Fair Optimal Stopping Policy for Matching with Mediator

Yang Liu

yangl@seas.harvard.edu, SEAS, Harvard University

Abstract

In this paper we study an optimal stopping policy for a multi-agent delegated sequential matching system with *fairness constraints*. We consider a setting where a mediator/decision maker matches a sequence of arriving assignments to multiple groups of agents, with agents being grouped according to certain sensitive attributes that needs to be protected. The decision maker aims to maximize total rewards that can be collected from above matching process (from all groups), while making the matching fair among groups. We discuss two types of fairness constraints: (i) each group has a certain expected deadline before which the match needs to happen; (ii) each group would like to have a guaranteed share of average reward from the matching. We present the exact characterization of fair optimal strategies. Example is provided to demonstrate the computation efficiency of our solution.

1 INTRODUCTION

We consider a delegated multi-agent sequential matching problem [Derman et al., 1972, Moscarini, 2005, Le Ny et al., 2006], where the decision maker (mediator) helps decide on matching a set of sequentially arriving assignments to a set of arriving groups of agents. This setting can model various application scenarios in a sequential and dynamical system:

- *Ride sharing*: UBER matches arriving ride requests to drivers from different groups (according to race, gender, neighborhood, income etc.)

- *Job matching*: On employment market, the mediator matches arriving job opportunities to job candidates from different groups (again race, gender, or social status, background etc).

Serving as the proxy mediator, the decision maker receives reward from each successful matching, which realization depends on the quality of matching. For example, the proxy extracts subsidies proportionally from the payments issued for each matched and completed task.

This decision making problem for sequential matching has been studied and has been shown to be solvable elegantly under an optimal stopping rule solution framework [Ferguson, 2006]. This line of works typically focus on finding the optimal stopping strategy by matching agents with assignments judiciously such that the total average reward (over time) being collected is maximized. The optimal matching policy is shown to have a stopping structure, with the core idea of skipping certain matching if the observed reward realization is low. Often the optimal stopping policy has a simple threshold structure w.r.t. the reward realization.

Recently fairness issues have raised a lot of discussions and concerns on designing and implementing different decision making policies, e.g., see discussion of fairness in social decision making [Radke et al., 2012], algorithmic bias in data mining [Hajian et al., 2016], fairness in a bandit decision making framework [Joseph et al., 2016], and fair action within a Markovian decision making framework [Jabbari et al., 2016]. Our problem setting shares similar concerns. For example, when different groups (differ in sensitive features, such as gender, race etc) of agents have different chances of seeing a high reward matching, the decision maker may bias towards *not* matching such agents. Or consider another example where different agents may have different matching constraints that are supposed to be protected under certain policy. Yet from technical perspectives, much less results has been developed towards understanding the issue of fairness constraints for this multi-agent optimal sequential matching problems.

In this paper we set out to explore this topic. We will devote more to the algorithmic aspects of such a problem, rather than the policy level in that we will characterize the optimal stopping policy under a set of *given* fairness constraints within the optimal stopping rule's framework. Specifically we consider the following *fair* multi-agent matching problem. Suppose there are multiple groups of agents trying to be matched with a sequence of arriving assignments in discrete time slots. Agents's availabilities follow an IID process. The available agent will be allowed to be matched to an assignment. Over the successful matches both the decision maker and agents will accumulate rewards. Rewards follow certain distribution and are generated according to an IID process over matching periods. Therefore due to the fact that the reward statistics vary dynamically, the decision maker needs to decide whether to match or not when facing the opportunity depending on the currently observed reward of the potential matching, in order to improve the total rewards collected. For this classical stopping rule setting, we are going to show that the optimal matching strategy gives up matching opportunity when the observed reward statistics is low, and individual groups of agents may suffer from losing matching opportunities. We consider two fairness constrained versions of above problem. For the first one, each group of agents has a deadline, which the group's expected matching time cannot go beyond of. In the second case, each group of agents would like a certain amount of average reward (over the waiting time) to be guaranteed from the matching.

1.1 OUR RESULTS AND CONTRIBUTIONS

Our contributions summarize as follows. (1) We formulate the fairness constrained multi-agent sequential matching problems into a set of constrained optimal stopping ones, and show their equivalences with a set of constrained optimization problems. (2) We characterize the fair optimal stopping policy under delay and reward constraints respectively. Our characterization also reveals the fundamental limits in fair optimal stopping problems, e.g. our results help characterize the feasible region w.r.t. individual fairness constraints. Properties of the derived optimal strategies are also studied. (3) We demonstrate the efficiency of computing the optimal stopping strategy within our solution framework. To our best knowledge, none of the existing works has tackled this type of matching/stopping problem with groups of agents facing individual fairness constraints.

1.2 RELATED WORKS

Fairness issues in decision making and machine learning systems, or more generally in Artificial Intelligence,

have attracted an increasing amount of attentions recently [Hardt, 2011, Dwork et al., 2012, Radke et al., 2012, Joseph et al., 2016, Kleinberg et al., 2017]. More specifically, [Radke et al., 2012] studied the effect of fairness concerns in social decision making processes, while [Joseph et al., 2016] analyzed fairness in candidate option selection within the classical Multi-Armed Bandit decision making framework. While a growing number of works [Hardt, 2011, Dwork et al., 2012, Chouldechova, 2017, Hardt et al., 2016] have studied the fairness issues for one-shot decision making problems, much less has been understood for a fair decision making process. Our results try to fill in this direction.

On a higher level, our matching problem can also be viewed as a multi-agent scheduling problem. [Ferguson, 2006, Xu et al., 2013, Zheng et al., 2009] have established the equivalence between a large category of such a multi-agent scheduling problem with the classical optimal stopping rule problem, where the optimal policy has proven to have a simple characterizable threshold structure. [Ferguson, 2006] gives a good summary of the stopping rule problems for different types (e.g., selfish v.s. collaborative) of agents/users. In [Kennedy, 1982] an optimal stopping problem for multi-agents under delay constraint is considered. Nonetheless the above work considered only the *average delay* (average over all agents) constraint, while the issues with individual *fairness* constraints have not been addressed in this type of scheduling problems.

The rest of our paper is organized as follows. In Section 2, we detail our model and revisit the optimal stopping rule based multi-agent sequential matching problem. Formulations of the sequential matching problems with delay and reward fairness constraints are presented in Section 3. We then state the solutions for above problems in Section 4. In Section 5, we present an specific example to demonstrate the efficiency of computing the optimal strategy. Section 6 concludes the paper.

2 PRELIMINARIES

Consider the following sequential delegated assignment problem. We have $M > 1$ groups of agents facing a sequence of arriving assignments and denote them by the set $\mathcal{U} = \{1, 2, ..., M\}$. We can view the agents as companies looking for job candidates (assignments), or freelancers (UBER drivers) looking for job offers (ride requests). They form groups according to certain attributes, for instance race, gender, or residency. The decision maker (DM) is delegating the matching procedure. Note that the group to be matched is the same as there is an arriving agent from that group to be matched.

The decision making process proceeds as follows. At each time n (discrete time slot), there is an arrival of possible assignment (deterministically) to be matched. We model the availabilities of groups (of agents) as a stochastic process with a constant expected inter-arrival time t_c that $\tau := \mathbb{E}[t_c]$ stays as a constant. Denote by p_i^s the probability that there is an agent from group i is ready to be matched to an assignment upon an arrival time, and $\sum_{i \in \mathcal{U}} p_i^s = 1$ [1].

Both agents and DM are collecting rewards from the matching process. Per each match, reward is generated according to a random variable $R_i \in \mathbb{R}^+$ (non-negative real numbers) for group i, with its cumulative distribution function (CDF) $F_{R_i}(x)$ defined on $x \in \mathbb{R}^+$. Jointly the DM observes reward R:

$$R \sim \sum_i p_i^s F_{R_i}(x) := F_R(x), x \in \mathbb{R}^+.$$

Before each matching, the DM observes the current reward realization $R(\omega)$ (ω denotes a realization of R). Then he makes a decision between two options : *stop* and *continue*. By *stop* the DM will proceed to match the agent with the assignment, which takes one unit time slot to finish. As a consequence, a total reward $\delta \cdot R(\omega), 1 > \delta > 0$ will be collected. δ can be thought of the fraction DM charges for his service. By *continue*, the DM will forgo the current matching opportunity, and the above process repeats by itself. Clearly by *continue* a potentially higher reward may be collected in the future if the current matching happens to be a bad realization. The trade-off is also clear as due to availabilities, coordinating for a new matching takes away non-negligible amount of resources (time in our case). For simplicity of analysis, we assume the matching process repeats over time, and the system clock resets to 0 after each successful match.
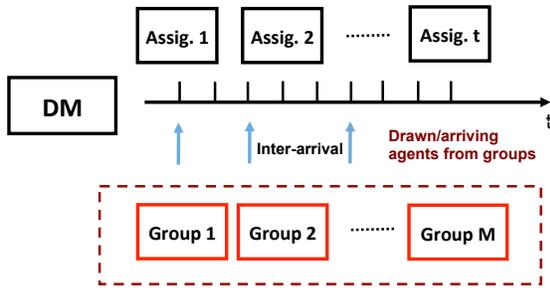

Figure 1: Illustration of the system. The delegated decision maker makes decision on matching arriving assignments to groups of agents.

Example: Consider the ride-sharing example. When ride-sharing company (e.g. UBER) schedules drivers, t_c models the coordination cost, such as time for scheduling drivers (sending notification, making calls etc). The reward R captures the per ride earning. Each individual driver has either an expected delay in receiving a job (delay constraints), or an expected hourly payment (reward constraints).

2.1 UN-CONSTRAINED SEQUENTIAL MATCHING: REVISIT

We now rigorously describe our sequential matching problem with the DM's objective being maximizing the average return of reward over successful matches. For each such successful match, denote by π a matching policy $\pi := \{\eta_1, \eta_2, \cdots \eta_{\gamma(\pi)}\}$, where η_k denotes the k-th action taken by the k-th to-be-matched agent for this particular matching cycle. Notice we have

$$\eta_k = continue, \forall k = 1, \cdots, \gamma(\pi) - 1,$$

and $\eta_{\gamma(\pi)} = stop$. Note decision is only made upon the availability of groups of agents, and $\gamma(\pi)$ denotes the stopping time at which the process terminates with decision being *stop and match*. Denote by $R(k)$ as the realized reward of the matching at k-th decision epoch, and we assume within each matching period, the sequence of rewards $\{R(k)\}_k$ forms an IID process.

Let $R^\pi(k)$ denotes the matching reward obtained during the k-th decision epoch under policy π. Clearly for $k < \gamma(\pi)$ we will have $R^\pi(k) = 0$, that is no reward will be collected when no matching happened, and $R^\pi(\gamma(\pi)) = R(\gamma(\pi))$. Let T_k^π denotes the total amount of time that has been spent up to decision epoch k under policy π, which includes the inter-arrival times (denoting as t_n):

$$T_k^\pi = \begin{cases} \sum_{n=1}^k t_n, \forall k < \gamma(\pi). \\ \sum_{n=1}^k t_n + 1, \ k = \gamma(\pi). \end{cases}$$

The goal is to maximize the average return of reward (average reward per time unit) over the duration of this decision process. This is rigorously stated as follows:

$$J^* = \max_{\pi \in \Pi} \mathbb{E}\left[\frac{\sum_{k=1}^{\gamma(\pi)} \delta \cdot R^\pi(k)}{T_{\gamma(\pi)}^\pi}\right]$$
$$= \max_{\pi \in \Pi} \frac{\mathbb{E}[R^\pi(\gamma(\pi))]}{\mathbb{E}[T_{\gamma(\pi)}^\pi]}, \quad (1)$$

with Π denoting the admissible set of policies. The second equality is due to renewal theory [Ferguson, 2006]. Maximizing this rate-of-return leads to the problem of choosing a stopping policy π to maximize the ratio defined in Eqn.(1).

Classical results [Ferguson, 2006] have established that the optimal strategy for above problem is simply a threshold based stopping policy. Further the threshold can be easily obtained by solving a fixed point equation:

Theorem 2.1 (Chapter 6 of [Ferguson, 2006], [Zheng et al., 2009])**.** *The optimal action for each matching of deciding between* stop *and* continue *is given by a stopping rule: the state space of the reward can be divided into a stopping set Δ^s and continuation set Δ^c, such that whenever the reward is observed to be in either set above, the corresponding action (*stop *v.s.* continue*) is taken. Furthermore these two sets are given by the following threshold property ($\forall$ steps n):*

$$\Delta^s = \{R(n) : R(n) \geq x^*\}, \tag{2}$$

and this threshold x^* at time n is given by the unique solution to the following fixed point equation:

$$\mathbb{E}[R(n) - x]^+ = x \cdot \tau . \tag{3}$$

The optimal stopping time N^* is then given as

$$N^* = \min\{n \geq 1 : R(n) \geq x^*\} . \tag{4}$$

For detailed proof please refer to [Chapter 6, [Ferguson, 2006]],[Zheng et al., 2009] and we will not restate. Also the following iterative algorithm has been developed to compute the exact threshold:

$$x(k+1) := \frac{\int_{x(k)}^{\infty} y\,dF_R(y)}{\tau + [1 - F_R(x(k))]} , \tag{5}$$

with an arbitrarily chosen initial belief $x(0)$. Convergence is defined as observing negligible successive change and is guaranteed [Zheng et al., 2009].

3 FAIR SEQUENTIAL MATCHING

The above results imply that implementing the optimal decision rule via setting an universal threshold may bias towards groups that tend to have higher realized reward statistically. We define our fairness constrained sequential matching problems in this section.

3.1 DELAY FAIR MATCHING

Now we consider a fair matching setting. We start with a matching problem with delay constraints. For each group i, we assume the expected delay before matching one of its agents should be no larger than a certain constant $\nu_i > 0$, and denote $\nu = [\nu_1, \nu_2, ..., \nu_M]$ as the vector form of the constraints. Such fairness constraints may come from regulations, or may come as functions of different groups' protected attributes. Note setting the constraints blindly equal is not regarded as fair. For instance, even for people with the same educational background , which might be a strong indicator of success, fairness constraints may not be necessarily the same for all individuals, as it is likely less so for individuals from a protected subgroup that has fewer financial resources. In fact, such fairness constraints should also take into consideration of the differences in individuals' ability, as well as their commitments for using it. This is similar to the notion of Rawlsian fairness [Braybrooke et al., 1963]. We are not getting into details on how such fairness constraints are generated – instead we assume they are certain outputs of a fairness policy or regulation, and we would like to implement them in our decision making setting. So we will take them as given.

We shorthand the stopping time $\gamma(\pi)$ as N. Denote T_N as the total time spent for one successful match with stopping time N. We first have the following lemma formulating the fairness constrained matching problem.

Lemma 3.1. *The optimal matching strategy under individual group delay constraints is given by a stopping rule with threshold strategy $\mathbf{x}^* = [x_1^*, ..., x_M^*]$ such that for each to-be-matched agent from group i, the decision will be matching if observed reward is no less than x_i^*. $\mathbf{x}^*$ can be characterized by solving the following constrained optimization problem:*

(PD) $\quad \max_{\mathbf{x}} \quad \mathbb{E}[R(N)] - \sum_i p_i^s x_i \mathbb{E}[T_N]$
$\quad\quad s.t. \quad \frac{1 - \sum_{j \in \mathcal{U}} p_j^s F_R(x_j)}{p_i^s[1 - F_R(x_i)]} \mathbb{E}[T_N] \leq \nu_i, \forall i \in \mathcal{U}.$
$\quad\quad\quad\quad x_i \geq 0, \forall i \in \mathcal{U} .$

The proof first borrows results from optimal stopping theory, via which we can transfer the rate-of-return objective to the one in (PD). The constraints are given by formulating a system of equations characterizing agents' delay in matching.

Proof. Following results from [Ferguson, 2006, Tan et al., 2010] we know the rate-of-return problem is equivalent with the optimal stopping problem, which can be equivalently formulated as solving the following optimization problem to obtain a set of thresholds $\mathbf{x} = \{x_1, ..., x_M\}$:

$$\sup_{\mathbf{x}} \mathbb{E}[R(N) - \sum_{i \in \Omega} p_i^s x_i T_N] .$$

Showing this is equivalent with showing $\sup \mathbb{E}[R(N) - \sum_{i \in \Omega} p_i^s x_i T_N]$ returns the optimal rate of reward, which can be done similarly as in [Ferguson, 2006]. We then have the constrained matching problem stated as follows:

$$\max_{\mathbf{x}} \mathbb{E}[R(N)] - \sum_{i \in \mathcal{U}} p_i^s x_i \mathbb{E}[T_N]$$

subject to

$$x_i \geq 0, \ D(N, x_i, \mathbf{x}_{-i}) \leq \nu_i, \forall i \in \mathcal{U},$$

where $D(N, x_i, \mathbf{x}_{-i})$ denotes the expected delay for group i before a (successful)-match, with the threshold policy $\mathbf{x}$ and stopping time N. Now we quantify $D(\cdot)$. Notice the following equation holds:

$$\begin{aligned}
&D(N, x_i, \mathbf{x}_{-i}) \\
&= \tau + p_i^s[(1 - F_R(x_i)) + F_R(x_i)D(N, x_i, \mathbf{x}_{-i})] \\
&\quad + (1 - p_i^s) \sum_{j \neq i} p^s(j|\bar{i})[F_R(x_j)D(N, x_i, \mathbf{x}_{-i}) \\
&\quad + (1 - F_R(x_j))(1 + D(N, x_i, \mathbf{x}_{-i}))] \,,
\end{aligned}$$

where $p^s(j|\bar{i})$ is the probability agent from group $j \neq i$ to be matched (upon the arrival of agents), conditional on the fact group i is not the one to be matched.

We explain the implication of RHS of above equation. The first term corresponds to the expected inter-arrival time. The second term corresponds to the case group i is to be matched: with probability $1 - F_R(x_i)$ agent from group i will be matched immediately while the DM will forgo the match with probability $F_R(x_i)$. The terms associated with the third part correspond to the events when agent from group j other than i is to be matched. From above, re-arrange we have

$$D(N, x_i, \mathbf{x}_{-i}) = \frac{\sum_{j \in \mathcal{U}} p_j^s(1 - F_R(x_j)) + \tau}{p_i^s(1 - F_R(x_i))} \,.$$

Similarly we have (details of explanation omitted)

$$\begin{aligned}
\mathbb{E}[T_N] = \tau &+ \sum_{j \in \mathcal{U}} p_j^s(1 - F_R(x_j)) \\
&+ \sum_{j \in \mathcal{U}} p_j^s F_R(x_j)\mathbb{E}[T_N] \,,
\end{aligned}$$

which further this gives us

$$\mathbb{E}[T_N] = \frac{\sum_{j \in \mathcal{U}} p_j^s(1 - F_R(x_j)) + \tau}{1 - \sum_{j \in \mathcal{U}} p_j^s F_R(x_j)} \,.$$

We thus establish the following

$$D(N, x_i, \mathbf{x}_{-i}) = \frac{1 - \sum_{j \in \mathcal{U}} p_j^s F_R(x_j)}{p_i^s(1 - F_R(x_i))}\mathbb{E}[T_N] \,. \quad (6)$$

Moreover since the thresholds policy can only be non-negative, we have established that the constrained optimal stopping problem is equivalent with (PD).

$\square$

In the constraints, when $1 - F_R(x_i) = 0$ we will follow the convention that $1/0 = +\infty$. The only case that gives

us a $0/0$ in the constraints is when $F_R(x_j) = 1, \forall j$. This is an extreme case stating that the DM will only match a pair when the realization of R reaches its maximum (zero measure event for continuous random variable). We rule out this case.

3.2 REWARD FAIR MATCHING

Now we consider another fair setting with each group requiring to be guaranteed for an average rate of returned reward (e.g., minimum hourly wage). Denote individual group's average (per time unit) expected reward demand as $\mu = [\mu_1, ..., \mu_M]$. Re-denote $\mu_i := \frac{\mu_i}{1 - \delta}$: since each agent only collects $1 - \delta$ fraction of the realized reward of each matching, this is a normalization step to simplify the notation. We similarly prove the following:

Lemma 3.2. *The optimal matching strategy under individual fairness reward constraints is given by a stopping rule policy with thresholds* $\mathbf{x}^* = [x_1^*, ..., x_M^*]$, *which can be characterized by solving the following optimization problem:*

$$\begin{aligned}
\text{(PR)} \quad \max_{\mathbf{x}} \quad &\mathbb{E}[R(N)] - \sum_{i \in \mathcal{U}} p_i^s x_i \mathbb{E}[T_N] \\
\text{s.t.} \quad &\mu_i \frac{1 - \sum_{j \in \mathcal{U}} p_j^s F_R(x_j)}{p_i^s[1 - F_R(x_i)]}\mathbb{E}[T_N] \leq \mathbb{E}[R(N)], \forall i. \\
&x_i \geq 0, \forall i \in \mathcal{U}.
\end{aligned}$$

Proof. By renewal theory, with stopping time N, agent from group i's average expected reward is given by

$$(1 - \delta)\frac{\mathbb{E}[R(N)]}{D(N, x_i, \mathbf{x}_{-i})} \,.$$

Then similar with last section, we have the following problem formulation for obtaining $\mathbf{x}^*$:

$$\begin{aligned}
\max_{\mathbf{x}} \quad &\mathbb{E}[R(N)] - \sum_{i \in \mathcal{U}} p_i^s x_i \mathbb{E}[T_N] \\
\text{s.t.} \quad &\frac{\mathbb{E}[R(N)T]}{D(N, x_i, \mathbf{x}_{-i})} \geq \mu_i, \forall i. \\
&x_i \geq 0, \forall i \in \mathcal{U}.
\end{aligned}$$

Remember μ_i is normalized by $1 - \delta$ in above. Since we have (Eqn.(6))

$$D(N, x_i, \mathbf{x}_{-i}) = \frac{1 - \sum_{j \in \mathcal{U}} p_j^s F_R(x_j)}{p_i^s[1 - F_R(x_i)]}\mathbb{E}[T_N]$$

re-arrange the terms we have the constraints being equivalent with

$$\mu_i \frac{1 - \sum_{j \in \mathcal{U}} p_j^s F_R(x_j)}{p_i^s[1 - F_R(x_i)]}\mathbb{E}[T_N] - \mathbb{E}[R(N)] \leq 0 \,,$$

which finishes the proof. $\square$

4 FAIR OPTIMAL MATCHING

We studied the structure of optimal solutions for delay fair matching, and have shown that solving the fair matching problems is equivalent with solving a set of constrained optimization problems. In this section we present the details towards characterizing the optimal solutions to (PD). We obtain very similar results as for the reward fair matching. The proof is also similar to the case with delay constraints – we leave out the details for a concise presentation.

4.1 WHEN IS FAIRNESS UN-ACHIEVABLE?

We start by presenting a characterization result. Denote

$$\nu_{\mathcal{U}} := (\sum_{i \in \Delta} \frac{1}{\nu_i})^{-1}.$$

$\nu_{\mathcal{U}}$ is the geometric mean of all groups' constraints on average matching delay. We first argue $\nu_{\mathcal{U}}$ controls the feasibility region of the fairness problem.

Lemma 4.1. *(PD) is feasible only if $\nu_{\mathcal{U}} \geq 1 + \tau$.*

Proof. To see this, denote by $D_i(N, \mathbf{x})$ the expected delay for group i before each of its matching with matching policy (stopping time) N and thresholds $\mathbf{x}$, we will have (Eqn.(6) of Proof for Lemma 3.1)

$$D_i(N, \mathbf{x}) = \frac{1 - \sum_{j \in \mathcal{U}} p_j^s F_R(x_j)}{p_i^s(1 - F_R(x_i))} \mathbb{E}[T_N] .$$

From above we easily obtain

$$\sum_{i \in \mathcal{U}} \frac{1}{D_i(N, \mathbf{x})} = \frac{1}{\mathbb{E}[T_N]} .$$

For a feasible solution we have $D_i(N, \mathbf{x}) \leq \nu_i$, and thus

$$\sum_{i \in \mathcal{U}} \frac{1}{D_i(N, \mathbf{x})} \geq \sum_{i \in \mathcal{U}} \frac{1}{\nu_i} .$$

Then we arrive at a necessary condition:

$$\frac{1}{\sum_i \frac{1}{\nu_i}} \geq \mathbb{E}[T_N] \geq 1 + \tau .$$

The second inequality is due to the fact each matching attempt takes up at least 1 (the matching time)+ τ (interarrival time) time units.   □

This result can also be seen as an impossibility results stating the region where the fairness profiles are unable to satisfy. Therefore for this part of study we focus on the case when $\nu_{\mathcal{U}} \geq 1 + \tau$.

4.2 RESULTS

We first introduce a parameter in helping characterize the optimal stopping policy:

Lemma 4.2. *There exists a unique solution for $\nu \geq 1+\tau$ for the following equation:*

$$F_R^{-1}(1 - \frac{\tau}{\nu - 1}) = \frac{1}{\tau} \mathbb{E}[R - F_R^{-1}(1 - \frac{\tau}{\nu - 1})]^+ .$$

Denote the unique solution as ν^*. Simple algebras show that the above fact holds. Intuitively speaking (as well as we will show later) ν^* is a threshold determining the strictness of the set of constraints.[2] When $\nu_{\mathcal{U}} < \nu^*$, we will see the constraints are more strict; while when $\nu_{\mathcal{U}} \geq \nu^*$, the set of delay constraints will be easier to satisfy.

When constraints are tight $(\nu_{\mathcal{U}} < \nu^*)$**:** When constraints are tight we have the following theorem:

Theorem 4.3. *When $\nu_{\mathcal{U}} < \nu^*$, there exists a computable constant $x_{\mathcal{U}}^*$ (computable in $O(1)$ time) such that the optimal solution $\mathbf{x}^* = [x_1^*, ..., x_M^*]$ for (PD) is given by the solution to the following system of equations:*

$$\begin{cases} \frac{1-F_R(x_i)}{1-F_R(x_j)} = \frac{p_j^s \nu_j}{p_i^s \nu_i}, \forall i \neq j, i, j \in \mathcal{U}. \\ \sum_{i \in \mathcal{U}} p_i^s x_i = x_{\mathcal{U}}^* . \end{cases}$$

when the following condition holds

$$\sum_{j \in \mathcal{U}} p_j^s F_R^{-1}\left(1 - \frac{\min_{i \in \mathcal{U}} p_i^s \nu_i}{p_j^s \nu_j}\right) \leq x_{\mathcal{U}}^* .$$

This strikingly clean structure of optimal solution has a very intuitive structural property: roughly speaking the probability of matching (with the assigned threshold) should be inversely proportional to different groups' availability probability, as well as their delay constraints $p_i^s \nu_i$. This result also looks similar to proportional fairness [Kelly et al., 1998], may it be in a resource allocation or a cooperative game setting.

When constraints are loose $(\nu_{\mathcal{U}} \geq \nu^*)$**:** Consider now $\nu_{\mathcal{U}} \geq \nu^*$, in which case the constraints are regarded as being loose (the larger the $\nu_{\mathcal{U}}$ is, the larger the ν_is can be.) The idea for solving this case is simple: we can then afford to make a set of strictly tighter constraints out from the one we have, and reuse the results from the case when $\nu < \nu^*$.

---

[2]For details please refer to the proof.

First denote $\tilde{\nu} = [\tilde{\nu}_1, ..., \tilde{\nu}_M]$ as the solution for the following optimization problem.

$$\text{(TIGHT)} \quad \min_{\tilde{\nu}} \quad \sum_{j \in \mathcal{U}} p_j^s F_R^{-1}(1 - \frac{\min_{i \in \mathcal{U}} p_i^s \tilde{\nu}_i}{p_j^s \tilde{\nu}_j})$$
$$\text{s.t.} \quad 0 \leq \tilde{\nu}_i \leq \nu_i, \forall i \in \mathcal{U}$$
$$\sum_{j \in \mathcal{U}} \frac{1}{\tilde{\nu}_j} = \frac{1}{\nu^* - \epsilon},$$

where $\epsilon > 0$ is an arbitrarily small quantity. This is a "tightening" step that we will refer to. The optimization step is to relax the conditions for the solutions to be non-negative as much as we can. Notice technically we should have $\tilde{\nu}_i > 0$. However $\tilde{\nu}_i = 0$ clearly violates the second equality constraint with an appropriately selected ϵ, which leads to the current equivalent formulation.

Theorem 4.4. *When $\nu_{\mathcal{U}} \geq \nu^*$, there exists a computable constant x_ν^* (computable in $O(1)$ time) such that one optimal solution $\mathbf{x}^* = [x_1^*, ..., x_M^*]$ for (PD) is given by the solution to the following system of equations:*

$$\begin{cases} \frac{1 - F_R(x_i)}{1 - F_R(x_j)} = \frac{p_j^s \tilde{\nu}_j}{p_i^s \tilde{\nu}_i}, \forall i \neq j, i, j \in \mathcal{U}. \\ \sum_{i \in \mathcal{U}} p_i^s x_i = x_\nu^*. \end{cases}$$

when the following condition holds

$$\sum_{j \in \mathcal{U}} p_j^s F_R^{-1}\left(1 - \frac{\min_{i \in \mathcal{U}} p_i^s \tilde{\nu}_i}{p_j^s \tilde{\nu}_j}\right) \leq x_\nu^*.$$

The following remarks end this subsection: (1) Our results extend the ones reported in [Kennedy, 1982]. In particular if there is only one global delay constraint ν, the individual group delay becomes ν/p_i^s (the delay before matching of each particular group). Then $\nu_{\mathcal{U}}$ will degenerate to ν and the results degenerate to the ones reported in [Kennedy, 1982]. Further Theorem 4.3 and 4.4 become the two cases that have been discussed in above work. (2) The second observation is when the constraints are loose ($\nu_{\mathcal{U}} \geq \nu^*$), we could simply make the constraint set to be more strict and solve it according to the case when constraints are tight. However this "tightening" step requires solving another optimization problem. We will show later with specific reward distributions the optimization problem can be solved efficiently. (3) The closed-form of the system of equations characterizing the optimal solution in Theorem 4.3 and 4.4 is not clear so far in terms of computation complexity, e.g., linearity, convexity etc. We will discuss this matter in Section 5.

4.3 MAIN IDEAS OF THE PROOF

We provide the main ideas of the proof for Theorem 4.3 and 4.4[3]. The proof builds on the Lagrange relaxation

---

[3]Full proof can be found in the supplementary materials.

and primal dual analysis of (PD), as well as the optimization formulation of an optimal stopping problem.

(Theorem 4.3) We will first introduce the Lagrange relaxation for (PD):

$$V^* := \mathbb{E}[R(N)] -$$
$$\sum_i (p_i^s x_i + \lambda_i \frac{1 - \sum_j p_j^s F_R(x_j)}{p_i^s[1 - F_R(x_i)]}) \mathbb{E}[T_N]. \quad (7)$$

Following standard stopping rule results [Ferguson, 2006] the optimal stopping time for maximizing V^* is:

$$N^* = \min\{n \geq 1 :$$
$$R(n) \geq \sum_i (p_i^s x_i + \lambda_i \frac{1 - \sum_j p_j^s F_R(x_j)}{p_i^s[1 - F_R(x_i)]}) + V^*.$$

The intuition is that the stopping time is characterized by when the excessive expected return of reward is no less than V^*, which characterizes the expected excessive return from the stopping policy. Further the optimal set of thresholds satisfy the following fixed-point equation:

$$\mathbb{E}\left[R(N) - \sum_i (p_i^s x_i + \lambda_i \frac{1 - \sum_j p_j^s F_R(x_j)}{p_i^s[1 - F_R(x_i)]}) - V^*\right]^+$$
$$= \sum_i (p_i^s x_i + \lambda_i \frac{1 - \sum_j p_j^s F_R(x_j)}{p_i^s[1 - F_R(x_i)]})\tau. \quad (8)$$

The LHS of above equation characterizes the expected additional reward for not stopping. While the RHS is the additional reward collected when stop.

Meanwhile with above characterization, the expected stopping time can be written as (geometric distribution):

$$\mathbb{E}[T_N] = 1 +$$
$$\frac{\tau}{1 - F_R(\sum_i (p_i^s x_i + \lambda_i \frac{1 - \sum_j p_j^s F_R(x_j)}{p_i^s[1 - F_R(x_i)]}) + V^*)}. \quad (9)$$

From KKT condition on the set of threshold policies.

$$\lambda_i^* \frac{1 - \sum_{j \in \mathcal{U}} p_j^s F_R(x_j)}{p_i^s[1 - F_R(x_i)]} \mathbb{E}[T_N] - \lambda_i^* \nu_i = 0, \forall i \in \mathcal{U}. \quad (10)$$

Plug Eqn.(9) into Eqn.(10), we will be able to show

$$p_i^s[1 - F_R(x_i)]\nu_i = C(\mathbf{x}, \lambda),$$

where $C(\mathbf{x}, \lambda)$ is a constant that is independent of group index i. From above we know that for $\lambda_i^*, \lambda_j^* > 0, i \neq j$ we have

$$\frac{1 - F_R(x_i)}{1 - F_R(x_j)} = \frac{p_j^s \nu_j}{p_i^s \nu_i}.$$

This give us the first set of equations in Theorem 4.3. For our primal and dual problems we have

$$L(\mathbf{x}^*, \lambda^*) := V^* + \lambda^{*\mathbf{T}}\nu = 0 . \tag{11}$$

$\lambda^{*\mathbf{T}}\nu$ can further be calculated as follows:

$$\lambda_i^* \nu_i = \lambda_i^* \frac{1 - \sum_j p_j^s F_R(x_j)}{p_i^s[1 - F_R(x_i)]} \cdot \frac{p_i^s[1 - F_R(x_i)]}{1 - \sum_j p_j^s F_R(x_j)} \nu_i .$$

Plug in V^* (Eqn.(7)) and above into Eqn. (11) we have

$$\exists\, x_{\mathcal{U}}^*, \text{ s.t. } \sum_{i \in \mathcal{U}} p_i^s x_i = x_{\mathcal{U}}^* .$$

(Theorem 4.4) Consider $\nu_{\mathcal{U}} \geq \nu^*$, i.e., not all constraints are active. Notice since $\frac{1}{\sum_{i \in \mathcal{U}} \frac{1}{\nu_i}} \geq \nu^*$, we could find a set of $\tilde{\nu}$ such that

$$0 \leq \tilde{\nu}_i \leq \nu_i, \forall i \in \mathcal{U}, \text{ and } \frac{1}{\sum_{i \in \mathcal{U}} \frac{1}{\tilde{\nu}_i}} < \nu^* .$$

In particular choose a small $\epsilon > 0$ and set $\frac{1}{\sum_{i \in \mathcal{U}} \frac{1}{\tilde{\nu}_i}} = \nu^* - \epsilon$. Under this case the average optimal reward rate x_ν^* ($O(1)$ computable) associated with $\nu^* - \epsilon$ is achieved. Since we have a set of more strict constraints we know the original constraints $[\nu_1, ..., \nu_M]$ will all be met if the new one gets satisfied. The rest of the proof is similar to the one for Theorem 4.3 and will be omitted.

Meanwhile we would like to choose the set of $\tilde{\nu}$ that will make the constraint

$$\sum_{i \in \mathcal{U}} F_R^{-1}\left(1 - \frac{\min_{j \in \mathcal{U}} p_j^s \tilde{\nu}_j}{p_i^s \tilde{\nu}_i}\right) \leq x_\nu^*$$

less strict, i.e., we choose the set of $\tilde{\nu}$ minimizing $\sum_{i \in \mathcal{U}} F_R^{-1}\left(1 - \frac{\min_{j \in \mathcal{U}} p_j^s \tilde{\nu}_j}{p_i^s \tilde{\nu}_i}\right)$ and we have the following optimization problem for charactering $[\tilde{\nu}_1, ..., \tilde{\nu}_M]$:

$$\min_{\tilde{\nu}} \quad \sum_{j \in \mathcal{U}} p_j^s F_R^{-1}\left(1 - \frac{\min_{i \in \mathcal{U}} p_i^s \tilde{\nu}_i}{p_j^s \tilde{\nu}_j}\right)$$
$$\text{s.t.} \quad 0 \leq \tilde{\nu}_i \leq \nu_i, \forall i \in \mathcal{U}$$
$$\sum_{j \in \mathcal{U}} \frac{1}{\tilde{\nu}_j} = \frac{1}{\nu^* - \epsilon}.$$

4.4 REWARD FAIR MATCHING

We show the results we proved for the delay fairness case can be largely extended to the reward fairness case. First denote the maximum reward rate that can be achieved by the group of agents for un-constrained rate-of-return problem as x^* (solution to Eqn.(3)). Obviously when $\sum_{i \in \mathcal{U}} \mu_i > x^*$, there is no solution to the constrained system. This is because we know the objective value of a constrained setting can never exceed the one for its un-constrained counterpart. We then focus on the case with $\sum_{i \in \mathcal{U}} \mu_i \leq x^*$. We have the following theorems leading to the characterization of the fair optimal solution.

When constraints are tight $(\sum_{i \in \mathcal{U}} \mu_i = x^*)$:

Theorem 4.5. *When* $\sum_{i \in \mathcal{U}} \mu_i = x^*$, *the optimal solution* $\mathbf{x}^* = [x_1^*, ..., x_M^*]$ *for* (PR) *is given by the solution to the following system of equations:*

$$\begin{cases} \frac{1 - F_R(x_i)}{1 - F_R(x_j)} = \frac{p_j^s \mu_i}{p_i^s \mu_j}, \forall i \neq j, i, j \in \mathcal{U}. \\ \sum_{i \in \mathcal{U}} p_i^s x_i = x^* . \end{cases}$$

when the following condition holds

$$\sum_{j \in \mathcal{U}} p_j^s F_R^{-1}\left(1 - \frac{\mu_j/p_j^s}{\max_{i \in \mathcal{U}} \mu_i/p_i^s}\right) \leq x^* .$$

When constraints are loose $(\sum_{i \in \mathcal{U}} \mu_i < x^*)$: When $\sum_i \mu_i < x^*$, we similarly have a less restricted scenario. The constraints will be firstly tightened. Denote $\tilde{\mu} = [\tilde{\mu}_1, ..., \tilde{\mu}_M]$ as the solution for the following optimization problem:

$$\min_{\tilde{\mu}} \quad \sum_{j \in \mathcal{U}} p_j^s F_R^{-1}\left(1 - \frac{\tilde{\mu}_j/p_j^s}{\max_{i \in \mathcal{U}} \tilde{\mu}_i/p_i^s}\right)$$
$$\text{s.t.} \quad \tilde{\mu}_i \geq \mu_i, \forall i \in \mathcal{U}.$$
$$\sum_{j \in \mathcal{U}} \tilde{\mu}_j = x^*.$$

And then we have the following set of results:

Theorem 4.6. *When* $\sum_i \mu_i < x^*$, *one optimal solution* $\mathbf{x}^* = [x_1^*, ..., x_M^*]$ *for* (PR) *is given by the solution to the following system of equations:*

$$\begin{cases} \frac{1 - F_R(x_i)}{1 - F_R(x_j)} = \frac{p_j^s \tilde{\mu}_i}{p_i^s \tilde{\mu}_j}, \forall i \neq j, i, j \in \mathcal{U} . \\ \sum_{i \in \mathcal{U}} p_i^s x_i = x^* . \end{cases}$$

when the following condition holds

$$\sum_{j \in \mathcal{U}} p_j^s F_R^{-1}\left(1 - \frac{\tilde{\mu}_j/p_j^s}{\max_{i \in \mathcal{U}} \tilde{\mu}_i/p_i^s}\right) \leq x^* .$$

It is worth to note though (PD) and (PR) appear to be different, their solution structures are similar. The two problems can then share the same solver with setting different parameters (replacing $(\{\nu_i\}, x_{\mathcal{U}}^*)$ with $(\{\mu_i\}, x^*)$ to reach (PR) from (PD)).

5 COMPUTATION AND EXAMPLES

Our main results have reduced the original constrained stopping rule problem to solving a set of closed-form equations. Despite the clean structure, solving the problem will further rely on the reward distributions, where the computation issue merits a further clarification. We could show these equations reduce to system of *linear*

equations for simple distributions such as exponential distribution and uniform distribution etc. Take exponential distribution for example (for some ρ): $F_R(x) = 1 - \exp(-\rho x), \forall x \geq 0$, with $F_R^{-1}(y) = -\log(1 - y)/\rho$, $0 \leq y \leq 1$. We take the delay constrained case for demonstration, and we start with the case $\nu_{\mathcal{U}} < \nu^*$. From $\frac{1-F_R(x_i)}{1-F_R(x_j)} = \frac{p_j^s \nu_j}{p_i^s \nu_i}$ we know

$$x_i - x_j = -1/\rho \cdot \log p_j^s \nu_j / p_i^s \nu_i . \qquad (12)$$

Meanwhile we know $\sum_{i \in \mathcal{U}} p_i^s x_i = x_{\mathcal{U}}^*$. Combining this and Eqn.(12) returns us a linear system of equations for all x_i; we can then solve for $\mathbf{x}^*$. This can be done in polynomial time.

For the case $\nu_{\mathcal{U}} \geq \nu^*$, we need to solve the optimization problem (TIGHT). Make the following substitute: $y_i = \frac{\min_{j \in \mathcal{U}} p_j^s \tilde{\nu}_j}{p_i^s \tilde{\nu}_i}$, $y_{\min} = \min_{j \in \mathcal{U}} p_j^s \tilde{\nu}_j$, and the constraint $\sum_{j \in \mathcal{U}} \frac{1}{\tilde{\nu}_j} = \frac{1}{\nu^* - \epsilon}$ then becomes $\sum_{j \in \mathcal{U}} p_j^s \cdot y_j = \frac{y_{\min}}{\nu^* - \epsilon}$. And the constraints $\tilde{\nu}_i \leq \nu_i, \forall i \in \mathcal{U}$ become $\frac{y_{\min}}{p_i^s y_i} \leq \nu_i, \forall i \in \mathcal{U}$, which is further equivalent with $y_{\min} \leq p_i^s y_i \nu_i$. Hence (TIGHT) becomes equivalent with the one below:

$$\begin{aligned} \min \quad & -\textstyle\sum_{i \in \mathcal{U}} p_j^s \cdot \log y_i / \rho \\ \text{s.t.} \quad & 0 \leq y_{\min} \leq p_i^s y_i \nu_i, \forall i \in \mathcal{U} \\ & \textstyle\sum_{j \in \mathcal{U}} p_j^s \cdot y_j = \frac{y_{\min}}{\nu^* - \epsilon}. \end{aligned}$$

Since $-\log x$ is a convex function, the above programming problem is convex and can be solved efficiently.

A sub-optimal but easy-to-compute solution It is worth to mention for the case with $\nu_{\mathcal{U}} \geq \nu^*$ we can also easily obtain a feasible solution that is sub-optimal by setting: $\tilde{\nu} = \frac{\nu^* - \epsilon}{\nu_{\mathcal{U}}} \nu$, and it is easy to verify under this case solving the system of equations reduces to the case with $\nu_{\mathcal{U}} < \nu^*$, which is much easier to solve.

In summary, solving for the optimal solution only requires solving (1) a linear system of equation and (2) a convex optimization problem. Existing solver or methods (e.g., interior point etc) can then be applied to obtain a solution in polynomial time.

5.1 SIMULATION

We numerically simulate a simple two groups scenario. For parameter settings please refer to caption of Figure 2. The figure pictures the heat map for expected reward under different delay constraints: dark blue indicates infeasible area, and dark red implies reaching the maximum. Along with relaxing constraints, our solution gradually converges to the un-constrained optimal ones. It is clear our solution help characterize the trade-off between fairness and social optimality.
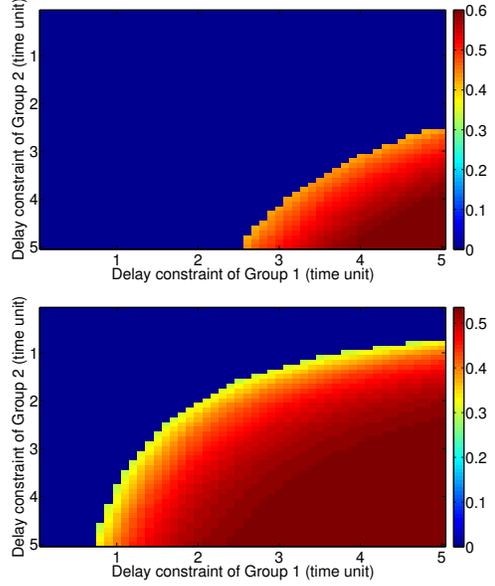

Figure 2: Fairness satisfiability & reward (heat map) for different combination of delay constraints ν_1, ν_2. $p_1^s = p_2^s = 1/2, \rho = 0.05$ (parameter for exp. distribution). Top: $\tau = 1$, max reward = 0.6006 (un-constrained). Bottom: $\tau = 0.5$, max reward = 0.5299 (un-constrained).

6 CONCLUSION

In this paper, we discussed optimal multi-agent matching strategies with individual fairness constraints. We first show the equivalence between this problem and the locally constrained optimal stopping problem, and then a set of constrained programings. We considered two different types of constraints: one with matching deadlines (delay) and the other with reward requirements. We characterized the fair optimal solutions for above two problems respectively. Via practical example, we demonstrated that our solutions can be computed efficiently. The reported results can also be leveraged to obtain scheduling policies with other fairness concerns, by setting the local constraints appropriately.

Our work aims to kick off fairness studies in the sequential decision making framework. One interesting and immediate followup would be to study a fair policy when the decisions made in the past will in fact affect different groups' status (e.g., reward statistics, availability etc) in the future.

References

Cyrus Derman, Gerald J Lieberman, and Sheldon M Ross. A sequential stochastic assignment problem. *Management Science*, 18(7):349–355, 1972.

Giuseppe Moscarini. Job matching and the wage distribution. *Econometrica*, 73(2):481–516, 2005.

Jerome Le Ny, Munther Dahleh, and Eric Feron. Multi-agent task assignment in the bandit framework. In *Decision and Control, 2006 45th IEEE Conference on*, pages 5281–5286. IEEE, 2006.

Thomas S. Ferguson. Optimal stopping and applications. *Mathematics Department, UCLA*, 2006.

Sina Radke, Berna Güroğlu, and Ellen RA de Bruijn. There's something about a fair split: intentionality moderates context-based fairness considerations in social decision-making. *PloS One*, 7(2):e31491, 2012.

Sara Hajian, Francesco Bonchi, and Carlos Castillo. Algorithmic bias: From discrimination discovery to fairness-aware data mining. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 2125–2126. ACM, 2016.

Matthew Joseph, Michael Kearns, Jamie Morgenstern, and Aaron Roth. Fairness in learning: Classic and contextual bandits. *arXiv preprint arXiv:1605.07139*, 2016.

Shahin Jabbari, Matthew Joseph, Michael Kearns, Jamie Morgenstern, and Aaron Roth. Fair learning in markovian environments. *arXiv preprint arXiv:1611.03071*, 2016.

Moritz AW Hardt. *A study of privacy and fairness in sensitive data analysis*. PhD thesis, Princeton University, 2011.

Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *ITCS*, pages 214–226. ACM, 2012.

Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. Inherent trade-offs in the fair determination of risk scores. *ITCS*, 2017.

Alexandra Chouldechova. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *arXiv preprint arXiv:1703.00056*, 2017.

Moritz Hardt, Eric Price, Nati Srebro, et al. Equality of opportunity in supervised learning. In *Advances in Neural Information Processing Systems*, pages 3315–3323, 2016.

Yuhua Xu, Alagan Anpalagan, Qihui Wu, Liang Shen, Zhan Gao, and Jinglong Wang. Decision-theoretic distributed channel selection for opportunistic spectrum access: Strategies, challenges and solutions. *Communications Surveys & Tutorials, IEEE*, 15(4):1689–1713, 2013.

Dong Zheng, Weiyan Ge, and Junshan Zhang. Distributed opportunistic scheduling for ad hoc networks with random access: an optimal stopping approach. *IEEE Trans. Inf. Theor.*, 55(1):205–222, January 2009. ISSN 0018-9448.

DP Kennedy. On a constrained optimal stopping problem. *Journal of Applied Probability*, pages 631–641, 1982.

David Braybrooke, Charles Edward Lindblom, et al. strategy of decision; policy evaluation as a social process. 1963.

Sheu-Sheu Tan, Dong Zheng, Junshan Zhang, and James Zeidler. Distributed opportunistic scheduling for ad-hoc communications under delay constraints. In *INFOCOM'10*, pages 2874–2882, Piscataway, NJ, USA, 2010. IEEE Press. ISBN 978-1-4244-5836-3.

Frank P Kelly, Aman K Maulloo, and David KH Tan. Rate control for communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research society*, 49(3):237–252, 1998.

$$L(x, \lambda) = \mathbb{E}[R(N)] - \sum_{i \in \mathcal{U}} p_i^s x_i \mathbb{E}[T_N] - \sum_{i \in \mathcal{U}} \lambda_i \left(\frac{1 - \sum_j p_j^s F_R(x_j)}{p_i^s[1 - F_R(x_i)]} \mathbb{E}[T_N] - \nu_i \right)$$

$$= \mathbb{E}[R(N)] - \sum_{i \in \mathcal{U}} \left(p_i^s x_i + \lambda_i \frac{1 - \sum_j p_j^s F_R(x_j)}{p_i^s[1 - F_R(x_i)]} \right) \mathbb{E}[T_N] + \sum_{i \in \mathcal{U}} \lambda_i \nu_i \, . \tag{13}$$

$$0 = -\lambda_i^* \nu_i + \lambda_i^* \frac{1 - \sum_j p_j^s F_R(x_j)}{p_i^s[1 - F_R(x_i)]} \cdot \left[\frac{\tau}{1 - F_R(\sum_i (p_i^s x_i + \lambda_i^* \frac{1 - \sum_j p_j^s F_R(x_j)}{p_i^s[1 - F_R(x_i)]}) + V^*)} + 1 \right] \, . \tag{14}$$

$$\Rightarrow p_i^s[1 - F_R(x_i)]\nu_i = \left(1 - \sum_j p_j^s F_R(x_j) \right) \cdot \left[\frac{\tau}{1 - F_R(\sum_i (p_i^s x_i + \lambda_i^* \frac{1 - \sum_j p_j^s F_R(x_j)}{p_i^s[1 - F_R(x_i)]}) + V^*)} + 1 \right] \, . \tag{15}$$

Supplementary materials: Proof of Theorem 4.3

To solve the constrained optimization problem, we introduce its corresponding Lagrange relaxation in Eqn.(13). Denote

$$V^* := \mathbb{E}[R(N)] - \sum_i \left(p_i^s x_i + \lambda_i \frac{1 - \sum_j p_j^s F_R(x_j)}{p_i^s[1 - F_R(x_i)]} \right) \mathbb{E}[T_N] \, .$$

Following previous work [Ferguson, 2006] we know the optimal stopping time for maximizing V^* is given by

$$N^* = \min\{n \geq 1 : R(n) \geq \sum_i (p_i^s x_i + \lambda_i \frac{1 - \sum_j p_j^s F_R(x_j)}{p_i^s[1 - F_R(x_i)]}) + V^*\} \, ,$$

and the optimal set of thresholds satisfy the following fixed-point equation:

$$\mathbb{E}\left[R(N) - \sum_i (p_i^s x_i + \lambda_i \frac{1 - \sum_j p_j^s F_R(x_j)}{p_i^s[1 - F_R(x_i)]}) - V^* \right]^+ = \sum_i (p_i^s x_i + \lambda_i \frac{1 - \sum_j p_j^s F_R(x_j)}{p_i^s[1 - F_R(x_i)]})\tau \, . \tag{16}$$

Again following the results of [Ferguson, 2006], we are able to characterize the expected stopping time as follows,

$$\mathbb{E}[T_N] = \frac{\tau}{1 - F_R(\sum_i (p_i^s x_i + \lambda_i \frac{1 - \sum_j p_j^s F_R(x_j)}{p_i^s[1 - F_R(x_i)]}) + V^*)} + 1 \, . \tag{17}$$

Consider the dual problem $L(\mathbf{x}, \lambda) = V^* + \lambda^\mathbf{T}\nu$. By complementary slackness we have:

$$\lambda_i^* \frac{1 - \sum_{j \in \mathcal{U}} p_j^s F_R(x_j)}{p_i^s[1 - F_R(x_i)]} \mathbb{E}[T_N] - \lambda_i^* \nu_i = 0, \forall i \in \mathcal{U}. \tag{18}$$

Substitute (17) into (18) we have $\forall i \in \mathcal{U}$, we have Eqn.(14).

We first consider active constraints, i.e., constraints with $\lambda_i^* > 0$. (recall in the dual problem we have the variables $\lambda_i^* \geq 0, \forall i \in \mathcal{U}$.) For these active constraints we will get Eqn.(15). Notice the RHS of above equation does not depend on any specific index i. Therefore we have

$$p_i^s[1 - F_R(x_i)]\nu_i = C(\mathbf{x}, \lambda^*) \, ,$$

where $C(\mathbf{x}, \lambda^*)$ is a constant that is independent of indices. Moreover for $\lambda_i^*, \lambda_j^* > 0, i \neq j$ we have

$$\frac{1 - F_R(x_i)}{1 - F_R(x_j)} = \frac{p_j^s \nu_j}{p_i^s \nu_i} \, .$$

Denote $\nu(\mathbf{x}) := \frac{p_i^s[1-F_R(x_i)]\nu_i}{1-\sum_j p_j^s F_R(x_j)}$, we will get

$$\sum_i \left(p_i^s x_i + \lambda_i^* \frac{1-\sum_j p_j^s F_R(x_j)}{p_i^s[1-F_R(x_i)]} \right) + V^* = F_X^{-1}(1 - \frac{\tau}{\nu(\mathbf{x})-1}) \ .$$

Plug above to Eqn.(16), the following establishes

$$\mathbb{E}\Big[R(N) - F_R^{-1}(1 - \frac{\tau}{\nu(\mathbf{x})-1})\Big]^+ = \sum_i (p_i^s x_i + \lambda_i^* \frac{1-\sum_j p_j^s F_R(x_j)}{p_i^s[1-F_R(x_i)]})\tau \ ,$$

and

$$V^* = F_R^{-1}(1 - \frac{\tau}{\nu(\mathbf{x})-1}) - \frac{1}{\tau}\cdot\mathbb{E}\Big[R(N) - F_R^{-1}(1 - \frac{\tau}{\nu(\mathbf{x})-1})\Big]^+ \ .$$

Also for our primal and dual problems we have

$$0 = L(\mathbf{x}^*, \lambda^*) = V^* + \lambda^{*\mathbf{T}}\nu \ .$$

Consider the $\lambda^{*\mathbf{T}}\nu$ term. Notice

$$\lambda_i^* \nu_i = \lambda_i^* \frac{1-\sum_j p_j^s F_R(x_j)}{p_i^s[1-F_R(x_i)]} \cdot \frac{p_i^s[1-F_R(x_i)]}{1-\sum_j p_j^s F_R(x_j)}\nu_i$$

$$= \nu(\mathbf{x}) \cdot \lambda_i^* \frac{1-\sum_j p_j^s F_R(x_j)}{p_i^s[1-F_R(x_i)]} \ .$$

Therefore we have

$$\lambda^{*\mathbf{T}}\nu = \nu(\mathbf{x}) \cdot \sum_i \lambda_i^* \frac{1-\sum_j p_j^s F_R(x_j)}{p_i^s[1-F_R(x_i)]}$$

$$= \nu(\mathbf{x}) \cdot \Big[\frac{\mathbb{E}[R(N) - F_R^{-1}(1 - \frac{\tau}{\nu(\mathbf{x})-1})]^+}{\tau} - \sum_i p_i^s x_i\Big] \ .$$

Moreover we have Eqn.(19):

$$F_R^{-1}(1 - \frac{\tau}{\nu(\mathbf{x})-1}) - \frac{1}{\tau}\cdot\mathbb{E}\Big[R(N) - F_R^{-1}(1 - \frac{\tau}{\nu(\mathbf{x})-1})\Big]^+$$

$$+ \nu(\mathbf{x}) \cdot \left\{ \frac{\mathbb{E}\Big[R(N) - F_R^{-1}(1 - \frac{\tau}{\nu(\mathbf{x})-1})\Big]^+}{\tau} - \sum_i p_i^s x_i \right\} = 0 \ . \tag{19}$$

Notice the following holds:

$$\nu(\mathbf{x}) = \frac{p_i^s[1-F_R(x_i)]\nu_i}{1-\sum_j p_j^s F_R(x_j)} = \frac{p_i^s[1-F_R(x_i)]\nu_i}{\sum_j p_j^s(1-F_R(x_j))} \ ,$$

where the second equality is due to $\sum_j p_j^s = 1$. Substitute in

$$1 - F_R(x_j) = \frac{p_i^s\nu_i}{p_j^s\nu_j}(1 - F_R(x_i))$$

we have

$$\nu(\mathbf{x}) = \frac{p_i^s[1 - F_R(x_i)]\nu_i}{\sum_j p_j^s \frac{p_i^s \nu_i}{p_j^s \nu_j}(1 - F_R(x_i))} = \frac{1}{\sum_{j \in \mathcal{U}} \frac{1}{\nu_j}} = \nu_{\mathcal{U}} ,$$

which is a constant. From which we have

$$F_R^{-1}(1 - \frac{\tau}{\nu(\mathbf{x}) - 1}) = C_1 ,$$

$$\frac{1}{\tau} \cdot \mathbb{E}[R(N) - F_R^{-1}(1 - \frac{\tau}{\nu(\mathbf{x}) - 1})]^+ = C_2 ,$$

with C_1, C_2 being some constant. Notice $C_1 - C_2$ is our previously defined function g evaluated at $\nu(\mathbf{x})$. Therefore we have

$$C_1 - C_2 + \nu_{\mathcal{U}}(C_2 - \sum_i p_i^s x_i) = 0 .$$

Theoretically with the other equations $\frac{1 - F_R(x_i)}{1 - F_R(x_j)} = \frac{p_j^s \nu_j}{p_i^s \nu_i}$ we can solve $\mathbf{x}^*$. The above equation can be reduced to

$$\sum_{i \in \mathcal{U}} p_i^s x_i = x_{\mathcal{U}}^* .$$

The rest is to check whether the corresponding λ^*s are positive. This is equivalent with deciding whether the current constraints are active or not. In fact from previous derivations we know

$$\lambda^T \nu = \frac{\mathbb{E}[R(N) - F_R^{-1}(1 - \frac{\tau}{\nu_{\mathcal{U}} - 1})]^+}{\tau} - \sum_i p_i^s x_i^*$$

therefore we could derive a positive set of λ if the following holds,

$$\frac{\mathbb{E}[R(N) - F_R^{-1}(1 - \frac{\tau}{\nu_{\mathcal{U}} - 1})]^+}{\tau} - \sum_i p_i^s x_i^* > 0$$

which further gives us

$$\frac{1}{\tau}\mathbb{E}[R(N) - F_R^{-1}(1 - \frac{\tau}{\nu_{\mathcal{U}} - 1})]^+ - F_R^{-1}(1 - \frac{\tau}{\nu_{\mathcal{U}} - 1}) > 0 ,$$

since ν^* is the solution for above equation when equality holds, and by simple algebra we could show for $\nu_{\mathcal{U}} < \nu^*$ the above condition holds.

Now we show under what condition the solutions $\mathbf{x}^*$ is non-negative. Note from

$$\frac{1 - F_R(x_i)}{1 - F_R(x_j)} = \frac{p_j^s \nu_j}{p_i^s \nu_i} ,$$

we know

$$x_i = F_R^{-1}\left(1 - [1 - F_R(x_j)]\frac{p_j^s \nu_j}{p_i^s \nu_i}\right) .$$

Not hard to see the smallest x_j associated with the minimum $p_j^s \nu_j$. Therefore

$$\sum_{i \in \mathcal{U} \backslash j} F_R^{-1}\left(1 - [1 - F_R(x_j)]\frac{\min_{j \in \mathcal{U}} p_j^s \nu_j}{p_i^s \nu_i}\right) + x_j = x_{\mathcal{U}}^* .$$

In order to make sure $\mathbf{x}^*$ is non-negative, we need $x_j \geq 0$, i.e.,

$$\sum_{i \in \mathcal{U}} F_R^{-1}\left(1 - \frac{\min_{j \in \mathcal{U}} p_j^s \nu_j}{p_i^s \nu_i}\right) \leq x_{\mathcal{U}}^* .$$