Supplement to "On Efficient Optimal Transport: An Analysis of Greedy and Accelerated Mirror Descent Algorithms"

In this Supplementary material, we first establish several key properties of APDAMD algorithm for the general setup in (12) in Section A. Then we provide the proofs for all the lemmas, theorems and propositions in the main context in Section B. Finally, in Section C, we provide some additional experimental results on synthetic and real MNIST images.

A. Properties of the APDAMD algorithm

In this section we present several important properties of the APDAMD algorithm that can be used later for regularized OT problems. First, we prove the following result regarding the number of line search iterations in the APDAMD algorithm:

Lemma A.1. *For the APDAMD algorithm, the number of line search iterations in the line search strategy is finite. Furthermore, the total number of gradient oracle calls after the k-th iteration is bounded as*

$$N_k \leq 4k + 4 + \frac{2\log\left(\frac{\|A\|_1^2}{2\eta}\right) - 2\log(L^0)}{\log 2}. \tag{17}$$

Proof. We follow Jiang et al. (2018) but we provide the proof details for the reader's convenience. First, we observe that multiplying M^k by two will not stop until the line search stopping criterion is satisfied. Therefore, we must have

$$M^k \geq \frac{\|A\|_1^2}{2\eta}.$$

By using Lemma 4.1, we obtain that the number of line search iterations in the line search strategy is finite. Letting i_j denote the total number of multiplication at the j-th iteration, we have

$$i_0 \ \leq \ 1 + \frac{\log\left(\frac{M^0}{L^0}\right)}{\log 2}, \qquad i_j \ \leq \ 2 + \frac{\log\left(\frac{M^j}{M^{j-1}}\right)}{\log 2}.$$

Furthermore, $M^j \leq \frac{\|A\|_1^2}{\eta}$ must hold. Otherwise, we have

$$\frac{M^j}{2} \ \geq \ \frac{\|A\|_1^2}{\eta},$$

which implies that the line search stopping criterion will be satisfied with $\frac{M^j}{2}$ and proceed to the line search in the next iteration. Therefore, the total number of line search can be bounded by

$$\begin{aligned}
\sum_{j=0}^{k} i_j \ &\leq \ 1 + \frac{\log\left(\frac{M^0}{L^0}\right)}{\log 2} + \sum_{j=1}^{k}\left(2 + \frac{\log\left(\frac{M^j}{M^{j-1}}\right)}{\log 2}\right) \\
&\leq \ 2k + 1 + \frac{\log\left(M^k\right) - \log(L^0)}{\log 2} \\
&\leq \ 2k + 1 + \frac{\log\left(\frac{\|A\|_1^2}{2\eta}\right) - \log(L^0)}{\log 2}.
\end{aligned}$$

Since each line search contains two gradient oracle calls, we conclude (17). $\square$

The next lemma presents a property of the dual objective function at the iterates of the APDAMD algorithm.

Lemma A.2. *For each iteration k of the APDAMD algorithm and any $z \in \mathbb{R}^n$, we have*

$$\begin{aligned}
&\bar{\alpha}^k \varphi(\lambda^k) \\
&\leq \sum_{j=0}^{k}\left[\alpha^j\left(\varphi(\mu^j) + \left\langle \nabla\varphi(\mu^j), z - \mu^j\right\rangle\right)\right] + \|z\|_\infty^2 .
\end{aligned} \tag{18}$$

Proof. We follow the proof path in Dvurechensky et al. (2018) with ℓ_∞-norm instead of ℓ_2-norm. First, we claim that

$$\begin{aligned}
\alpha^{k+1}\left\langle \nabla\varphi(\mu^{k+1}), z^k - z\right\rangle \ &\leq \ \bar{\alpha}^{k+1}\left(\varphi(\mu^{k+1}) - \varphi(\lambda^{k+1})\right) \\
&\quad + B_\phi(z, z^k) - B_\phi(z, z^{k+1}),
\end{aligned} \tag{19}$$

for any $z \in \mathbb{R}^n$. Indeed, it follows from the optimality condition in the mirror descent step that, for any $z \in \mathbb{R}^n$, we have

$$\left\langle \nabla\varphi(\mu^{k+1}) + \frac{\nabla\phi(z^{k+1}) - \nabla\phi(z^k)}{\alpha^{k+1}}, z - z^{k+1}\right\rangle \ \geq \ 0. \tag{20}$$

Recall the celebrated generalized triangle inequality for the Bregman divergence:

$$\begin{aligned}
&B_\phi(z, z^k) - B_\phi(z, z^{k+1}) - B_\phi(z^{k+1}, z^k) \\
&\quad = \left\langle \nabla\phi(z^{k+1}) - \nabla\phi(z^k), z - z^{k+1}\right\rangle .
\end{aligned} \tag{21}$$

Therefore, we have

$$\begin{aligned}
&\alpha^{k+1}\left\langle \nabla\varphi(\mu^{k+1}), z^k - z\right\rangle \qquad (22)\\
&\quad = \alpha^{k+1}\left\langle \nabla\varphi(\mu^{k+1}), z^k - z^{k+1}\right\rangle \\
&\qquad +\alpha^{k+1}\left\langle \nabla\varphi(\mu^{k+1}), z^{k+1} - z\right\rangle \\
&\quad \overset{(20)}{\leq} \alpha^{k+1}\left\langle \nabla\varphi(\mu^{k+1}), z^k - z^{k+1}\right\rangle \\
&\qquad + \left\langle \nabla\phi(z^{k+1}) - \nabla\phi(z^k), z - z^{k+1}\right\rangle \\
&\quad \overset{(21)}{=} \alpha^{k+1}\left\langle \nabla\varphi(\mu^{k+1}), z^k - z^{k+1}\right\rangle \\
&\qquad +B_\phi(z, z^k) - B_\phi(z, z^{k+1}) - B_\phi(z^{k+1}, z^k) \\
&\quad \leq \alpha^{k+1}\left\langle \nabla\varphi(\mu^{k+1}), z^k - z^{k+1}\right\rangle \\
&\qquad +B_\phi(z, z^k) - B_\phi(z, z^{k+1}) - \frac{1}{2\gamma}\left\|z^{k+1} - z^k\right\|_\infty^2,
\end{aligned}$$

where the last inequality comes from the fact that ϕ is $\frac{1}{\gamma}$-strongly convex with respect to ℓ_∞-norm. Furthermore, we observe from the update formula of μ^{k+1} and λ^{k+1} that

$$\lambda^{k+1} - \mu^{k+1} = \frac{\alpha^{k+1}}{\bar{\alpha}^{k+1}}(z^{k+1} - z^k), \qquad (23)$$

and the update formula of α^{k+1} and $\bar{\alpha}^{k+1}$ yields

$$\gamma M^k (\alpha^{k+1})^2 = \bar{\alpha}^k + \alpha^{k+1} = \bar{\alpha}^{k+1}. \qquad (24)$$

Therefore, we have

$$\begin{aligned}
&\alpha^{k+1}\langle \nabla\varphi(\mu^{k+1}), z^k - z^{k+1}\rangle \\
&\qquad \overset{(23)}{=} \bar{\alpha}^{k+1}\langle \nabla\varphi(\mu^{k+1}), \mu^{k+1} - \lambda^{k+1}\rangle.
\end{aligned}$$

In addition, the following equality holds:

$$\begin{aligned}
\left\|z^{k+1} - z^k\right\|_\infty^2 &\overset{(23)}{=} \left(\frac{\bar{\alpha}^{k+1}}{\alpha^{k+1}}\right)^2 \left\|\mu^{k+1} - \lambda^{k+1}\right\|_\infty^2 \\
&\overset{(24)}{=} \gamma M^k \bar{\alpha}^{k+1}\left\|\mu^{k+1} - \lambda^{k+1}\right\|_\infty^2.
\end{aligned}$$

Plugging all the above equations into (22) yields that

$$\begin{aligned}
&\alpha^{k+1}\left\langle \nabla\varphi(\mu^{k+1}), z^k - z\right\rangle \\
&\quad \leq \bar{\alpha}^{k+1}\left\langle \nabla\varphi(\mu^{k+1}), \mu^{k+1} - \lambda^{k+1}\right\rangle + B_\phi(z, z^k) \\
&\quad - B_\phi(z, z^{k+1}) - \frac{\bar{\alpha}^{k+1}M^k}{2}\left\|\mu^{k+1} - \lambda^{k+1}\right\|_\infty^2 \\
&\quad = \bar{\alpha}^{k+1}\Bigg(\left\langle \nabla\varphi(\mu^{k+1}), \mu^{k+1} - \lambda^{k+1}\right\rangle \\
&\quad - \frac{M^k}{2}\left\|\mu^{k+1} - \lambda^{k+1}\right\|_\infty^2\Bigg) + B_\phi(z, z^k) - B_\phi(z, z^{k+1}) \\
&\quad \leq \bar{\alpha}^{k+1}\left(\varphi(\mu^{k+1}) - \varphi(\lambda^{k+1})\right) \\
&\quad + B_\phi(z, z^k) - B_\phi(z, z^{k+1}),
\end{aligned}$$

where the last inequality comes from the stopping criterion in the line search strategy. Therefore, we conclude the desired inequality (19).

The next step is to bound the iterative objective gap, i.e., for $z \in \mathbb{R}^n$,

$$\begin{aligned}
&\bar{\alpha}^{k+1}\varphi(\lambda^{k+1}) - \bar{\alpha}^k\varphi(\lambda^k) \qquad (25)\\
&\quad \leq \alpha^{k+1}\left(\varphi(\mu^{k+1}) + \left\langle \nabla\varphi(\mu^{k+1}), z - \mu^{k+1}\right\rangle\right) \\
&\qquad +B_\phi(z, z^k) - B_\phi(z, z^{k+1}),
\end{aligned}$$

Indeed, we observe from the update formula of μ^{k+1} that

$$\begin{aligned}
&\alpha^{k+1}\left(\mu^{k+1} - z^k\right) \qquad (26)\\
&\quad \overset{(24)}{=} \left(\bar{\alpha}^{k+1} - \bar{\alpha}^k\right)\mu^{k+1} - \alpha^{k+1}z^k \\
&\quad = \alpha^{k+1}z^k + \bar{\alpha}^k\lambda^k - \bar{\alpha}^k\mu^{k+1} - \alpha^{k+1}z^k \\
&\quad = \bar{\alpha}^k\left(\lambda^k - \mu^{k+1}\right).
\end{aligned}$$

Thus, we have

$$\begin{aligned}
&\alpha^{k+1}\left\langle \nabla\varphi(\mu^{k+1}), \mu^{k+1} - z\right\rangle \\
= \; &\alpha^{k+1}\left\langle \nabla\varphi(\mu^{k+1}), \mu^{k+1} - z^k\right\rangle \\
&\qquad +\alpha^{k+1}\left\langle \nabla\varphi(\mu^{k+1}), z^k - z\right\rangle \\
\overset{(26)}{=} \; &\bar{\alpha}^k\left\langle \nabla\varphi(\mu^{k+1}), \lambda^k - \mu^{k+1}\right\rangle \\
&\qquad +\alpha^{k+1}\left\langle \nabla\varphi(\mu^{k+1}), z^k - z\right\rangle \\
= \; &D.
\end{aligned}$$

Furthermore, given the results of (19) and (24), the following results hold:

$$\begin{aligned}
D \;&\leq\; \bar{\alpha}^k\left(\varphi(\lambda^k) - \varphi(\mu^{k+1})\right) \\
&+\; \alpha^{k+1}\left\langle \nabla\varphi(\mu^{k+1}), z^k - z\right\rangle \\
&\overset{(19)}{\leq}\; \bar{\alpha}^k\left(\varphi(\lambda^k) - \varphi(\mu^{k+1})\right) \\
&+\; \bar{\alpha}^{k+1}\left(\varphi(\mu^{k+1}) - \varphi(\lambda^{k+1})\right) + B_\phi(z, z^k) \\
&-\; B_\phi(z, z^{k+1}) \\
&\overset{(24)}{=}\; \bar{\alpha}^k\varphi(\lambda^k) - \bar{\alpha}^{k+1}\varphi(\lambda^{k+1}) \\
&+\; \alpha^{k+1}\varphi(\mu^{k+1}) + B_\phi(z, z^k) - B_\phi(z, z^{k+1}).
\end{aligned}$$

Summing up (25) over $k = 0, 1, \ldots, N-1$ yields that

$$\begin{aligned}
&\bar{\alpha}^N\varphi(\lambda^N) - \bar{\alpha}^0\varphi(\lambda^0) \\
&\quad \leq \sum_{k=0}^{N-1}\left[\alpha^{k+1}\left(\varphi(\mu^{k+1}) + \left\langle \nabla\varphi(\mu^{k+1}), z - \mu^{k+1}\right\rangle\right)\right] \\
&\quad + B_\phi(z, z^0) - B_\phi(z, z^N).
\end{aligned}$$

Finally, we observe that $\alpha^0 = \bar{\alpha}^0 = 0$, $B_\phi(z, z^N) \geq 0$ and ϕ is 1-smooth with respect to ℓ_∞-norm, and conclude that, for any $z \in \mathbb{R}^n$

$$
\begin{aligned}
\bar{\alpha}^N \varphi(\lambda^N) \quad & \leq \quad \sum_{k=0}^{N} \left[\alpha^k \left(\varphi(\mu^k) + \left\langle \nabla\varphi(\mu^k), z - \mu^k \right\rangle\right)\right] \\
& \quad\quad + B_\phi(z, z^0) \\
& \leq \quad \sum_{k=0}^{N} \left[\alpha^k \left(\varphi(\mu^k) + \left\langle \nabla\varphi(\mu^k), z - \mu^k \right\rangle\right)\right] \\
& \quad\quad + \left\|z - z^0\right\|_\infty^2 \\
& \overset{z^0=0}{=} \quad \sum_{k=0}^{N} \left[\alpha^k \left(\varphi(\mu^k) + \left\langle \nabla\varphi(\mu^k), z - \mu^k \right\rangle\right)\right] \\
& \quad\quad + \|z\|_\infty^2 .
\end{aligned}
$$

The desired inequality (18) directly follows by changing the counter from k to i and the iteration count N to k. □

The final lemma provides us with a key lower bound for the accumulating parameter.

Lemma A.3. *For each iteration k of the APDAMD algorithm, we have*

$$
\bar{\alpha}^k \ \geq \ \frac{\eta(k+1)^2}{8\gamma \|A\|_1^2}. \tag{27}
$$

Proof. For $k = 1$, we have

$$
\bar{\alpha}^1 \ \overset{(24)}{=} \ \alpha^1 \ \overset{(24)}{=} \ \frac{1}{\gamma M^1} \ \geq \ \frac{\eta}{2\gamma \|A\|_1^2},
$$

where $M^1 \leq \frac{2\|A\|_1^2}{\eta}$ has been proven in Lemma A.1. So (27) holds true for $k = 1$. Then we proceed to prove (27) for $k \geq 1$ by using the mathematical induction. Indeed, we have

$$
\begin{aligned}
\bar{\alpha}^{k+1} \quad & = \quad \bar{\alpha}^k + \alpha^{k+1} \\
& = \quad \bar{\alpha}^k + \frac{1 + \sqrt{1 + 4\gamma M^k \bar{\alpha}^k}}{2\gamma M^k} \\
& = \quad \bar{\alpha}^k + \frac{1}{2\gamma M^k} + \sqrt{\frac{1}{4\left(\gamma M^k\right)^2} + \frac{\bar{\alpha}^k}{\gamma M^k}} \\
& \geq \quad \bar{\alpha}^k + \frac{1}{2\gamma M^k} + \sqrt{\frac{\bar{\alpha}^k}{\gamma M^k}} \\
& \geq \quad \bar{\alpha}^k + \frac{\eta}{4\gamma \|A\|_1^2} + \sqrt{\frac{\eta \bar{\alpha}^k}{2\gamma \|A\|_1^2}},
\end{aligned}
$$

where $M^k \leq \frac{2\|A\|_1^2}{\eta}$ was also proven in Lemma A.1. Now, we assume that (27) hold true for k. Then, we find that

$$
\begin{aligned}
\bar{\alpha}^{k+1} \quad & \geq \quad \frac{\eta(k+1)^2}{8\gamma \|A\|_1^2} + \frac{\eta}{4\gamma \|A\|_1^2} + \sqrt{\frac{\eta^2(k+1)^2}{16\gamma^2 \|A\|_1^4}} \\
& = \quad \frac{\eta}{8\gamma \|A\|_1^2} \left[(k+1)^2 + 2 + 2(k+1)\right] \\
& \geq \quad \frac{\eta(k+2)^2}{8\gamma \|A\|_1^2},
\end{aligned}
$$

which implies that (27) holds true for $k + 1$. □

B. Technical Proofs

In this section, we provide the proofs for the remaining results in the paper.

B.1. Proof of Lemma 3.1

By the definition, we have

$$
\begin{aligned}
f(u, v) &= \mathbf{1}^\top B(u, v)\mathbf{1} - \langle u, r\rangle - \langle v, l\rangle \\
&= \sum_{i,j=1}^{n} e^{u_i + v_j - \frac{C_{ij}}{\eta}} - \sum_{i=1}^{n} r_i u_i - \sum_{j=1}^{n} l_j v_j
\end{aligned}
$$

The gradients of f at (u^k, v^k) are

$$
\begin{aligned}
\nabla_u f(u^k, v^k) \quad & = \quad B(u^k, v^k)\mathbf{1} - r, \\
\nabla_v f(u^k, v^k) \quad & = \quad B(u^k, v^k)^\top \mathbf{1} - l.
\end{aligned}
$$

Therefore, the quantity E^k can be rewritten as

$$
E^k \ = \ \|\nabla_u f(u^k, v^k)\|_1 + \|\nabla_v f(u^k, v^k)\|_1.
$$

By using the fact that f is convex and globally minimized at (u^*, v^*), we have

$$
\begin{aligned}
f(u^k, v^k) - f(u^*, v^*) \leq & \ (u^k - u^*)^\top \nabla_u f(u^k, v^k) \\
& + (v^k - v^*)^\top \nabla_v f(u^k, v^k).
\end{aligned}
$$

Applying Hölder's inequality yields

$$
\begin{aligned}
& f(u^k, v^k) - f(u^*, v^*) \\
& \quad \leq \quad \|u^k - u^*\|_\infty \|\nabla_u f(u^k, v^k)\|_1 \\
& \quad + \quad \|v^k - v^*\|_\infty \|\nabla_v f(u^k, v^k)\|_1 \\
& \quad = \quad \left(\|u^k - u^*\|_\infty + \|v^k - v^*\|_\infty\right) E^k.
\end{aligned} \tag{28}
$$

Thus it suffices to show that

$$
\|u^k - u^*\|_\infty + \|v^k - v^*\|_\infty \ \leq \ 2\|u^*\|_\infty + 2\|v^*\|_\infty .
$$

The next result is the key observation that makes our analysis work for the Greenkhorn algorithm. We use an induction argument to establish the following bound:

$$
\begin{aligned}
& \max\{\|u^k - u^*\|_\infty, \ \|v^k - v^*\|_\infty\} \\
& \qquad\qquad \leq \max\{\|u^0 - u^*\|_\infty, \ \|v^0 - v^*\|_\infty\}.
\end{aligned} \tag{29}
$$

It is easy to verify (29) for $k = 0$. Assuming that it holds true for $k = k_0 \geq 0$, we show that it also holds true for $k = k_0 + 1$. Without loss of generality, let I be the index chosen at the $k_0 + 1$-th iteration. Then we have

$$\|u^{k_0+1} - u^*\|_\infty \leq \max\{\|u^{k_0} - u^*\|_\infty, |u_I^{k_0+1} - u_I^*|\}, \tag{30}$$

$$\|v^{k_0+1} - v^*\|_\infty = \|v^{k_0} - v^*\|_\infty. \tag{31}$$

By the updating formula for $u_I^{k_0+1}$ and the optimality condition for u_I^*, we have

$$e^{u_I^{k_0+1}} = \frac{r_I}{\sum_{j=1}^n e^{-\frac{C_{Ij}}{\eta} + v_j^{k_0}}}, \qquad e^{u_I^*} = \frac{r_I}{\sum_{j=1}^n e^{-\frac{C_{Ij}}{\eta} + v_j^*}}.$$

This implies that

$$\begin{aligned} |u_I^{k_0+1} - u_I^*| &= \left|\log\left(\frac{\sum_{j=1}^n e^{-C_{Ij}/\eta + v_j^{k_0}}}{\sum_{j=1}^n e^{-C_{Ij}/\eta + v_j^*}}\right)\right| \qquad (32) \\ &\leq \|v^{k_0} - v^*\|_\infty, \end{aligned}$$

where the inequality comes from the following inequality:

$$\frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i} \leq \max_{1\leq j\leq n} \frac{a_i}{b_i}, \quad \forall a_i, b_i > 0.$$

Combining (30) and (32) yields

$$\|u^{k_0+1} - u^*\|_\infty \leq \max\{\|u^{k_0} - u^*\|_\infty, \|v^{k_0} - v^*\|_\infty\}. \tag{33}$$

Therefore, we conclude that (29) holds true for $k = k_0 + 1$ by combining (31) and (33). Since $u^0 = v^0 = 0$, (29) implies that

$$\begin{aligned} & \|u^k - u^*\|_\infty + \|v^k - v^*\|_\infty \qquad (34) \\ \leq\ & 2\left(\|u^0 - u^*\|_\infty + \|v^0 - v^*\|_\infty\right) \\ =\ & 2\|u^*\|_\infty + 2\|v^*\|_\infty. \end{aligned}$$

Finally, we obtain the result (7) by combining (28) and (34).

B.2. Proof of Lemma 3.2

First, we claim that there exists an optimal solution pair (u^*, v^*) such that

$$\max_{1\leq i\leq n} u_i^* \geq 0 \geq \min_{1\leq i\leq n} u_i^*. \tag{35}$$

Indeed, since the function f is convex with respect to (u, v), the set of optima of problem (5) is not empty. Thus, we can choose an optimal solution $(\tilde{u}^*, \tilde{v}^*)$ where

$$\begin{aligned} +\infty &> \max_{1\leq i\leq n} \tilde{u}_i^* \geq \min_{1\leq i\leq n} \tilde{u}_i^* > -\infty, \\ +\infty &> \max_{1\leq i\leq n} \tilde{v}_i^* \geq \min_{1\leq i\leq n} \tilde{v}_i^* > -\infty. \end{aligned}$$

Given the optimal solution $(\tilde{u}^*, \tilde{v}^*)$, we let (u^*, v^*) be

$$\begin{aligned} u^* &= \tilde{u}^* - \frac{\max_{1\leq i\leq n} u_i^* + \min_{1\leq i\leq n} u_i^*}{2}\mathbf{1}, \\ v^* &= \tilde{v}^* + \frac{\max_{1\leq i\leq n} u_i^* + \min_{1\leq i\leq n} u_i^*}{2}\mathbf{1}. \end{aligned}$$

and observe that (u^*, v^*) satisfies (35). It now suffices to show that (u^*, v^*) is optimal; i.e., $f(u^*, v^*) = f(\tilde{u}^*, \tilde{v}^*)$. Since $\mathbf{1}^\top r = \mathbf{1}^\top l = 1$, we have

$$\langle u^*, r\rangle = \langle \tilde{u}^*, r\rangle, \quad \langle v^*, l\rangle = \langle \tilde{v}^*, l\rangle.$$

Therefore, we conclude that

$$\begin{aligned} f(u^*, v^*) &= \sum_{i,j=1}^n e^{-C_{ij}/\eta + u_i^* + v_j^*} - \langle u^*, r\rangle - \langle v^*, l\rangle \\ &= \sum_{i,j=1}^n e^{-C_{ij}/\eta + \tilde{u}_i^* + \tilde{v}_j^*} - \langle \tilde{u}^*, r\rangle - \langle \tilde{v}^*, l\rangle \\ &= f(\tilde{u}^*, \tilde{v}^*). \end{aligned}$$

The next step is to establish the following bounds:

$$\max_{1\leq i\leq n} u_i^* - \min_{1\leq i\leq n} u_i^* \leq \frac{\|C\|_\infty}{\eta} - \log\left(\min_{1\leq i,j\leq n}\{r_i, l_j\}\right), \tag{36}$$

$$\max_{1\leq i\leq n} v_i^* - \min_{1\leq i\leq n} v_i^* \leq \frac{\|C\|_\infty}{\eta} - \log\left(\min_{1\leq i,j\leq n}\{r_i, l_j\}\right). \tag{37}$$

Indeed, for each $1 \leq i \leq n$, we have

$$\begin{aligned} e^{-\|C\|_\infty/\eta + u_i^*}\left(\sum_{j=1}^n e^{v_j^*}\right) &\leq \sum_{j=1}^n e^{-C_{ij}/\eta + u_i^* + v_j^*} \\ &= [B(u^*, v^*)\mathbf{1}]_i = r_i \leq 1, \end{aligned}$$

implying that

$$u_i^* \leq \frac{\|C\|_\infty}{\eta} - \log\left(\sum_{j=1}^n e^{v_j^*}\right). \tag{38}$$

On the other hand, we have

$$\begin{aligned} e^{u_i^*}\left(\sum_{j=1}^n e^{v_j^*}\right) &\geq \sum_{j=1}^n e^{-C_{ij}/\eta + u_i^* + v_j^*} \\ &= [B(u^*, v^*)\mathbf{1}]_i = r_i \geq \min_{1\leq i,j\leq n}\{r_i, l_j\}, \end{aligned}$$

implying that

$$u_i^* \geq \log\left(\min_{1\leq i,j\leq n}\{r_i, l_j\}\right) - \log\left(\sum_{j=1}^n e^{v_j^*}\right). \tag{39}$$

Combining (38) and (39) yields (36). In addition, (37) can be proved by a similar argument.

Finally, we proceed to prove that (8) holds true. We first assume that

$$\max_{1\leq i\leq n} v_i^* \ \geq \ 0, \quad \max_{1\leq i\leq n} u_i^* \ \geq \ 0 \ \geq \ \min_{1\leq i\leq n} u_i^*.$$

The optimality condition implies that

$$\sum_{i,j=1}^{n} e^{-\frac{C_{ij}}{\eta}+u_i^*+v_j^*} = 1,$$

and

$$\max_{1\leq i\leq n} u_i^* + \max_{1\leq i\leq n} v_i^* \ \leq \ \log\left(\max_{1\leq i,j\leq n} e^{C_{ij}/\eta}\right) \ = \ \frac{\|C\|_\infty}{\eta}.$$

Equipped with the assumptions $\max_{1\leq i\leq n} u_i^* \geq 0$ and $\max_{1\leq i\leq n} v_i^* \geq 0$, we have

$$0 \ \leq \ \max_{1\leq i\leq n} u_i^* \ \leq \ \frac{\|C\|_\infty}{\eta}, \qquad 0 \ \leq \ \max_{1\leq i\leq n} v_i^* \ \leq \ \frac{\|C\|_\infty}{\eta}. \tag{40}$$

Combining (40) with (36) and (37) yields

$$\begin{aligned}
\min_{1\leq i\leq n} u_i^* \ &\geq \ -\frac{\|C\|_\infty}{\eta} + \log\left(\min_{1\leq i,j\leq n}\{r_i, l_j\}\right), \\
\min_{1\leq i\leq n} v_i^* \ &\geq \ -\frac{\|C\|_\infty}{\eta} + \log\left(\min_{1\leq i,j\leq n}\{r_i, l_j\}\right).
\end{aligned}$$

We conclude (8) by putting together the above inequalities.

We proceed to the alternative scenario, where

$$\max_{1\leq i\leq n} v_i^* \ \leq \ 0, \quad \max_{1\leq i\leq n} u_i^* \ \geq \ 0 \ \geq \ \min_{1\leq i\leq n} u_i^*.$$

Combining with (36) yields

$$\begin{aligned}
\max_{1\leq i\leq n} u_i^* \ &\leq \ \frac{\|C\|_\infty}{\eta} - \log\left(\min_{1\leq i,j\leq n}\{r_i, l_j\}\right) \\
\min_{1\leq i\leq n} u_i^* \ &\geq \ -\frac{\|C\|_\infty}{\eta} + \log\left(\min_{1\leq i,j\leq n}\{r_i, l_j\}\right).
\end{aligned}$$

Similar to (39), we have

$$\begin{aligned}
\min_{1\leq i\leq n} v_i^* \ &\geq \ \log\left(\min_{1\leq i,j\leq n}\{r_i, l_j\}\right) - \log\left(\sum_{i=1}^{n} e^{u_i^*}\right) \\
&\geq \ 2\log\left(\min_{1\leq i,j\leq n}\{r_i, l_j\}\right) - \log(n) - \frac{\|C\|_\infty}{\eta},
\end{aligned}$$

and again we conclude that (8) holds.

B.3. Proof of Lemma 3.6

We observe that

$$\begin{aligned}
& f(u^k, v^k) - f(u^{k+1}, v^{k+1}) \\
\geq \ & \frac{1}{2n}\left(\rho\left(r, B(u^k, v^k)\mathbf{1}\right) + \rho\left(c, B(u^k, v^k)^\top\mathbf{1}\right)\right) \\
\geq \ & \frac{1}{14n}\left(\|r - B(u^k, v^k)\mathbf{1}\|_1^2 + \|c - B(u^k, v^k)^\top\mathbf{1}\|_1^2\right),
\end{aligned}$$

where the first inequality comes from Lemma 5 in Altschuler et al. (2017) and the fact that the row or column update is chosen in a greedy manner, and the second inequality comes from Lemma 6 in Altschuler et al. (2017). Therefore, by the definition of E^k, we conclude (10).

B.4. Proof of Theorem 3.7

Denote $\delta_k = f(u^k, v^k) - f(u^*, v^*)$. Based on the results of Corollary 3.3 and Lemma 3.6, we have

$$\delta_k - \delta_{k+1} \ \geq \ \max\left\{\frac{\delta_k^2}{448nR^2}, \frac{(\varepsilon')^2}{28n}\right\},$$

where $E_k \geq \varepsilon'$ as soon as the stopping criterion is not fulfilled. In the following step we apply a switching strategy introduced by Dvurechensky et al. (2018). More specifically, given any $k \geq 1$, we have two estimates:

(i) Considering the process from the first iteration and the k-th iteration, we have

$$\frac{\delta_{k+1}}{448nR^2} \ \leq \ \frac{1}{k + \frac{448nR^2}{\delta_1^2}}.$$

The above inequality directly leads to

$$k \ \leq \ 1 + \frac{448nR^2}{\delta_k} - \frac{448nR^2}{\delta_1}.$$

(ii) Considering the process from the $(k+1)$-th iteration to the $(k+m)$-th iteration for $\forall m \geq 1$, we have

$$\delta_{k+m} \ \leq \ \delta_k - \frac{(\varepsilon')^2 m}{28n}.$$

The above result demonstrates that

$$m \ \leq \ \frac{28n}{(\varepsilon')^2}\left(\delta_k - \delta_{k+m}\right).$$

We then minimize the sum of these two estimates by an optimal choice of a tradeoff parameter $s \in (0, \delta_1]$:

$$\begin{aligned}
k \ &\leq \ \min_{0<s\leq\delta_1}\left(2 + \frac{448nR^2}{s} - \frac{448nR^2}{\delta_1} + \frac{28ns}{(\varepsilon')^2}\right) \\
&= \ \begin{cases} 2 + \dfrac{224nR}{\varepsilon'} - \dfrac{448nR^2}{\delta_1}, & \delta_1 \geq 4R\varepsilon', \\ 2 + \dfrac{28n\delta_1}{(\varepsilon')^2}, & \delta_1 \leq 4R\varepsilon'. \end{cases}
\end{aligned}$$

This implies that $k \leq 2 + \frac{112nR}{\varepsilon'}$ in both cases. Therefore, we conclude that the number of iterations k satisfies (11).

B.5. Proof of Theorem 3.8

We follow the same steps as in the proof of Theorem 1 in Altschuler et al. (2017) and obtain

$$\begin{aligned} & \left\langle C, \hat{X} \right\rangle - \left\langle C, X^* \right\rangle \\ \leq \ & 2\eta \log(n) + 4\left(\left\| \tilde{X}\mathbf{1} - r \right\|_1 + \left\| \tilde{X}^\top \mathbf{1} - l \right\|_1 \right) \|C\|_\infty \\ \leq \ & \frac{\varepsilon}{2} + 4\left(\left\| \tilde{X}\mathbf{1} - r \right\|_1 + \left\| \tilde{X}^\top \mathbf{1} - l \right\|_1 \right) \|C\|_\infty , \end{aligned}$$

where $\hat{X}$ is the output of Algorithm 2, X^* is a solution to the optimal transport problem and $\tilde{X}$ is the matrix returned by the Greenkhorn algorithm (Algorithm 1) with $\tilde{r}$, $\tilde{l}$ and $\varepsilon'/2$ in Step 3 of Algorithm 2. The last inequality in the above display holds since $\eta = \frac{\varepsilon}{4\log(n)}$. Furthermore, we have

$$\begin{aligned} & \left\| \tilde{X}\mathbf{1} - r \right\|_1 + \left\| \tilde{X}^\top \mathbf{1} - l \right\|_1 \\ \leq \ & \left\| \tilde{X}\mathbf{1} - \tilde{r} \right\|_1 + \left\| \tilde{X}^\top \mathbf{1} - \tilde{l} \right\|_1 + \|r - \tilde{r}\|_1 + \left\| l - \tilde{l} \right\|_1 \\ \leq \ & \frac{\varepsilon'}{2} + \frac{\varepsilon'}{4} + \frac{\varepsilon'}{2} \ = \ \varepsilon'. \end{aligned}$$

We conclude that $\langle C, \hat{X} \rangle - \langle C, X^* \rangle \leq \varepsilon$ from that $\varepsilon' = \frac{\varepsilon}{8\|C\|_\infty}$. The remaining task is to analyze the complexity bound. It follows from Theorem 3.7 that

$$\begin{aligned} k \ \leq \ & 2 + \frac{112nR}{\varepsilon'} \\ \leq \ & 2 + \frac{96n\|C\|_\infty}{\varepsilon}\left(\frac{\|C\|_\infty}{\eta} + \log(n) \right. \\ & \left. -2\log\left(\min_{1 \leq i,j \leq n} \{r_i, l_j\} \right) \right) \\ \leq \ & 2 + \frac{96n\|C\|_\infty}{\varepsilon}\left(\frac{4\|C\|_\infty \log(n)}{\varepsilon} + \log(n) \right. \\ & \left. -2\log\left(\frac{\varepsilon}{64n\|C\|_\infty} \right) \right) \\ = \ & \mathcal{O}\left(\frac{n\|C\|_\infty^2 \log(n)}{\varepsilon^2} \right). \end{aligned}$$

Therefore, the total iteration complexity of the Greenkhorn algorithm can be bounded by $\mathcal{O}\left(\frac{n\|C\|_\infty^2 \log(n)}{\varepsilon^2}\right)$. Combining with the fact that each iteration of Greenkhorn algorithm requires $\mathcal{O}(n)$ arithmetic operations yields a total amount of arithmetic operations equal to $\mathcal{O}\left(\frac{n^2\|C\|_\infty^2 \log(n)}{\varepsilon^2}\right)$. On the other hand, $\tilde{r}$ and $\tilde{l}$ in Step 2 of Algorithm 2 can be found in $\mathcal{O}(n)$ arithmetic operations (cf. Algorithm 2 in Altschuler et al. (2017)), requiring $O(n^2)$ arithmetic operations. We conclude that the total number of arithmetic operations required for the Greenkhorn algorithm is $\mathcal{O}\left(\frac{n^2\|C\|_\infty^2 \log(n)}{\varepsilon^2}\right)$.

B.6. Proof of Lemma 4.1

The proof shares the same spirit with that used in Theorem 1 in Nesterov (2005). In particular, we first show that

$$\|\nabla\varphi(\lambda_1) - \nabla\varphi(\lambda_2)\|_1 \ \leq \ \frac{\|A\|_1^2}{\eta}\|\lambda_1 - \lambda_2\|_\infty . \quad (41)$$

Indeed, from the definition of $\nabla\varphi(\lambda)$, we have

$$\begin{aligned} \|\nabla\varphi(\lambda_1) - \nabla\varphi(\lambda_2)\|_1 &= \|Ax(\lambda_1) - Ax(\lambda_2)\|_1 \quad (42) \\ &\leq \|A\|_1 \|x(\lambda_1) - x(\lambda_2)\|_1 . \end{aligned}$$

We also observe from the strong convexity of f that

$$\begin{aligned} & \eta\|x(\lambda_1) - x(\lambda_2)\|_1^2 \\ \leq \ & \langle \nabla f(x(\lambda_1)) - \nabla f(x(\lambda_2)), x(\lambda_1) - x(\lambda_2) \rangle \\ = \ & \left\langle A^\top\lambda_2 - A^\top\lambda_1, x(\lambda_1) - x(\lambda_2) \right\rangle \\ \leq \ & \|\lambda_1 - \lambda_2\|_\infty \|Ax(\lambda_1) - Ax(\lambda_2)\|_1 \\ \leq \ & \|A\|_1 \|x(\lambda_1) - x(\lambda_2)\|_1 \|\lambda_1 - \lambda_2\|_\infty , \end{aligned}$$

which implies

$$\|x(\lambda_1) - x(\lambda_2)\|_1 \ \leq \ \frac{\|A\|_1}{\eta}\|\lambda_1 - \lambda_2\|_\infty . \quad (43)$$

We conclude (41) by combining (42) and (43). To this end, we have

$$\begin{aligned} & \varphi(\lambda_1) - \varphi(\lambda_2) - \langle \nabla\varphi(\lambda_2), \lambda_1 - \lambda_2 \rangle \\ = \ & \int_0^1 \langle \nabla\varphi(t\lambda_1 + (1-t)\lambda_2) - \nabla\varphi(\lambda_2), \lambda_1 - \lambda_2 \rangle \ dt \\ \leq \ & \left(\int_0^1 \|\nabla\varphi(t\lambda_1 + (1-t)\lambda_2) - \nabla\varphi(\lambda_2)\|_1 \ dt \right) \\ & \times \|\lambda_1 - \lambda_2\|_\infty \\ \overset{(41)}{\leq} \ & \left(\int_0^1 t \ dt \right) \frac{\|A\|_1^2}{\eta}\|\lambda_1 - \lambda_2\|_\infty^2 \\ = \ & \frac{\|A\|_1^2}{2\eta}\|\lambda_1 - \lambda_2\|_\infty^2 . \end{aligned}$$

This completes the proof of the lemma.

B.7. Proof of Theorem 4.2

From Lemma A.2, we have

$$\begin{aligned}
& \bar{\alpha}^k \varphi(\lambda^k) \\
\leq \; & \min_{z \in \mathbb{R}^n} \left\{ \sum_{j=0}^{k} \left[\alpha^j \left(\varphi(\mu^j) + \left\langle \nabla\varphi(\mu^j), z - \mu^j \right\rangle\right)\right] \right. \\
& \left. + \|z\|_\infty^2 \right\} \\
\leq \; & \min_{z \in B_\infty(2\widehat{R})} \left\{ \sum_{j=0}^{k} \left[\alpha^j \left(\varphi(\mu^j) + \left\langle \nabla\varphi(\mu^j), z - \mu^j \right\rangle\right)\right] \right. \\
& \left. + \|z\|_\infty^2 \right\},
\end{aligned}$$

where $\widehat{R} = \eta(R + 1/2)$ is the upper bound for ℓ_∞-norm of optimal solutions of dual regularized OT problem (15) and $B_\infty(r)$ is defined as

$$B_\infty(r) := \left\{\lambda \in \mathbb{R}^n \mid \|\lambda\|_\infty \leq r\right\}.$$

This implies that

$$\begin{aligned}
\bar{\alpha}^k \varphi(\lambda^k) \leq \; & \min_{z \in B_\infty(2\widehat{R})} \left\{ \sum_{j=0}^{k} \left[\alpha^j \left(\varphi(\mu^j) \right.\right.\right. \\
& \left.\left.\left. + \left\langle \nabla\varphi(\mu^j), z - \mu^j \right\rangle\right)\right]\right\} + 4\widehat{R}^2.
\end{aligned}$$

By the definition of the dual objective function $\varphi(\lambda)$, we further have

$$\begin{aligned}
& \varphi(\mu^j) + \left\langle \nabla\varphi(\mu^j), z - \mu^j \right\rangle \\
= \; & \left\langle \mu^j, b - Ax(\mu^j) \right\rangle - f(x(\mu^j)) \\
+ \; & \left\langle z - \mu^j, b - Ax(\mu^j) \right\rangle \\
= \; & -f(x(\mu^j)) + \left\langle z, b - Ax(\mu^j) \right\rangle.
\end{aligned}$$

Therefore, we conclude that

$$\begin{aligned}
& \bar{\alpha}^k \varphi(\lambda^k) \\
\leq \; & \min_{z \in B_\infty(2\widehat{R})} \left\{ \sum_{j=0}^{k} \left[\alpha^j \left(\varphi(\mu^j) + \left\langle \nabla\varphi(\mu^j), z - \mu^j \right\rangle\right)\right] \right\} \\
& + 4\widehat{R}^2 \\
\leq \; & 4\widehat{R}^2 - \bar{\alpha}^k f(x^k) + \min_{z \in B_\infty(2\widehat{R})} \left\{\bar{\alpha}^k \left\langle z, b - Ax^k \right\rangle\right\} \\
= \; & 4\widehat{R}^2 - \bar{\alpha}^k f(x^k) - 2\bar{\alpha}^k \widehat{R} \left\|Ax^k - b\right\|_1,
\end{aligned}$$

where the second inequality comes from the convexity of f and the last equality comes from the fact that ℓ_1-norm is the dual norm of ℓ_∞-norm. That is to say,

$$f(x^k) + \varphi(\lambda^k) + 2\widehat{R} \left\|Ax^k - b\right\|_1 \leq \frac{4\widehat{R}^2}{\bar{\alpha}^k}.$$

By the definition of $\varphi(\lambda)$ and the fact that λ^* is an optimal solution, we have

$$\begin{aligned}
f(x^k) + \varphi(\lambda^k) \; & \geq \; f(x^k) + \varphi(\lambda^*) \\
& = \; f(x^k) + \langle \lambda^*, b \rangle \\
& + \; \max_{x \in \mathbb{R}^n} \left\{-f(x) - \left\langle A^\top \lambda^*, x \right\rangle\right\} \\
& \geq \; f(x^k) + \langle \lambda^*, b \rangle - f(x^k) - \left\langle \lambda^*, Ax^k \right\rangle \\
& = \; \left\langle \lambda^*, b - Ax^k \right\rangle \\
& \geq \; -\widehat{R} \left\|Ax^k - b\right\|_1,
\end{aligned}$$

where the last inequality comes from the Hölder inequality and $\|\lambda\|_\infty \leq \widehat{R}$. We conclude that

$$\left\|Ax^k - b\right\|_1 \leq \frac{4\widehat{R}}{\bar{\alpha}^k} \overset{(27)}{\leq} \frac{32\gamma(R + 1/2)\|A\|_1^2}{(k+1)^2},$$

and obtain the desired bound on the number of iterations k required to satisfy the bound $\left\|A \text{ vec}(X^k) - b\right\|_1 \leq \varepsilon'$.

B.8. Proof of Theorem 4.3

We follow the same steps as those in the proof of Theorem 1 in Altschuler et al. (2017) and obtain

$$\begin{aligned}
& \left\langle C, \hat{X} \right\rangle - \langle C, X^* \rangle \\
\leq \; & 2\eta \log(n) + 4\left(\left\|\tilde{X}\mathbf{1} - r\right\|_1 + \left\|\tilde{X}^\top \mathbf{1} - l\right\|_1\right) \|C\|_\infty \\
\leq \; & \frac{\varepsilon}{2} + 4\left(\left\|\tilde{X}\mathbf{1} - r\right\|_1 + \left\|\tilde{X}^\top \mathbf{1} - l\right\|_1\right) \|C\|_\infty,
\end{aligned}$$

where $\hat{X}$ is the output of Algorithm 3, X^* is a solution to the optimal transport problem and $\tilde{X}$ is the matrix returned by the APDAMD algorithm (Algorithm 4) with $\tilde{r}$, $\tilde{l}$ and $\varepsilon'/2$ in Step 3 of this algorithm. The last inequality in this display holds since $\eta = \frac{\varepsilon}{4\log(n)}$. Furthermore, we have

$$\begin{aligned}
& \left\|\tilde{X}\mathbf{1} - r\right\|_1 + \left\|\tilde{X}^\top \mathbf{1} - l\right\|_1 \\
\leq \; & \left\|\tilde{X}\mathbf{1} - \tilde{r}\right\|_1 + \left\|\tilde{X}^\top \mathbf{1} - \tilde{l}\right\|_1 + \|r - \tilde{r}\|_1 + \left\|l - \tilde{l}\right\|_1 \\
\leq \; & \frac{\varepsilon'}{2} + \frac{\varepsilon'}{4} + \frac{\varepsilon'}{2} = \varepsilon'.
\end{aligned}$$

We conclude that $\langle C, \hat{X} \rangle - \langle C, X^* \rangle \leq \varepsilon$ given that $\varepsilon' = \frac{\varepsilon}{8\|C\|_\infty}$. The remaining step is to analyze the complexity bound. We obtain from Lemma 3.2 and $\tilde{r}$ and $\tilde{l}$ in Algorithm 3 that

$$\begin{aligned}
R \; & = \; \frac{\|C\|_\infty}{\eta} + \log(n) - 2\log\left(\min_{1 \leq i,j \leq n} \left\{\tilde{r}_i, \tilde{l}_j\right\}\right) \quad (44) \\
& \leq \; \frac{4\|C\|_\infty \log(n)}{\varepsilon} + \log(n) - 2\log\left(\frac{\varepsilon}{64n\|C\|_\infty}\right).
\end{aligned}$$

Since $\|A\|_1$ equals to the maximum ℓ_1-norm of a column of A and each column of A contains only two nonzero elements which are equal to one, we have $\|A\|_1 = 2$. We conclude by Lemma A.1 and Theorem 4.2 that

$$\begin{aligned} N_k &\leq 4k + 4 + \frac{2\log\left(\frac{\|A\|_1^2}{2\eta}\right) - 2\log(L^0)}{\log 2} \\ &\leq 8 + 16\sqrt{2}\,\|A\|_1\sqrt{\frac{\gamma(R+1/2)}{\varepsilon'}} \\ &\quad + \frac{2\log\left(\frac{\|A\|_1^2}{2\eta}\right) - 2\log(L^0)}{\log 2} \\ &= 8 + \frac{256}{\sqrt{\varepsilon}}\sqrt{\gamma(R+1/2)\,\|C\|_\infty \log(n)} \\ &\quad + \frac{2\log\left(\frac{\log(n)}{\varepsilon}\right)}{\log 2}. \end{aligned}$$

Plugging (44) into the above inequality yields that

$$\begin{aligned} & N_k \\ \leq\; & \frac{256}{\sqrt{\varepsilon}}\left(\frac{4\,\|C\|_\infty \log(n)}{\varepsilon} + \log(n)\right. \\ & \left. -2\log\left(\frac{\varepsilon}{64n\,\|C\|_\infty}\right) + \frac{1}{2}\right)^{1/2}\left(\sqrt{\gamma\,\|C\|_\infty \log(n)}\right) \\ & + \frac{2\log\left(\frac{\log(n)}{\varepsilon}\right)}{\log 2} + 8 \\ =\; & \mathcal{O}\left(\frac{\sqrt{\gamma}\,\|C\|_\infty \log(n)}{\varepsilon}\right). \end{aligned}$$

Therefore, the total number of iterations for the APDAMD algorithm can be bounded by $\mathcal{O}\left(\frac{\sqrt{\gamma}\|C\|_\infty \log(n)}{\varepsilon}\right)$. Combined with the fact that each iteration of APDAMD algorithm requires $O(n^2)$ arithmetic operations we find that the total number of arithmetic operations is $\mathcal{O}\left(\frac{n^2\sqrt{\gamma}\|C\|_\infty \log(n)}{\varepsilon}\right)$. Furthermore, $\tilde{r}$ and $\tilde{l}$ in Step 2 of Algorithm 3 can be found in $\mathcal{O}(n)$ arithmetic operations and Algorithm 2 in (Altschuler et al., 2017) requires $\mathcal{O}(n^2)$ arithmetic operations. Therefore, we conclude that the total number of arithmetic operations is $\mathcal{O}\left(\frac{n^2\sqrt{\gamma}\|C\|_\infty \log(n)}{\varepsilon}\right)$.

B.9. Proof of Proposition 5.2

Given the choices of r, l and η, we can rewrite the dual function $\varphi(\alpha, \beta)$ in (15) as follows:

$$\begin{aligned} \varphi(\alpha,\beta) = & \frac{\varepsilon}{4e\log(n)}\sum_{1\leq i,j\leq n} e^{-\frac{4\log(n)}{\varepsilon}(1-\alpha_i-\beta_j)} \\ & - \frac{\sum_{i=1}^n \alpha_i}{n} - \frac{\sum_{i=1}^n \beta_i}{n}. \end{aligned}$$

Since (α^*, β^*) is the optimal solution of dual regularized OT problem (15), we have

$$\begin{aligned} & e^{\frac{4\log(n)\alpha_i^*}{\varepsilon}}\sum_{j=1}^n e^{-\frac{4\log(n)}{\varepsilon}\left(1-\beta_j^*\right)} \qquad (45) \\ = \; & e^{\frac{4\log(n)\beta_i^*}{\varepsilon}}\sum_{j=1}^n e^{-\frac{4\log(n)}{\varepsilon}\left(1-\alpha_j^*\right)} \\ = \; & \frac{e}{n}, \quad \forall i \in [n]. \end{aligned}$$

This implies that $\alpha_i^* = \alpha_j^*$ and $\beta_i^* = \beta_j^*$ for all $i, j \in [n]$. So we can define A and B such that

$$A \equiv e^{\frac{4\log(n)\alpha_i^*}{\varepsilon}}, \quad B \equiv e^{\frac{4\log(n)\beta_i^*}{\varepsilon}}.$$

By the optimality condition (45), $ABe^{-4\log(n)/\varepsilon} = e/n^2$. Equivalently, $AB = \frac{e^{\frac{4\log(n)}{\varepsilon}+1}}{n^2}$. So we have

$$\begin{aligned} \alpha_i^* + \beta_i^* &= \frac{\varepsilon\left(\log(A) + \log(B)\right)}{4\log(n)} \\ &= \frac{\varepsilon}{4\log(n)}\left(\frac{4\log(n)}{\varepsilon} + 1 - 2\log(n)\right) \\ &= 1 + \frac{\varepsilon}{4\log(n)} - \frac{\varepsilon}{2}. \end{aligned}$$

Therefore, we conclude that

$$\begin{aligned} \|(\alpha^*, \beta^*)\|_2 &\geq \sqrt{\frac{\sum_{i=1}^n(\alpha_i^* + \beta_i^*)^2}{2}} \\ &= \sqrt{\frac{n}{2}}\left(1 + \frac{\varepsilon}{4\log(n)} - \frac{\varepsilon}{2}\right) \\ &\gtrsim n^{1/2}. \end{aligned}$$

As a consequence, we achieve the conclusion of the proposition.

B.10. Proof of Proposition 5.3

The proof of Proposition 5.3 is a modification of the proof for Theorem 4 in Dvurechensky et al. (2018). Therefore, we only give a proof sketch to ease the presentation. More specifically, we follow the argument of Theorem 4 in Dvurechensky et al. (2018) and obtain that the number of iterations for Algorithm 5 required to reach the tolerance ε is

$$\begin{aligned} k \leq \; & \max\left\{\mathcal{O}\left(\min\left\{\frac{n^{1/4}\sqrt{\overline{R}\,\|C\|_\infty \log(n)}}{\varepsilon},\right.\right.\right. \qquad (46) \\ & \left.\left.\left.\frac{\overline{R}\,\|C\|_\infty \log(n)}{\varepsilon^2}\right\}\right), \mathcal{O}\left(\frac{\overline{R}\sqrt{\log n}}{\varepsilon}\right)\right\}. \end{aligned}$$

Plugging the tight upper bound $\overline{R} \leq \sqrt{n}$ into (46) yields that

$$k = \mathcal{O}\left(\frac{\sqrt{n}\|C\|_\infty \log(n)}{\varepsilon}\right).$$

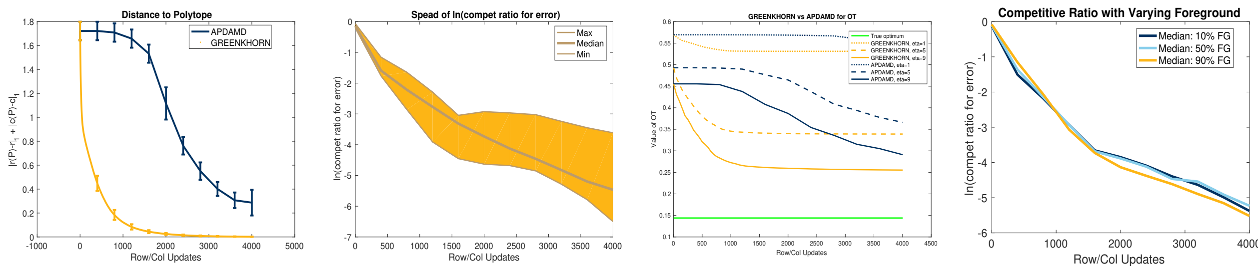

Figure 3. Performance of the Greenkhorn and APDAMD algorithms on the synthetic images. All the four images correspond to those in the figure in the main context, showing that the Greenkhorn algorithm is faster than the APDAMD algorithm in terms of iterations. Note that $\log(d(P_G)/d(P_{MD}))$ on ten random pairs of images is consistently used, where $d(P_G)$ and $d(P_{MD})$ refer to the Greenkhorn and APDAMD algorithms, respectively.

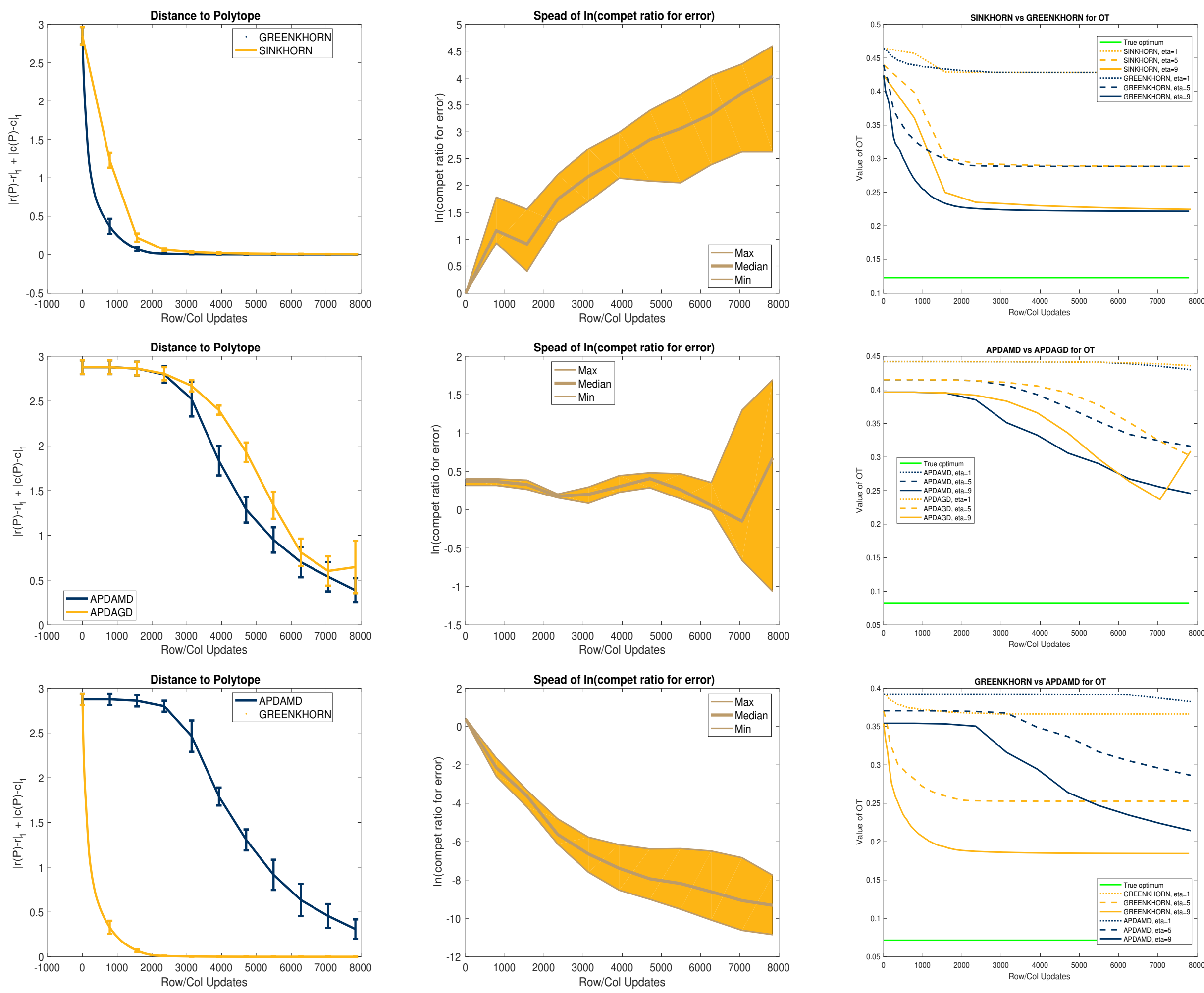

Figure 4. Performance of the Sinkhorn, Greenkhorn, APDAGD and APDAMD algorithms on the MNIST real images. In the first row of images, we compare the Sinkhorn and Greenkhorn algorithms in terms of iteration counts. The leftmost image specifies the distances $d(P)$ to the transportation polytope for two algorithms; the middle image specifies the maximum, median and minimum of competitive ratios $\log(d(P_S)/d(P_G))$ on ten random pairs of MNIST images, where P_S and P_G stand for the outputs of Sinkhorn and Greenkhorn, respectively; the rightmost image specifies the values of regularized OT with varying regularization parameter $\eta \in \{1, 5, 9\}$. In addition, the second and third rows of images present comparative results for APDAGD versus APDAMD and Greenkhorn versus APDAMD. In summary, the experimental results on the MNIST images are consistent with that on the synthetic images.

Since each iteration of the APDAGD algorithm requires $O(n^2)$ arithmetic operations, the total number of arithmetic operations is bounded by $\mathcal{O}\left(\frac{n^{5/2}\|C\|_\infty \log(n)}{\varepsilon}\right)$. Furthermore, $\tilde{r}$ and $\tilde{l}$ in Step 2 of Algorithm 5 can be found in $\mathcal{O}(n)$ arithmetic operations and Algorithm 2 in Altschuler et al. (2017) requires $\mathcal{O}(n^2)$ arithmetic operations. Therefore, we conclude that the total number of arithmetic operations required is $\mathcal{O}\left(\frac{n^{5/2}\sqrt{\|C\|_\infty} \log(n)}{\varepsilon}\right)$.

C. Further experiments

In this section, we present some additional experimental results. In particular, we compare the Greenkhorn and APDAMD algorithms on synthetic image and conduct the extensive comparative experiments with the Greenkhorn and APDAMD algorithms on real images from MNIST Digits dataset[2]. We use essentially the same baseline algorithms and evaluation metrics as in the synthetic images in Section 6.

Image processing: The MNIST dataset consists of 60,000 images of handwritten digits of size 28 by 28 pixels. To understand better the dependence on n for our algorithms, we add a very small noise term (10^{-6}) to all the zero elements in the measures and then normalize them such that their sum becomes one.

Experimental results: We present the experimental results on the comparison between the Greenkhorn and APDAMD algorithms on synthetic images in Figure 3. Despite the worse dependence of ε in the complexity bound, the Greenkhorn algorithm seems more practical than the APDAMD algorithm. The possible reason is the advantage of the coordinate descent algorithm over the gradient descent algorithm, which is worthy further exploration.

Figure 4 presents the experimental results on the real images with different choices of regularization parameters as well as the coverage ratio of the foreground on the real images. The Greenkhorn algorithm is the fastest among all the candidate algorithms in terms of iteration count. Also, the APDAMD algorithm outperforms the APDAGD algorithm in terms of robustness and efficiency. All the results on real images are consistent with those on the synthetic images.

[2]http://yann.lecun.com/exdb/mnist/