## A. Proof of Theorem 1

Proof. since $\mathrm{spt}\pi^* \subset \mathcal{N}$, its easy to verify that $\pi^* \in \Pi^{\mathcal{N}}(\mu,\nu)$, so that $\Pi^{\mathcal{N}}(\mu,\nu)$ is nonempty. and the problem $\min_{\pi\in\Pi^{\mathcal{N}}(\mu,\nu)} \hat{E}_\varepsilon(\pi)$ is feasible.

Now let $\hat{\pi}^*(\varepsilon_l)$ be the solution of (7) for $\varepsilon = \varepsilon_l$. Since $\Pi^{\mathcal{N}}(\mu,\nu)$ is bounded, we can extract a sequence (that we do not relabel for the sake of simplicity) such that $\hat{\pi}^*(\varepsilon_l) \to \hat{\pi}^*(l \to +\infty)$. Since $\Pi^{\mathcal{N}}(\mu,\nu)$ is closed, $\hat{\pi}^* \in \Pi^{\mathcal{N}}(\mu,\nu)$. Since $\pi^*$ is an optimal solution of unregularized problem, one has

$$\begin{aligned} 0 &\le \langle c, \hat{\pi}^*(\varepsilon_l)\rangle - \langle c, \pi^*\rangle \\ &= \sum_{(i,j)\in\mathcal{N}_0} c_{ij}\hat{\pi}^*_{ij}(\varepsilon_l) - \sum_{(i,j)\in\mathcal{N}_0} c_{ij}\pi^*_{ij} \\ &= \sum_{(i,j)\in\mathcal{N}} c_{ij}\hat{\pi}^*_{ij}(\varepsilon_l) - \sum_{(i,j)\in\mathcal{N}} c_{ij}\pi^*_{ij} \end{aligned} \tag{17}$$

Since $\hat{\pi}^*(\varepsilon_l)$ is the solution of (7),one has

$$\sum_{(i,j)\in\mathcal{N}} c_{ij}\hat{\pi}^*_{ij}(\varepsilon_l) - \varepsilon_l \sum_{(i,j)\in\mathcal{N}} \hat{\pi}^*_{ij}(\varepsilon_l)\log\frac{1}{\hat{\pi}^*_{ij}(\varepsilon_l)} \le \sum_{(i,j)\in\mathcal{N}} c_{ij}\pi^*_{ij} - \varepsilon_l \sum_{(i,j)\in\mathcal{N}} \pi^*_{ij}\log\frac{1}{\pi^*_{ij}} \tag{18}$$

By combining (17) and (18) together we have

$$\begin{aligned} 0 &\le \langle c, \hat{\pi}^*(\varepsilon_l)\rangle - \langle c, \pi^*\rangle \\ &\le \varepsilon_l\Big(\sum_{(i,j)\in\mathcal{N}} \hat{\pi}^*_{ij}(\varepsilon_l)\log\frac{1}{\hat{\pi}^*_{ij}(\varepsilon_l)} - \sum_{(i,j)\in\mathcal{N}} \pi^*_{ij}\log\frac{1}{\pi^*_{ij}}\Big) \\ &= \varepsilon_l\Big(\sum_{(i,j)\in\mathcal{N}_0} \hat{\pi}^*_{ij}(\varepsilon_l)\log\frac{1}{\hat{\pi}^*_{ij}(\varepsilon_l)} - \sum_{(i,j)\in\mathcal{N}_0} \pi^*_{ij}\log\frac{1}{\pi^*_{ij}}\Big) \\ &= H(\hat{\pi}^*(\varepsilon_l)) - H(\pi^*) \end{aligned} \tag{19}$$

Since the $H$ is continues, taking the limit $l \to +\infty$ in (19) shows that $\langle c, \hat{\pi}^*\rangle = \langle c, \pi^*\rangle$, so that $\hat{\pi}^*$ is a feasible candidate of the following optimization problem

$$\begin{aligned} \min_{\pi\in\Pi(\mu,\nu)} \quad & -H(\pi) \\ s.t. \quad & \langle c, \pi\rangle = L_c(\mu,\nu). \end{aligned} \tag{20}$$

Dividing by $\varepsilon_l$ in (19) we have

$$0 \le H(\hat{\pi}^*(\varepsilon_l)) - H(\pi^*) \tag{21}$$

Taking the limits $l \to +\infty$ shows that $H(\pi^*) \le H(\hat{\pi}^*)$, which shows that $\hat{\pi}^*$ is a solution of (20). Since the solution to the problem (20) is unique by strict convexity of $-H$, we have $\hat{\pi}^* = \pi^*$ and the whole sequence is converging. Since $x\log\frac{1}{x}$ is a continuous and bounded for $x \in [0,1]$, we have

$$\begin{aligned} \lim_{\varepsilon\to 0}\min_{\pi\in\Pi(\mu,\nu)} \hat{E}_\varepsilon(\pi) &= \lim_{\varepsilon\to 0}\sum_{(i,j)\in\mathcal{N}} \pi^*_{ij}(\varepsilon)c_{ij} - \\ &\quad \lim_{\varepsilon\to 0}\varepsilon\sum_{(i,j)\in\mathcal{N}} \pi^*_{ij}(\varepsilon)\log\frac{1}{\pi^*_{ij}(\varepsilon)} \\ &= \sum_{(i,j)\in\mathcal{N}} \pi^*_{ij}c_{ij} \\ &= \sum_{(i,j)\in\mathcal{N}_0} \pi^*_{ij}c_{ij} = \langle c, \pi^*\rangle = L_c(\mu,\nu), \end{aligned} \tag{22}$$

which finishes the proof.

### A.1. Proof of Theorem 2

We extend the restricted kernel $K$ to the whole space $\mathcal{N}_0$. Define

$$\tilde{K}_{ij} := \begin{cases} K_{ij} & (i,j)\in\mathcal{N}, \\ 0 & (i,j)\in\mathcal{N}_0\setminus\mathcal{N}. \end{cases} \tag{23}$$

Since $\mathrm{spt}\pi^* \subset \mathcal{N}$ and $\tilde{K}_{ij}$ is strictly positive for all $(i,j)\in\mathcal{N}$, it is easy to check that $r_i(\tilde{K}) > 0$ for all $i\in[n]$ and $c_j(\tilde{K}) > 0$ for all $j\in[m]$, so that the iteration in Algorithm 1 can be extended to the whole space $\mathcal{N}_0$ with

$$\begin{aligned} u_i^{(k+1)} &\leftarrow \mu_i/r_i(\tilde{K}v^{(k)}), \text{ for all } i\in[n], \\ v^{(k+1)} &\leftarrow v^{(k)}, k \text{ is even}, \\ v_j^{(k+1)} &\leftarrow \nu_j/c_j(\tilde{K}^T u^{(k+1)}), \text{ for all } j\in[m], \\ u^{(k+1)} &\leftarrow u^{(k)}, k \text{ is odd}. \end{aligned} \tag{24}$$

According to the theorem of the RAS algorithm (see [23] Theorem 4.1), the iteration (24) is convergent and outputs a coupling $\pi$ satisfying $||r(\pi)-\mu||_2 + ||c(\pi)-\nu||_2 < \delta^2$. It is easy to check that $\pi_{ij} = 0$ for all $(i,j)\in\mathcal{N}_0\setminus\mathcal{N}$, which shows that $r(\pi) = r^{\mathcal{N}}(\pi)$ and $c(\pi) = c^{\mathcal{N}}(\pi)$. The number of iterations $K_\delta$ needed to scale $\tilde{K}$ to accuracy $\delta$ satisfies $K_\delta = O(\rho h(\delta)^{-2}\log(h/s))$, where $h = \sum_{(i,j)\in\mathcal{N}_0}\pi_{ij} = \sum_{(i,j)\in\mathcal{N}}\pi_{ij}$, $s = \min_{\{(i,j)\in\mathcal{N}_0|\pi_{ij}>0\}}\pi = \min_{(i,j)\in\mathcal{N}}\pi$ and $\rho = \max\{||r(\pi)||_\infty, ||c(\pi)||_\infty\} = \max\{||r^{\mathcal{N}}(\pi)||_\infty, ||c^{\mathcal{N}}(\pi)||_\infty\}$

### A.2. Proof of Theorem 3

Let $\mathcal{N} \subset \mathcal{N}_0$ be a subspace such that $\mathrm{spt}\pi^* \subset \mathcal{N}$. Algorithm 1 with stopping criterion $||r^{\mathcal{N}}(\pi^{(k)}) - \mu||_1 + ||c^{\mathcal{N}}(\pi^{(k)}) - \nu||_1 < \delta$ outputs a matrix $\pi$ satisfying $||r^{\mathcal{N}}(\pi) - \mu||_1 + ||c^{\mathcal{N}}(\pi) - \nu||_1 < \delta$ after $O((\delta)^{-2}\log(h/s))$ iterations.

Proof. (Pinskers Inequality). For any probability measures $p$ and $q$, $||p-q||_1 \le \sqrt{2\mathrm{KL}(p|q)}$.

Let $k^*$ be the first iteration such that $||r^{\mathcal{N}}(\pi^{(k^*)}) - \mu||_1 + ||c^{\mathcal{N}}(\pi^{(k^*)}) - \nu||_1 < \delta$. Pinskers inequality implies that for any $k < k^*$, we have

$$\delta^2 < \left(||r^{\mathcal{N}}(\pi^{(k)}) - \mu||_1 + ||c^{\mathcal{N}}(\pi^{(k)}) - \nu||_1\right)^2 <$$
$$\left(\sqrt{2\mathrm{KL}(\mu|r^{\mathcal{N}}(\pi^{(k)}))} + \sqrt{2\mathrm{KL}(\nu|c^{\mathcal{N}}(\pi^{(k)}))}\right)^2$$

Assume without loss of generality that $k$ is even, so that $c^{\mathcal{N}}(\pi^{(k)}) = \nu$ and $r^{\mathcal{N}}(\pi^{(k-1)}) = \mu$, which shows that

$$\begin{aligned}
&\left(\sqrt{2\mathrm{KL}(\mu|r^{\mathcal{N}}(\pi^{(k)}))} + \sqrt{2\mathrm{KL}(\nu|c^{\mathcal{N}}(\pi^{(k)}))}\right)^2 \\
&= 2\mathrm{KL}(\mu|r^{\mathcal{N}}(\pi^{(k)})) \\
&= 2\sum_{i=1}^{n} \mu_i \log \frac{\mu_i}{r_i^{\mathcal{N}}(\pi^{(k)})} \qquad (25)
\end{aligned}$$

We have

$$\begin{aligned}
\sum_{i=1}^{n} \mu_i \log \frac{\mu_i}{r_i^{\mathcal{N}}(\pi^{(k)})} &= \sum_{i=1}^{n} \mu_i \log \frac{\mu_i}{\sum_{\{j|(i,j)\in\mathcal{N}\}} u_i^{(k)} K_{ij} v_j^{(k)}} \\
&= \sum_{i=1}^{n} \mu_i \log \frac{\mu_i}{u_i^{(k)} \sum_{\{j|(i,j)\in\mathcal{N}\}} K_{ij} v_j^{(k)}} \\
&= \sum_{i=1}^{n} \mu_i \log \frac{u_i^{(k+1)}}{u_i^{(k)}} \\
&= \sum_{i=1}^{n} \mu_i \log \frac{u_i^{(k+1)}}{u_i^{(k)}} + \sum_{j=1}^{m} \nu_j \log \frac{v_j^{(k+1)}}{v_j^{(k)}}.
\end{aligned}$$

So that

$$\frac{1}{2}\delta^2 < \sum_{i=1}^{n} \mu_i \log \frac{u_i^{(k+1)}}{u_i^{(k)}} + \sum_{j=1}^{m} \nu_j \log \frac{v_j^{(k+1)}}{v_j^{(k)}}. \qquad (26)$$

Equation (26) holds for $k$ is odd. By taking $k = 1, \ldots, k+1$ and adding all the equality together we have

$$\frac{1}{2}k\delta^2 < \sum_{i=1}^{n} \mu_i \log u_i^{(k+1)} + \sum_{j=1}^{m} \nu_j \log v_j^{(k+1)} - \sum_{i=1}^{n} \mu_i \log u_i^{(1)} - \sum_{j=1}^{m} \nu_j \log v_j^{(1)}. \qquad (27)$$

Define the Lyapunov functiion $\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$

$$f(x,y) = \sum_{(i,j)\in\mathcal{N}} K_{ij} \mathrm{e}^{x_i+y_j} - \sum_{i=1}^{n} \mu_i x_i - \sum_{j=1}^{m} \nu_j y_j \qquad (28)$$

Let $f(x^*, y^*)$ the global minimum of $f$, we have

$$\sum_{(i,j)\in\mathcal{N}} K_{ij} \mathrm{e}^{x_i^*+y_j^*} = 1, \qquad (29)$$

$$x_i^* + y_j^* \le \log(\frac{1}{s}). \forall i \in [n], \forall j \in [m]. \qquad (30)$$

So that

$$\begin{aligned}
\frac{1}{2}k\delta^2 &< \sum_{i=1}^{n} \mu_i \log u_i^{(k+1)} + \sum_{j=1}^{m} \nu_i \log v_i^{(k+1)} \qquad (31)\\
&\quad - \sum_{i=1}^{n} \mu_i \log u_i^{(1)} - \sum_{j=1}^{m} \nu_i \log v_i^{(1)} \\
&= f(u^{(1)}, v^{(1)}) - f(u^{(k+1)}, v^{(k+1)}) \\
&\le f(u^{(1)}, v^{(1)}) - f(x^*, y^*) \\
&= (1 - \sum_{i=1}^{n} \mu_i \log u^{(1)}) - (1 - \sum_{i=1}^{n} \mu_i x_i^* - \sum_{j=1}^{m} \nu_j y_i^*) \\
&\le -\sum_{i=1}^{n} \mu_i \log \frac{\mu_i}{\sum_{j=1}^{m} K_{ij}} + \log\frac{1}{s} \\
&= \sum_{i=1}^{n} \mu_i \log \frac{\sum_{j=1}^{m} K_{ij}}{\mu_i} + \log\frac{1}{s} \\
&\quad \left(\text{since } \sum_{i=1}^{n} \mu_i x_i^* + \sum_{j=1}^{m} \nu_j y_i^* \le \log\frac{1}{v}\right) \\
&\le \log \sum_{i=1}^{n} \left(\mu_i \frac{\sum_{j=1}^{m} K_{ij}}{\mu_i}\right) + \log(\frac{1}{s}) \\
&\quad \text{(Jensen inequality)} \\
&= \log\frac{h}{s}.
\end{aligned}$$

So that the inequality (32) implies that we terminate in $k^* \le 2\delta^{-2}\log(h/s)$ steps, which finishes the proof.

To prove Theorem (4), we state the following lemmas.

**Lemma 1** *For $\pi \in \Pi(\mu,\nu)$, we have $0 \le H(\pi) \le n\log n$.*

**Lemma 2** *If $\pi_1, \pi_2 \in \mathbb{R}_+^{n\times m}$ are nonnegative matrices and $\lambda \in [0,1]$, we have*

$$H(\lambda\pi_1 + (1-\lambda)\pi_2) \le \lambda H(\pi_1) + (1-\lambda)H(\pi_2) + \max\{\|\pi_1\|_1, \|\pi_2\|_1\} \cdot h(\lambda)$$

*where $h(\lambda) = \lambda\log\frac{1}{\lambda} + (1-\lambda)\log\frac{1}{1-\lambda} \quad \forall\lambda, \in [0,1]$.*

**Lemma 3** $h(\lambda) \le \lambda(\log\frac{1}{\lambda} + 1), \quad \lambda \in [0,1]$

**Lemma 4** *Let $\pi^*$ be the optimal solution of the unregularized problem, $\hat{\pi}^*(\hat{\varepsilon})$ be the optimal solution of the restricted optimal transport (7). The gap between the $\pi^*$ and $\hat{\pi}^*(\hat{\varepsilon})$ can be bounded by*

$$\|\hat{\pi}^*(\hat{\varepsilon}) - \pi^*\|_1 \le 2\exp(-\Delta/\hat{\varepsilon} + 4\frac{\log(n)}{n^2})$$

Proof. Let $V$ be the vertices of $\Pi(\mu,\nu)$ and $O : \{\pi \in V | \langle\pi, c\rangle = \langle\pi^*, c\rangle\}$ be the set of optimal vertex solution of (3). Let the suboptimal set be defined as $O_s := \{\pi \in$

$V|\langle\pi, c\rangle > \langle\pi, c\rangle\}$. Since $\hat{\pi}^*(\hat{\varepsilon}) \in \Pi(\mu,\nu)$, there exist $\lambda \in (0,1)$, $\pi^* \in V$ and $\tilde{\pi} \in \text{span}(O_s)$ such that

$$\hat{\pi}^*(\hat{\varepsilon}) = (1-\gamma)\pi^* + \gamma\tilde{\pi}.$$

We define the gap between $\hat{\pi}^*(\hat{\varepsilon})$ and $\pi^*$ by

$$g(\varepsilon) := \langle\hat{\pi}^*(\hat{\varepsilon}), c\rangle - \langle\pi^*, c\rangle$$

The lower bound of $g(\varepsilon)$ could be estimated by

$$\begin{aligned} g(\varepsilon) &= \langle\hat{\pi}^*(\hat{\varepsilon}), c\rangle - \langle\pi^*, c\rangle \\ &= \langle\hat{\pi}^*(\hat{\varepsilon}) - \pi^*, c\rangle \\ &= \gamma\langle\pi^* - \tilde{\pi}, c\rangle \geq \gamma\Delta \end{aligned}$$

The upper bound of $g(\varepsilon)$ could be estimated by

$$\begin{aligned} g(\varepsilon) &= \langle\hat{\pi}^*(\hat{\varepsilon}), c\rangle - \langle\pi^*, c\rangle \\ &\leq \varepsilon H(\hat{\pi}^*(\hat{\varepsilon})) - \varepsilon H(\pi^*). \end{aligned} \tag{32}$$

According to Lemma2, we have

$$\begin{aligned} H(\hat{\pi}^*(\hat{\varepsilon})) &= H((1-\gamma)\pi^* + \gamma\tilde{\pi}) \\ &\leq (1-\gamma)H(\pi^*) + \gamma H(\tilde{\pi}) \\ &+ \max\{\|\pi^*\|_1, \|\tilde{\pi}\|_1\}h(\lambda). \end{aligned}$$

One can check that $\|\pi^*\|_1 = 1$, so that we have

$$H(\hat{\pi}^*(\hat{\varepsilon})) \quad \leq (1-\gamma)H(\pi^*) + \gamma H(\tilde{\pi}) + h(\lambda). \tag{33}$$

Combining (32) and (33), the upper bound of $g(\varepsilon)/\varepsilon$ could be estimated by

$$\begin{aligned} \tfrac{g(\varepsilon)}{\varepsilon} &\leq H(\hat{\pi}^*(\hat{\varepsilon})) - H(\pi^*) \\ &\leq \gamma|H(\tilde{\pi}) - H(\pi^*)| + h(\gamma). \end{aligned}$$

According to lemma1 and $\gamma \leq \frac{g(\varepsilon)}{\Delta}$, we have

$$\tfrac{g(\varepsilon)}{\varepsilon} \leq \tfrac{g(\varepsilon)}{\Delta}|H(\tilde{\pi}) - H(\pi^*)| + h(\tfrac{g(\varepsilon)}{\Delta}).$$

According to lemma 1 and lemma 3, we have

$$\tfrac{g(\varepsilon)}{\varepsilon} \leq \tfrac{g(\varepsilon)}{\Delta} n\log n + \tfrac{g(\varepsilon)}{\Delta}\log\tfrac{\Delta}{g(\varepsilon)} + \tfrac{g(\varepsilon)}{\Delta},$$

which shows that $\Delta\exp(-\Delta/\varepsilon + n\log n) \geq g(\varepsilon) \geq \Delta\gamma$ and

$$\begin{aligned} \|\hat{\pi}^*(\hat{\varepsilon}) - \pi^*\|_1 &\leq \gamma\|\pi^* - \tilde{\pi}\|_1 \\ &\leq \gamma\|\pi^*\|_1 + \|\tilde{\pi}\|_1 \\ &\leq 2\Delta\exp(-\Delta/\varepsilon + n\log n) \end{aligned}$$

### A.3. Proof of Theorem 4

*Proof.* First, we prove that $\Delta > 0$, that is, we show $\inf\limits_{\{\pi\in V:\langle c,\pi\rangle > \langle c,\pi^*\rangle\}} \langle c,\pi\rangle - \langle c,\pi^*\rangle > 0$. Since the set $V$ is the vertices of the set $\Pi(\mu,\nu)$, so that the set $V$ contains finite elements, which shows that the set $\{\pi \in V : \langle c,\pi\rangle > \langle c,\pi^*\rangle\}$ has finite elements and we have $\Delta > 0$.

Second, let $s = \min_{\{(i,j)\in\mathcal{N},\pi_{ij}>0\}} \pi_{ij}$, we show that for any $\varepsilon \leq \varepsilon_0 < \frac{\Delta}{n\log n - \log(s/2)}$, the following inequality holds:

$$\|\hat{\pi}(\varepsilon)^{(l+1)} - \pi^*\|_1 < \tfrac{s}{2} + \delta, \tag{34}$$

where $\hat{\pi}(\varepsilon)^{(l+1)}$ is the coupling generated by the Restricted-Sinkhorn (Algorithm 2, step 5). If the claim (34) does not hold, then there exits some $\hat{\varepsilon} < \frac{\Delta}{n\log n - \log(s/2)}$ such that

$$\|\hat{\pi}(\hat{\varepsilon})^{(l+1)} - \pi^*\|_1 \geq \tfrac{s}{2} + \delta, \tag{35}$$

Let $\hat{\pi}^*(\hat{\varepsilon})$ be the optimal solution of the restricted optimal transport (7), we have

$$\begin{aligned} \|\hat{\pi}(\hat{\varepsilon})^{(l+1)} - \pi^*\|_1 &= \|\hat{\pi}(\hat{\varepsilon})^{(l+1)} - \hat{\pi}^*(\hat{\varepsilon}) + \hat{\pi}^*(\hat{\varepsilon}) - \pi^*\|_1 \\ &\leq \|\hat{\pi}(\hat{\varepsilon})^{(l+1)} - \hat{\pi}^*(\hat{\varepsilon})\|_1 \\ &+ \|\hat{\pi}^*(\hat{\varepsilon}) - \pi^*\|_1 \\ &\leq \delta + \|\hat{\pi}^*(\hat{\varepsilon}) - \pi^*\|_1 \end{aligned} \tag{36}$$

According to Lemma 4, we have

$$\|\hat{\pi}^*(\hat{\varepsilon}) - \pi^*\|_1 \leq 2\exp(-\Delta/\hat{\varepsilon} + n\log n)$$

So that we have

$$\begin{aligned} \|\hat{\pi}^*(\hat{\varepsilon}) - \pi^*\|_1 &\leq \exp(-\tfrac{\Delta}{\hat{\varepsilon}} + n\log n) \\ &< \exp(-\tfrac{\Delta}{\frac{\Delta}{n\log n - \log(s/2)}} + n\log n) \\ &= \exp(\log(s/2)) \\ &= s/2 \end{aligned} \tag{37}$$

By combing (36) and (37), we have

$$\|\hat{\pi}(\hat{\varepsilon})^{(l+1)} - \pi^*\|_1 \quad < \quad \delta + \tfrac{s}{2}, \tag{38}$$

which contradicts the result of (35). So the claim (35) is incorrect and claim (34) holds.

So that for any $(i,j) \in \text{spt}\pi$, we have

$$\begin{aligned} \hat{\pi}_{ij}(\varepsilon)^{(l+1)} &\geq \pi^*_{ij} - \tfrac{s}{2} - \delta \\ &\geq s - \tfrac{s}{2} - \delta \\ &\geq \tfrac{s}{2} - \delta. \end{aligned}$$

Finally, let $\theta < (\frac{s}{2} - \delta)^{(1+c)}$, we have

$$\begin{aligned} K_{ij}^{\mathcal{N}^{(l)}}(\varepsilon^{(l+1)}, \alpha^{(l+1)}, \beta^{(l+1)}) &= \exp(\log(\hat{\pi}_{ij}^{(l+1)})(1+c)) \\ &\geq \exp(\log(\tfrac{s}{2} - \delta)(1+c)) \\ &= (\tfrac{s}{2} - \delta)^{(1+c)} > \theta, \end{aligned}$$

which shows that $K_{ij}^{\mathcal{N}^{(l)}}(\varepsilon^{(l+1)}, \alpha^{(l+1)}, \beta^{(l+1)}) > \theta$ for any $(i,j) \in \text{spt}\pi$.