APPENDIX

Algorithm 4 Adaptive Primal-Dual Accelerated Gradient Descent (APDAGD) [29, Algorithm 3]

Input: accuracy $\varepsilon_f, \varepsilon_{eq} > 0$; initial estimate L_0 such that $0 < L_0 < 2L$; initialize $M_{-1} = L_0$, $i_0 = k = \alpha_0 = \beta_0 = 0$, $\boldsymbol{\eta}_0 = \boldsymbol{\zeta}_0 = \boldsymbol{\lambda}_0 = \mathbf{0}$.

while $f(\hat{\mathbf{x}}_{k+1}) + \varphi(\boldsymbol{\eta}_{k+1}) \le \varepsilon_f, \|\mathbf{A}\hat{\mathbf{x}}_{k+1} - \mathbf{b}\|_2 \le \varepsilon_{eq}$ **do**
  while $\varphi(\boldsymbol{\eta}_{k+1}) \le \varphi(\boldsymbol{\lambda}_{k+1}) + \langle \nabla\varphi(\boldsymbol{\lambda}_{k+1}), \boldsymbol{\eta}_{k+1} - \boldsymbol{\lambda}_{k+1}\rangle + \frac{M_k}{2}\|\boldsymbol{\eta}_{k+1} - \boldsymbol{\lambda}_{k+1}\|_2^2$ **do**
    $M_k = 2^{i_k - 1} M_k$
    find α_{k+1} such that $\beta_{k+1} := \beta_k + \alpha_{k+1} = M_k \alpha_{k+1}^2$
    $\tau_k = \dfrac{\alpha_{k+1}}{\beta_{k+1}}$
    $\boldsymbol{\lambda}_{k+1} = \tau_k \boldsymbol{\zeta}_k + (1 - \tau_k)\boldsymbol{\eta}_k$
    $\boldsymbol{\zeta}_{k+1} = \boldsymbol{\zeta}_k - \alpha_{k+1}\nabla\varphi(\boldsymbol{\lambda}_{k+1})$
    $\boldsymbol{\eta}_{k+1} = \tau_k \boldsymbol{\zeta}_{k+1} + (1 - \tau_k)\boldsymbol{\eta}_k$
  end while
  $\hat{\mathbf{x}}_{k+1} = \tau_k \mathbf{x}(\boldsymbol{\lambda}_{k+1}) + (1 - \tau_k)\hat{\mathbf{x}}_k$
  $i_{k+1} = 0, k = k + 1$
end while

Output: $\hat{\mathbf{x}}_{k+1}, \boldsymbol{\eta}_{k+1}$.

Algorithm 5 CombiOT [18]

Input: Demands SA, Sources DB; Cost $\mathbf{C}$; accuracy ε.

$\mathbf{C} \leftarrow \varepsilon \lfloor \mathbf{C}/\varepsilon \rfloor$
$M \leftarrow \emptyset$
Dual weights $y(a) = 0 \quad \forall a \in \mathbb{A}, y(b) = \varepsilon \quad \forall b \in B$
$B' \leftarrow$ free vertices of B
while $\|B'\| > \varepsilon n$ **do**
  $E' \leftarrow$ set of admissible edges with at least one end point in B'.
  $A' = \{a | a \in A \text{ and } (a,b) \in E'\}$
  Compute maximal matching M' in $G'(A' \cup B', E')$
  $A'' \leftarrow a$s matched in both M and M'
  $M'' \leftarrow$ edges of M incident on some vertex of A''.
  Adds edges of M' and deletes edges of M'' in M.
  $y(a) \leftarrow y(a) - \varepsilon \quad \forall (a,b) \in M'$
  $y(b) \leftarrow y(b) + \varepsilon \quad \forall b \in B'$ free with respect to M'
end while
Arbitrary match free vertices and return resulting matching

Output: Transport plan F, dual values yA, yB, totalcost, iteration

A. Lagrangian duality

Consider an optimization problem as follows:

$$\min \quad f_0(x) \tag{19}$$
$$\text{s.t} \quad f_i(x) \le 0 \quad \forall i = 1, ..., n \tag{20}$$

Let p^* be the solution to (19). We further denote the Lagrangian of (19) to be:

$$L(\lambda, x) = f_0(x) + \sum_{i=1}^{n} \lambda_i f_i(x)$$

with $\lambda_i \ge 0 \quad \forall i \in \{1, ..., n\}$

Now, by setting $g(\lambda) = \min_x L(\lambda, x)$, we can formulate our problem into a dual version. Our target would be to find d^* satisfying:

$$d^* = \max_{\lambda} g(\lambda) \tag{21}$$

$g(\lambda)$ and (21) are known as the Lagrange dual function and the Lagrangian dual problem . Additionally, the solution d^* of (21) is called the dual solution and is always less than or equal to p^*, the primal solution . Strong duality holds when $p^* = d^*$. This occurs only if the optimization problem is convex and a strictly feasible point exists, i.e., a point x that strictly satisfies all constraints.

B. Lagrangian Dual of Quadratic Optimal Transport

Let $\mathbf{y} \in \mathbb{R}^n, \mathbf{z} \in \mathbb{R}^n$ and $t \in \mathbb{R}$ be the dual variables corresponding to the three equality constraints. The Lagrangian is

$$\begin{aligned}\mathscr{L}(\mathbf{x}, \gamma) &= \langle \mathbf{C}, \mathbf{X}\rangle + \gamma(\|\mathbf{X}\|^2 + \|\mathbf{p}\|^2 + \|\mathbf{q}\|^2) \\ &\quad - \langle \mathbf{y}, \mathbf{X}\mathbf{1} + \mathbf{p} - \mathbf{r}\rangle - \langle \mathbf{z}, \mathbf{X}^\top \mathbf{1} + \mathbf{q} - \mathbf{c}\rangle - t(\mathbf{1}^\top \mathbf{X}\mathbf{1} - s) \\ &= \langle \mathbf{y}, \mathbf{r}\rangle + \langle \mathbf{z}, \mathbf{c}\rangle + st + \langle \mathbf{C}, \mathbf{X}\rangle + \gamma\|\mathbf{X}\|^2 \langle \mathbf{y}, \mathbf{X}\mathbf{1}\rangle - \langle \mathbf{z}, \mathbf{X}^\top \mathbf{1}\rangle \\ &\quad - t\mathbf{1}^\top \mathbf{X}\mathbf{1} + \gamma\|\mathbf{p}\|^2 - \langle \mathbf{y}, \mathbf{p}\rangle + \gamma\|\mathbf{q}\|^2 - \langle \mathbf{z}, \mathbf{q}\rangle.\end{aligned}$$

So, we have the dual problem:

$$\begin{aligned}&\max_{\mathbf{y},\mathbf{z},t} F_\gamma(\mathbf{x} = (\mathbf{y}, \mathbf{z}, t)) \\ &= \left\{ \langle \mathbf{y}, \mathbf{r}\rangle + \langle \mathbf{z}, \mathbf{c}\rangle + ts + \min_{X_{i,j} \ge 0} \sum_{i,j=1}^{n} \gamma X_{i,j}^2 + X_{i,j}(C_{i,j} - y_i - z_j - t) \right. \\ &\quad \left. + \min_{p_i \ge 0} \sum_{i=1}^{n} p_i(\gamma p_i - y_i) + \min_{q_j \ge 0} \sum_{j=1}^{n} q_j(\gamma q_j - z_j) \right\}.\end{aligned} \tag{22}$$

Next, we minimize the Lagrangian w.r.t. $\mathbf{X}$, $\mathbf{p}$ and $\mathbf{q}$ according to the first order condition, which gives

$$X_{i,j} = -\min\left\{0, \frac{C_{i,j}-y_i-z_j-t}{2\gamma}\right\} = \max\left\{0, \frac{y_i+z_j+t-C_{i,j}}{2\gamma}\right\}, \tag{23}$$

$$\mathbf{p}_i = \max\left\{0, \frac{y_i}{2\gamma}\right\}, \tag{24}$$

$$\mathbf{q}_j = \max\left\{0, \frac{z_j}{2\gamma}\right\}. \tag{25}$$

Proof:

The optimal value for $X_{i,j} \geq 0$ is depending only on $X_{i,j}^2 + X_{i,j}(C_{i,j}-y_i-z_j-t)$, a quadratic function.
If $-\frac{C_{i,j}-y_i-z_j-t}{2\gamma} \geq 0$, we can choose $X_{i,j}$ to be that value. But else, the function increases on $\mathbb{R}_{\geq 0}$, so $X_{i,j}=0$ is optimal. Same arguments apply for finding optimal p_i, q_i since they also only depend on quadratic functions. ■

Plugging this back to the dual problem (22), we have

$$\begin{aligned}\max_{\mathbf{y},\mathbf{z},t}\Bigg\{ & g(\mathbf{x}=(\mathbf{y},\mathbf{z},t)) = \langle \mathbf{y},\mathbf{r}\rangle + \langle \mathbf{z},\mathbf{c}\rangle + ts \\ & - \sum_{i,j=1}^{n} \frac{\max(y_i+z_j+t-C_{i,j},0)^2}{4\gamma} \\ & - \sum_{i=1}^{n} \frac{\max(y_i,0)^2}{4\gamma} - \sum_{i=1}^{n} \frac{\max(z_i,0)^2}{4\gamma} \Bigg\}.\end{aligned} \tag{26}$$

C. Proof of Lemma 1

We first find upper bounds for y_i^* and z_j^*, to derive the lower bound and upper bound for t^*. Finally, we use upper bound for t^* to derive lower bound for y_i^*, z_j^*.
For all i such that y_i^* is positive

$$y_i^* = 2\gamma p_i \leq 2\gamma c_i$$

So in general, $y_i^* \leq \max\{0, 2\gamma c_i\} \leq 2\gamma c_i$.
Similarly, we also have $z_j^* \leq 2\gamma \|\mathbf{r}\|_\infty$.
Since $\mathbf{X}_\gamma$ has positive s as sum of all of its nonnegative entries, it is not a zero matrix. So there exist i, j such that $X_{i,j,\gamma} > 0$, and of course $X_{i,j,\gamma} \leq s$, then for that i, j we have

$$\begin{aligned} & 0 < -C_{ij} + y_i^* + z_j^* + t^* \leq 2\gamma s \\ \Rightarrow & t^* > C_{ij} - y_i^* - z_j^* > -2\gamma(\|\mathbf{c}\|_\infty + \|\mathbf{r}\|_\infty) \\ & \text{(based on the bounds for } y_i \text{ and } z_j \text{ found above)}\end{aligned}$$

Since $\|\mathbf{X}\|_1 = s < \min(\|\mathbf{r}\|_1, \|\mathbf{c}\|_1)$, $\mathbf{p}^*, \mathbf{q}^*$ are not zero vectors, which means there exist i, j such that p_i^*, q_j^* are positive numbers, and therefore, from (11) and (12), y_i, z_j are positive numbers too.
So we have

$$\begin{aligned} & X_{i,j} = \max\left\{0, \frac{y_i^*+z_j^*+t^*-C_{i,j}}{2\gamma}\right\} \leq s, \\ \Rightarrow & y_i^* + z_j^* + t^* - C_{ij} \leq \max(0, 2\gamma s) \leq 2\gamma s \\ \Rightarrow & t^* \leq 2\gamma s + C_{ij} - y_i^* - z_j^* \leq 2\gamma s + \|\mathbf{C}\|_\infty\end{aligned}$$

We conclude the proof with lower bounds for y_i^* and z_j^*

$$\begin{aligned} & \forall i, \text{if } y_i^* < 0, p_i^* = 0 \\ \Rightarrow \quad & \mathbf{X}\mathbf{1}_i = c_i = r_i > 0 \\ \Rightarrow \quad & \exists j, X_{ij} > 0 \Rightarrow y_i^* + z_j^* + t^* - C_{ij} > 0 \\ \Rightarrow \quad & y_i^* > C_{ij} - z_j^* - t^* \geq -2\gamma\|\mathbf{r}\|_\infty - 2\gamma s - \|\mathbf{C}\|_\infty.\end{aligned}$$

So,

$$\begin{aligned} y_i^* &> \min\{0, -2\gamma\|\mathbf{r}\|_\infty - 2\gamma s - \|\mathbf{C}\|_\infty\} \\ &\geq -2\gamma\|\mathbf{r}\|_\infty - 2\gamma s - \|\mathbf{C}\|_\infty.\end{aligned}$$

And similarly, $z_j^* > -2\gamma\|\mathbf{c}\|_\infty - 2\gamma s - \|\mathbf{C}\|_\infty$.
With $\gamma = 0.5\varepsilon/s^2$, terms involving γ are infinitesimal compared to $\|\mathbf{C}\|_\infty$. Therefore $\|\lambda\| = \|(\mathbf{y},\mathbf{z},t)\| = \mathcal{O}(\|\mathbf{C}\|_\infty)$.

D. Proof of Theorem 4.1

Let $\mathbf{x}^{\text{POT}} = (\mathbf{X}^{\text{POT}}, \mathbf{p}^{\text{POT}}, \mathbf{q}^{\text{POT}})$ be a solution of 5 and $\overline{\mathbf{x}} = \text{ROUND-POT}(\hat{\mathbf{x}}_k)$, we have

$$\begin{aligned} \langle \mathbf{d}, \mathbf{x}^{\text{POT}}\rangle + \gamma\left\|\mathbf{x}^{\text{POT}}\right\|^2 &\geq \langle \mathbf{d}, \mathbf{x}_f\rangle + \gamma\left\|\mathbf{x}_f\right\|^2 \\ &\geq \langle \mathbf{d}, \hat{\mathbf{x}}_k\rangle + \gamma\|\hat{\mathbf{x}}_k\|^2 - \frac{32R^2}{\gamma k^2} \\ \text{So }, \langle \mathbf{d}, \hat{\mathbf{x}}_k - \mathbf{x}^{\text{POT}}\rangle &\leq \gamma(\left\|\mathbf{x}^{\text{POT}}\right\|^2 - \|\hat{\mathbf{x}}_k\|^2) + \frac{32R^2}{\gamma k^2} \\ &\leq \gamma s^2 + \frac{32R^2}{\gamma k^2}\end{aligned}$$

From [15], Theorem 6, we have that:

$$\begin{aligned} \|\hat{\mathbf{x}}_k - \overline{\mathbf{x}}\|_1 &\leq 23\,\|\mathbf{A}\hat{\mathbf{x}}_k - \mathbf{b}\|_1 \\ &\leq 23\sqrt{2n+1}\,\|\mathbf{A}\hat{\mathbf{x}}_k - \mathbf{b}\|_2 \\ \langle \mathbf{d}, \overline{\mathbf{x}} - \hat{\mathbf{x}}_k\rangle &\leq \|\mathbf{d}\|_\infty \|\hat{\mathbf{x}}_k - \overline{\mathbf{x}}\|_1 \leq \|\mathbf{C}\|_\infty \cdot 23\sqrt{2n+1}\,\|\mathbf{A}\hat{\mathbf{x}}_k - \mathbf{b}\|_2 \\ &= \|\mathbf{C}\|_\infty \cdot 23\sqrt{2n+1}\frac{32R}{\gamma k^2}\end{aligned}$$

Adding them up, we have:

$$\begin{aligned} \langle \mathbf{C}, \overline{\mathbf{X}} - \mathbf{X}^{\text{POT}}\rangle &= \langle \mathbf{d}, \overline{\mathbf{x}} - \mathbf{x}^{\text{POT}}\rangle \\ &\leq \gamma s^2 + \frac{32R^2}{\gamma k^2}\left(1 + \frac{23\,\|\mathbf{C}\|_\infty\sqrt{2n+1}}{R}\right)\end{aligned}$$

Using AM-GM, we know the best γ to minimize the right-hand side to ε would satisfy $\gamma s^2 = 0.5\varepsilon$, or $\gamma = \varepsilon/2s^2$. The other component of the sum must also be 0.5ε. Therefore

$$\begin{aligned} 0.5\varepsilon &= \frac{32R^2}{\gamma k^2}\left(1+\frac{23\|\mathbf{C}\|_\infty\sqrt{2n+1}}{R}\right) \\ &= \frac{64(sR)^2}{\varepsilon k^2}\left(1+\frac{23\|\mathbf{C}\|_\infty\sqrt{2n+1}}{R}\right) \\ \Rightarrow k^2 &= \frac{128(sR)^2}{\varepsilon^2}\left(1+\frac{23\|\mathbf{C}\|_\infty\sqrt{2n+1}}{R}\right) \end{aligned}$$

Since $R = \mathcal{O}(\|\mathbf{C}\|_\infty\sqrt{n})$, the bracket is $\mathcal{O}(1)$. So the optimal number of loops is $k = \mathcal{O}(n^{\frac{1}{2}}\|\mathbf{C}\|_\infty/\varepsilon)$, combining with $\mathcal{O}(n^2)$ for each iteration of APDAGD gives the total complexity of $\mathcal{O}(n^{\frac{5}{2}}\|\mathbf{C}\|_\infty/\varepsilon)$.

E. Color Transfer Experiments

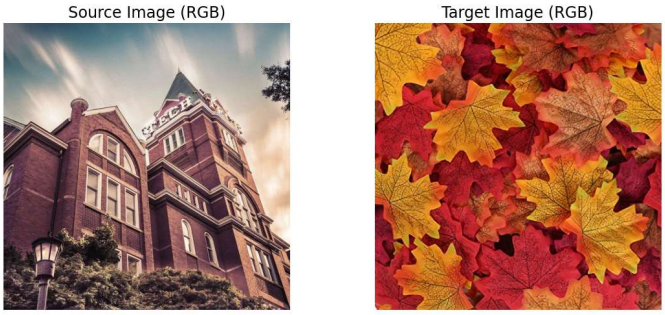

Fig. 4: Dataset for Color Transfer

Color transfer is an image editing technique that aligns the color palette of one image with another by transferring hue, saturation, and brightness while preserving contextual details. An RGB image is defined as $\mathbf{x} \in \mathbb{R}^{h\times w\times 3}$, where each pixel at coordinate (i,j) is represented as $x_{ij} = \{\text{red, green, blue}\}$ with values ranging from 0 to 255. This forms a color distribution across channels, and color transfer aims to match the source image's distribution to the target's. Our dataset consists of a pair of 256×256 images (Figure 4) with distinct color palettes. Instead of working in RGB, we convert images to LUV, where color is encoded in the U and V channels, reducing dimensionality from three (RGB) to two (UV) while better reflecting perceptual color differences. The conversion follows: $L = R+G+B$, $U = G/L$, and $V = B/L$. We extract color histograms with n bins by computing U and V frequencies, smoothing, and normalizing them to ensure $\|\mathbf{r}\|_1 = 1$ and $\|\mathbf{c}\|_1 = 1$. The cost matrix is defined as $C_{i,j} = \|a_i - b_j\|_2^2$, where a_i and b_j are histogram bin values.

The first experiment examines the transport mass, defined as $\alpha \times \min(\|r\|_1, \|c\|_1)$. Here, we vary α within the interval

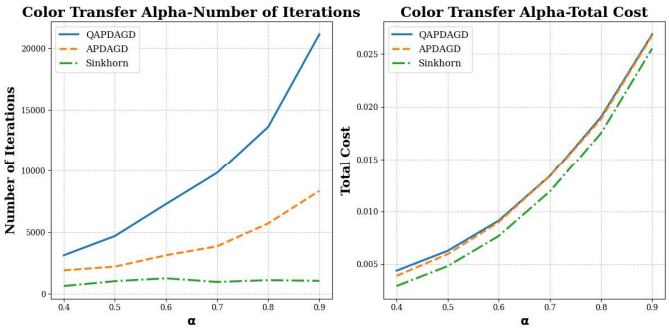

Fig. 5: Number of iterations and total cost for each value of alpha

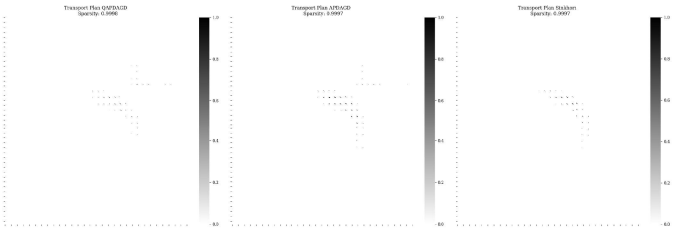

Fig. 6: Sparsity, number of iterations and total cost for each value of alpha

$[0.4, 0.9]$ and evaluate the number of iterations, and total cost for QAPDAGD, APDAGD, and Sinkhorn at each value of α in Figure 5. Additionally, we have also plotted out the sparsity of each method when $\alpha = 0.8$ in Figure 6. Overall, although QAPDAGD requires more iterations, its total cost remains comparable to other methods. Moreover, it achieves slightly better sparsity in the transportation plan compared to APDAGD and Sinkhorn.

In the second experiment, we set $s = 0.8\times\min(\|\mathbf{r}\|_1, \|\mathbf{c}\|_1)$ and tolerance $\varepsilon = 10^{-2}$, varying the sample size, i.e., the dimensions of the input and output images. Specifically, we set the image size to $n^2\times n^2$ for n ranging from 20 to 40. Figure 8 shows the total cost and computation time for each method across different sample sizes. Additionally, for $n = 25$, Figure 9 presents the heatmaps of the transportation plan for each method. In general, the total costs and sparsity of all methods are approximately the same. However, due to its inherent structure, QAPDAGD exhibits a higher runtime compared to other methods as the sample size increases.

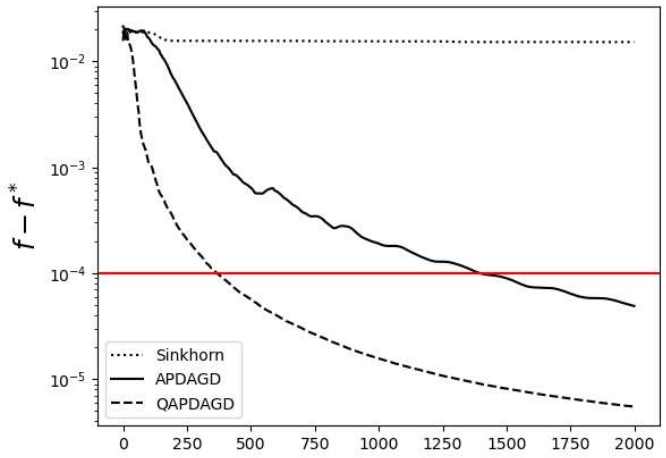

Fig. 7: Primal gap achieved by our QAPDAGD algorithm versus APDAGD and revised Sinkhorn for POT with same number of iterations and tolerance ε in red. The result transport plans are later used for Color Transfer.

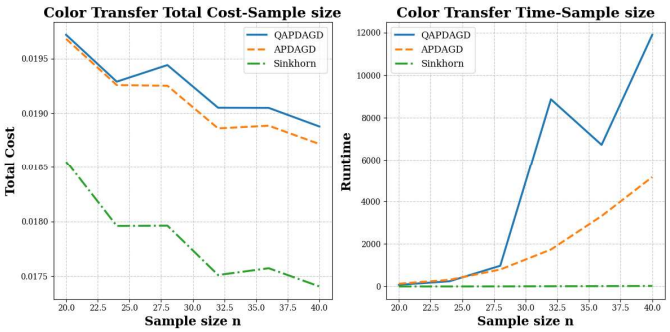

Fig. 8: Number of iterations and time taken for each value of sample size n.

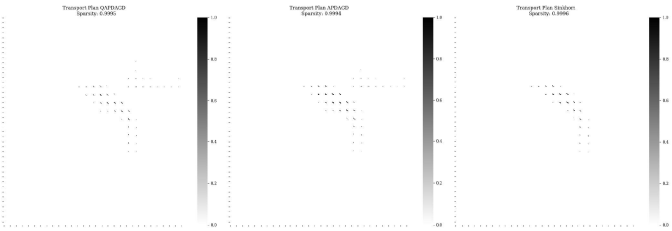

Fig. 9: The heatmaps of the transportation plans of each methods with the sample size $n = 25$.