A PROOF OF THEOREMS

A.1 Proof of Theorem 3

Proof. First, we write the original problem of Equation (7):

$$\begin{aligned} Y_{OPT} = \max_{x \in \mathcal{X}} \quad & a(n;\boldsymbol{x}) = \frac{1}{\boldsymbol{r}_n} \sum_{t=1}^{\boldsymbol{r}_n} \boldsymbol{s}_{u_t}^\top \boldsymbol{x}_t \\ \text{s.t.} \quad & \boldsymbol{E}_n^\top \geq \boldsymbol{M}_n^\top, \\ & \boldsymbol{E}_n^\top = \sum_{t=1}^{\boldsymbol{r}_n} \boldsymbol{A}^\top \boldsymbol{x}_t. \end{aligned}$$

Then, we use the Lagrange multiplier $\boldsymbol{\lambda} \in \mathbb{R}_+^{|\mathcal{P}|}$ to move the first constraint to the objective:

$$\begin{aligned} Y_{OPT} =& \max_{\boldsymbol{x} \in \mathcal{X}} \min_{\boldsymbol{\lambda} \geq 0} \quad a(n;\boldsymbol{x}) - \boldsymbol{\lambda}^\top (\boldsymbol{M}_n^\top - \boldsymbol{E}_n^\top) \\ \leq Y_{Dual} =& \min_{\boldsymbol{\lambda} \geq 0} \max_{\boldsymbol{x} \in \mathcal{X}} \quad a(n;\boldsymbol{x}) - \boldsymbol{\lambda}^\top (\boldsymbol{M}_n^\top - \boldsymbol{E}_n^\top) \\ \text{s.t.} \quad & \boldsymbol{E}_n^\top = \sum_{t=1}^{\boldsymbol{r}_n} \boldsymbol{A}^\top \boldsymbol{x}_t, \forall p \in \mathcal{P}, \end{aligned}$$

where $\boldsymbol{\lambda} \in \mathbb{R}_+^{|\mathcal{P}|}$ are the Lagrangian multiplier.

Since the Lagrangian multiplier $\boldsymbol{\lambda}$ is still hard to optimize since Equation A.1 is a constrained non-linear optimization problem w.r.t. $\boldsymbol{\lambda}$. Therefore, we adopt an adaptive penalty function $p_{\boldsymbol{\lambda}}(\boldsymbol{E}_n^\top)$ to transform the constraint into the objective, which is commonly used in linear programming:

$$\begin{aligned} W_{OPT} = \max_{x \in \mathcal{X}} \quad & a(n;\boldsymbol{x}) + p_{\boldsymbol{\lambda}}(\boldsymbol{E}) \\ \text{s.t.} \quad & \boldsymbol{E}_n^\top = \sum_{t=1}^{\boldsymbol{r}_n} \boldsymbol{A}^\top \boldsymbol{x}_t, \end{aligned}$$

where the penalty has the form:

$$p_{\boldsymbol{\lambda}}(\boldsymbol{E}_n^\top) = -\sum_{i=1}^{|\mathcal{P}|} \boldsymbol{\lambda}_i \left[\boldsymbol{M}_{n,i}^\top - \boldsymbol{E}_{n,i}^\top\right]_+,$$

where $\boldsymbol{\lambda} \in \mathcal{R}^{|\mathcal{P}|}$ is a violation penalty vector, and $\boldsymbol{\lambda}_p$ implies a penalty on the providers p for failing to meet the fairness constraint $\boldsymbol{E}_n^\top \geq \boldsymbol{M}_n^\top$.

Then, for the optimization problem in Equation A.1, we move the constraints to the objective using the Lagrange multipliers $\boldsymbol{\mu} \in \mathcal{R}^{|\mathcal{P}|}$:

$$\begin{aligned} W_{OPT} =& \max_{x_t \in \mathcal{X}} \min_{\boldsymbol{\mu} \in \mathcal{D}} \left[a(n;\boldsymbol{x}) + p_{\boldsymbol{\lambda}}(\boldsymbol{E}_n^\top) + \boldsymbol{\mu}^\top \left(\boldsymbol{E}_n^\top - \sum_{t=1}^{\boldsymbol{r}_n} \mathbf{A}^\top \boldsymbol{x}_t\right)\right] \\ \leq W_{Dual} =& \min_{\boldsymbol{\mu} \in \mathcal{D}} \left[\max_{\boldsymbol{x}_t \in \mathcal{X}} \left[a(n;\boldsymbol{x}) - \boldsymbol{\mu}^\top \sum_{t=1}^{\boldsymbol{r}_n} \boldsymbol{A}^\top \boldsymbol{x}_t\right] + \max_{\boldsymbol{E}_n^\top \leq \boldsymbol{\gamma}} \left(p_{\boldsymbol{\lambda}}(\boldsymbol{E}_n^\top) + \boldsymbol{\mu}^\top \boldsymbol{E}_n^\top\right)\right] \\ =& \min_{\boldsymbol{\mu} \in \mathcal{D}} \left[h^*\left(\boldsymbol{A}\boldsymbol{\mu}\right)\right] + p_{\boldsymbol{\lambda}}^*(-\boldsymbol{\mu}), \end{aligned}$$

where $\mathcal{D} = \{\boldsymbol{\mu} | p_{\boldsymbol{\lambda}}^*(-\boldsymbol{\mu}) \leq \infty\}$ is the feasible region of dual variable $\boldsymbol{\mu}$ for which the conjugate of the regularized is bounded. $h^*(\cdot)$ and $p_{\boldsymbol{\lambda}}^*(\cdot)$ are the conjugate functions:

$$\begin{aligned} h^*(\boldsymbol{A}\boldsymbol{\mu}) &= \max_{\boldsymbol{x}_t \in \mathcal{X}} \left[a(n;\boldsymbol{x}) - \boldsymbol{\mu}^\top \sum_{t=1}^{\boldsymbol{r}_n} \boldsymbol{A}^\top \boldsymbol{x}_t\right] \\ &= \max_{\boldsymbol{x}_t \in \mathcal{X}} \sum_{t=1}^{\boldsymbol{r}_n} \left[\boldsymbol{s}_{u_t}^\top \boldsymbol{x}_t / \boldsymbol{r}_n - \boldsymbol{\mu}^\top \boldsymbol{A}^\top \boldsymbol{x}_t\right]; \\ p_{\boldsymbol{\lambda}}^*(-\boldsymbol{\mu}) &= \max_{\boldsymbol{E}_n^\top \leq \boldsymbol{\gamma}} \left(p_{\boldsymbol{\lambda}}(\boldsymbol{E}_n^\top) + \boldsymbol{\mu}^\top \boldsymbol{E}_n^\top\right). \end{aligned}$$

Next, we specifically give the feasible region $\mathcal{D}$ of the dual variable $\boldsymbol{\mu}$, which is:

$$\mathcal{D} = \left\{\boldsymbol{\mu} \in \mathbb{R}^{|\mathcal{P}|} \mid \boldsymbol{\mu} \geq -\boldsymbol{\lambda}\right\},$$

where $\boldsymbol{\lambda}$ is the penalty vector. The proof of the feasible region of $\mathcal{D}$ can be seen in Appendix A.2. □

A.2 Proof of Lemma 1

Proof. Notice that $p_{\boldsymbol{\lambda}}(\boldsymbol{E}_n^\top) = -\sum_{i=1}^{|\mathcal{P}|} \max\{\boldsymbol{M}_{n,i}^\top - \boldsymbol{E}_{n,i}^\top, 0\}$

We can write the conjugate as:

$$\begin{aligned} p_{\boldsymbol{\lambda}}^*(-\boldsymbol{\mu}) &= \sup_{\boldsymbol{E}_n^\top \leq \boldsymbol{\gamma}} \left(p_{\boldsymbol{\lambda}}(\boldsymbol{E}_n^\top) + \boldsymbol{\mu}^\top \boldsymbol{E}_n^\top\right) \\ &= \sup_{\boldsymbol{E}_n^\top \leq \boldsymbol{\gamma}} \{-\boldsymbol{\lambda}^\top \max(\boldsymbol{M}_n^\top - \boldsymbol{E}_n^\top, 0) + \boldsymbol{\mu}^\top \boldsymbol{E}_n^\top\} \\ &= \boldsymbol{\mu}^\top \boldsymbol{M}_n^\top + \sup_{\boldsymbol{z} \leq \boldsymbol{\gamma} - \boldsymbol{M}_n^\top} \{\boldsymbol{\lambda}^\top \min(\boldsymbol{z}, 0) + \boldsymbol{\mu}^\top \boldsymbol{z}\} \\ &= \boldsymbol{\mu}^\top \boldsymbol{M}_n^\top + s(\boldsymbol{\mu}), \end{aligned}$$

where the second equation follows by performing the change of variables $\boldsymbol{z} = \boldsymbol{E}_n^\top - \boldsymbol{M}_n^\top$ and the last from setting:

$$s(\boldsymbol{\mu}) := \sup_{\boldsymbol{z} \leq \boldsymbol{\gamma} - \boldsymbol{M}_n^\top} \{\boldsymbol{\lambda}^\top \min(\boldsymbol{z}, 0) + \boldsymbol{\mu}^\top \boldsymbol{z}\}.$$

First, we proof that if $\boldsymbol{\mu} \notin \mathcal{D}$, the conjugate function $p_{\boldsymbol{\lambda}}^*(-\boldsymbol{\mu}) = \infty$. Suppose that $\boldsymbol{\mu} \leq -\boldsymbol{\lambda}$. For any $\boldsymbol{z} \leq 0$, we have that $s(\boldsymbol{\mu}) \geq (\boldsymbol{\mu}^\top + \boldsymbol{\lambda}^\top)\boldsymbol{z}$. Letting $\boldsymbol{z} \to -\infty$ yields that $s(\boldsymbol{\mu}) = \infty$, which makes $p_{\boldsymbol{\lambda}}^*(-\boldsymbol{\mu}) = \infty$.

Second, consider $\boldsymbol{\mu} \in [-\boldsymbol{\lambda}, 0]$. Write $s(\boldsymbol{\mu}) := \sup_{\boldsymbol{z} \leq \boldsymbol{\gamma} - \boldsymbol{M}_n^\top} \min((\boldsymbol{\lambda}^\top + \boldsymbol{\mu}^\top)\boldsymbol{z}, \boldsymbol{\mu}^\top \boldsymbol{z})$. The objective is decreasing for $\boldsymbol{z} \geq 0$ and increasing for $\boldsymbol{z} \leq 0$. Therefore, the optimal solution is $\boldsymbol{z} = 0$, i.e., $\boldsymbol{E}_n^\top = \boldsymbol{M}_n^\top$, and $s(\boldsymbol{\mu}) = 0$.

Third, for $\boldsymbol{\mu} > 0$ a similar argument shows that the objective is increasing in $\boldsymbol{z}$. Therefore, it is optimal to set $\boldsymbol{z} = \boldsymbol{\gamma} - \boldsymbol{M}_n^\top$, i.e., $\boldsymbol{E}_n^\top = \gamma$, which yields $s(\boldsymbol{\mu}) = (\boldsymbol{\gamma} - \boldsymbol{M}_n^\top)\boldsymbol{\mu}$.

Thus, for $\boldsymbol{\mu} \in \mathcal{D} = \left\{\boldsymbol{\mu} \in \mathbb{R}^{|\mathcal{P}|} \mid \boldsymbol{\mu} \geq -\boldsymbol{\lambda}\right\}$, we have the optimal dual variable is: $\boldsymbol{E}_{p,n}^*(-\boldsymbol{\mu}) = \boldsymbol{M}_{p,n}$ if $\boldsymbol{\mu}_p \in \left[-\boldsymbol{M}_{p,n}, 0\right)$ and $\boldsymbol{E}_{p,n}^*(-\boldsymbol{\mu}) = \boldsymbol{\gamma}_p$ if $\boldsymbol{\mu}_p \geq 0$.

□