_1} + c_{s_2})\alpha(1 - \beta))$.*

*The threshold $t^*$ is positive.*

Proof. The idea of the proof is to look at the Markov chain derived from the policy $u_t^\infty \in \Pi^{T_{\ell_3}}$. We show that the Markov chain is positive recurrent and we compute the stationary distribution. Using the strong law of large numbers for Markov chains, we show that the average cost $g^{u_t^\infty}$ is equal to the expected cost of the system under the stationary distribution. Then, we find an analytical form for the expected cost which depends on the threshold on $\ell_3$, i.e, $t$. Finally, we minimize the function over $t$. See details in [8, Appendix B]. □

## 4 GENERAL BIPARTITE GRAPHS

In this section, we investigate the optimal matching policy for general bipartite graphs. The natural extension of the optimal matching policy identified in the case N would be to prioritize the end edges. However, the concept of end edges only make sense for acyclic graphs. Therefore, we present first a heuristic for the case of acyclic bipartite graphs, that is an extension of the threshold type policy for the $N$-case (see Definition 3.4).

Our heuristic gives priority to the end edges and puts thresholds on vertices that are incident to those end edges. Then, the policy gives priority to the edges that are adjacent to the end edges and puts thresholds on vertices that are incident to those edges. We repeat this process until all edges are exhausted. This means that our heuristic first matches everything in the end edges. Then it matches everything in the adjacent edges that is above the thresholds (set on the incident vertices), and so on.

For general graphs, our heuristics uses the max-weight policy for the cycles and the heuristic for the acyclic graphs for the remaining of the graph. Priority is given to the nodes that are not a part of a

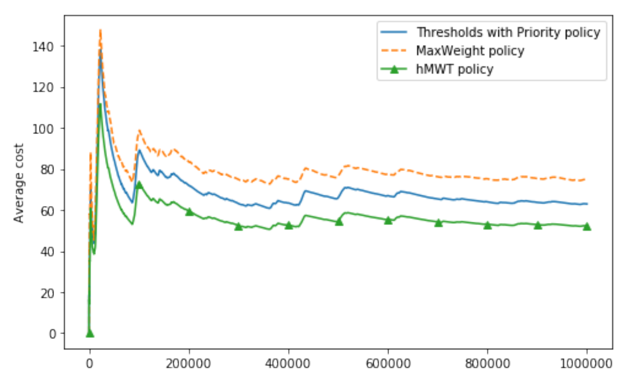


**Figure 3: NN model. Average cost of our heuristic (with a threshold of 0 in $d_1$, 9 in $s_3$ and 0 in $d_2$ and $s_2$), max-weight policy and hMWT the heavy-traffic optimal policy established in [6] (with parameters $D = \{d_3\}$, $\beta = 2$, $\kappa = 10$, $\theta = 1$, $\delta^+ = 0.01$ and $\bar{n}_u = 8$).**

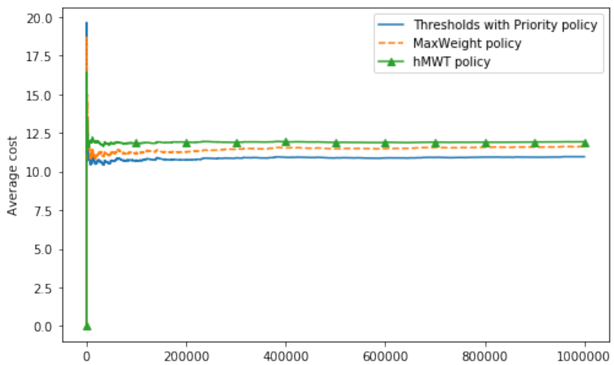


**Figure 4: NN model. Average cost of our heuristic (with a threshold of 0 in $d_1$, 1 in $s_3$ and 0 in $d_2$ and $s_2$), max-weight policy and hMWT the heavy-traffic optimal policy established in [6] (with parameters $D = \{d_3\}$, $\beta = 2$, $\kappa = 10$, $\theta = 1$, $\delta^+ = 0.01$ and $\bar{n}_u = 8$).**

cycle. Ties between two acyclic branches can be broken using the max-weight policy.

We consider the NN model, which is formed by three supply nodes and three demand nodes (see Figure 1). In this model, the end edges are $(d_1, s_1)$ and $(d_3, s_3)$. Thus, our heuristic first matches all $(d_1, s_1)$ and all $(d_3, s_3)$. Then, it matches every $(d_1, s_2)$ above the threshold in $d_1$ and every $(d_2, s_3)$ above the threshold in $s_3$. Finally, it matches every $(d_2, s_2)$ above the threshold in $d_2$ and $s_2$.

We compare the average cost of our heuristic with various policies in two different settings. We consider the max-weight policy which is well known in the literature and performs well in many applications. The decision rule for the max-weight policy is given as the solution of the following optimization problem:

$$u_{MW} = \text{argmax}\{u \cdot \nabla h(x) : u \in U_x\}, \ x \in \mathcal{X},$$

where $h(x) = \sum_{i \in \mathcal{D} \cup \mathcal{S}} c_i x_i^2$. We also consider the hMWT policy which was proved asymptotically optimal in heavy-traffic regime in [6].

For our fist example, we want to compare our heuristic with results from [6] and thus, we choose the same arrival rates and costs. This means that $\alpha_1 = \frac{3}{6}$, $\alpha_2 = \frac{2}{6}$, $\alpha_3 = \frac{1}{6}$, $\beta_1 = \frac{2}{6} - \frac{\delta}{2}$, $\beta_2 = \frac{3}{6} - \frac{\delta}{2}$ and $\beta_3 = \frac{1}{6} + \delta$ (with $\delta = 0.06$) and $c_{d_1} = c_{s_1} = 1$, $c_{d_2} = c_{s_2} = 2$ and $c_{d_3} = c_{s_3} = 3$. In Figure 3, we show the result of the numerical experiments we have performed for this model. As it can be observed, our heuristic performs better than max-weight and is not far from hMWT.

We tested our heuristic in a second example which is not close to a heavy-traffic regime. For the arrival rates, we choose similar rates as in the first example but with $\delta = 0.5$ and for the costs, we choose the same as in the first example. In Figure 4, we show that our heuristic outperforms the other policies.

The Python code for both examples can be found at [7].

## 5 CONCLUSION

The main result of this paper is the complete characterization of an optimal policy for the N-shaped bipartite matching model, that holds for any matching rates. We proved that an optimal matching policy is of threshold type for the diagonal edge and with priority to the end edges of the matching graph. In the case of a more general acyclic matching graph, prioritizing the end edges also seems to be a good policy. We proposed a simple heuristic matching policy for acyclic bipartite graphs. The generalization of the analytical results remains an open question, which is left for future work.

**Acknowledgments.** Funding from the French National Research Agency grant ANR-16-CE05-0008, the Marie Sklodowska-Curie grant agreement No 777778, the Basque Government, Spain, Consolidated Research Group Grant IT649-13, and the Spanish Ministry of Economy and Competitiveness project MTM2016-76329-R.

## REFERENCES

[1] Ivo Adan and Gideon Weiss. 2012. Exact FCFS Matching Rates for Two Infinite Multitype Sequences. *Operations Research* 60, 2 (2012), 475–489. https://doi.org/10.1287/opre.1110.1027
[2] Ivo J. B. F. Adan, Ana Busic, Jean Mairesse, and Gideon Weiss. 2018. Reversibility and Further Properties of FCFS Infinite Bipartite Matching. *Math. Oper. Res.* 43, 2 (2018), 598–621. https://doi.org/10.1287/moor.2017.0874
[3] Itai Ashlagi, Patrick Jaillet, and Vahideh H. Manshadi. 2013. Kidney Exchange in Dynamic Sparse Heterogenous Pools. In *Proceedings of the Fourteenth ACM Conference on Electronic Commerce (EC '13)*. ACM, New York, NY, USA, 25–26. https://doi.org/10.1145/2482540.2482565
[4] Siddhartha Banerjee, Yash Kanoria, and Pengyu Qian. 2018. The Value of State Dependent Control in Ridesharing Systems. *ArXiv e-prints* (March 2018). arXiv:1803.04959
[5] Ana Busic, Varun Gupta, and Jean Mairesse. 2013. Stability of the bipartite matching model. *Advances in Applied Probability* 45, 2 (2013), 351–378. http://arxiv.org/abs/1003.3477
[6] Ana Busic and Sean P. Meyn. 2014. Approximate optimality with bounded regret in dynamic matching models. *ArXiv e-prints* (2014). arXiv:1411.1044 http://arxiv.org/abs/1411.1044
[7] Arnaud Cadas. [n. d.]. Python Code. https://github.com/ArnaudCadas/DynamicMatchingModel.
[8] Arnaud Cadas, Ana Busic, and Josu Doncel. 2018. Optimal Control of Dynamic Bipartite Matching Models. *ArXiv e-prints* (Oct. 2018). arXiv:1810.08541
[9] René Caldentey, Edward H. Kaplan, and Gideon Weiss. 2009. FCFS infinite bipartite matching of servers and customers. *Advances in Applied Probability* 41, 3 (2009), 695–730. http://www.jstor.org/stable/27793900
[10] Jon Feldman, Aranyak Mehta, Vahab S. Mirrokni, and S. Muthukrishnan. 2009. Online Stochastic Matching: Beating 1-1/e. In *Proceedings of the 2009 50th Annual IEEE Symposium on Foundations of Computer Science*, IEEE Computer Society (Ed.). ACM Press, 117–126.
[11] United Network for Organ Sharing. [n. d.]. Online. https://unos.org/wp-content/uploads/unos/Living_Donation_KidneyPaired.pdf
[12] Itai Gurvich and Amy Ward. 2014. On the dynamic control of matching queues. *Stoch. Syst.* 4, 2 (2014), 479–523. https://doi.org/10.1214/13-SSY097
[13] Emmanuel Hyon and Alain Jean-Marie. 2012. Scheduling Services in a Queuing System with Impatience and Setup Costs. *Comput. J.* 55, 5 (2012), 553–563. https://doi.org/10.1093/comjnl/bxq096
[14] Patrick Jaillet and Xin Lu. 2014. Online Stochastic Matching: New Algorithms with Better Bounds. *Mathematics of Operations Research* 39, 3 (2014), 624–646. https://doi.org/10.1287/moor.2013.0621
[15] Jean Mairesse and Pascal Moyal. 2016. Stability of the stochastic matching model. *Journal of Applied Probability* 53, 4 (2016), 1064–1077. https://doi.org/10.1017/jpr.2016.65
[16] Vahideh H. Manshadi, Shayan Oveis Gharan, and Amin Saberi. 2012. Online Stochastic Matching: Online Actions Based on Offline Statistics. *Mathematics of Operations Research* 37, 4 (2012), 559–573. https://doi.org/10.1287/moor.1120.0551
[17] Mohammadreza Nazari and Alexander L. Stolyar. 2018. Reward maximization in general dynamic matching systems. *Queueing Systems* (12 Nov 2018). https://doi.org/10.1007/s11134-018-9593-y
[18] Martin L. Puterman. 2005. *Markov decision processes: discrete stochastic dynamic programming*. Wiley-Interscience.
[19] Robert J. Schalkoff. 1991. *Pattern Recognition: Statistical, Structural and Neural Approaches*. John Wiley & Sons, Inc., New York, NY, USA.
[20] Lenka Zdeborová, Aurélien Decelle, and Michael Chertkov. 2009. Message passing for optimization and control of a power grid: Model of a distribution system with redundancy. *Phys. Rev. E* 80 (Oct 2009), 046112. Issue 4. https://doi.org/10.1103/PhysRevE.80.046112