Appendix

A Proofs

A.1 Derivation of the smooth relaxed dual

Recall that

$$\mathrm{OT}_\Omega(\boldsymbol{a}, \boldsymbol{b}) = \min_{T \in \mathcal{U}(\boldsymbol{a}, \boldsymbol{b})} \sum_{j=1}^n \boldsymbol{t}_j^\top \boldsymbol{c}_j + \Omega(\boldsymbol{t}_j). \tag{16}$$

We now add Lagrange multipliers for the two equality constraints but keep the constraint $T \geq 0$ explicitly:

$$\mathrm{OT}_\Omega(\boldsymbol{a}, \boldsymbol{b}) = \min_{T \geq 0} \max_{\boldsymbol{\alpha} \in \mathbb{R}^m, \boldsymbol{\beta} \in \mathbb{R}^n} \sum_{j=1}^n \boldsymbol{t}_j^\top \boldsymbol{c}_j + \Omega(\boldsymbol{t}_j) + \boldsymbol{\alpha}^\top (T \mathbf{1}_n - \boldsymbol{a}) + \boldsymbol{\beta}^\top (T^\top \mathbf{1}_m - \boldsymbol{b}).$$

Since (16) is a convex optimization problem with only linear equality and inequality constraints, Slater's conditions reduce to feasibility [Boyd and Vandenberghe, 2004, §5.2.3] and hence strong duality holds:

$$\begin{aligned}
\mathrm{OT}_\Omega(\boldsymbol{a}, \boldsymbol{b}) &= \max_{\boldsymbol{\alpha} \in \mathbb{R}^m, \boldsymbol{\beta} \in \mathbb{R}^n} \min_{T \geq 0} \sum_{j=1}^n \boldsymbol{t}_j^\top \boldsymbol{c}_j + \Omega(\boldsymbol{t}_j) + \boldsymbol{\alpha}^\top (T \mathbf{1}_n - \boldsymbol{a}) + \boldsymbol{\beta}^\top (T^\top \mathbf{1}_m - \boldsymbol{b}) \\
&= \max_{\boldsymbol{\alpha} \in \mathbb{R}^m, \boldsymbol{\beta} \in \mathbb{R}^n} \sum_{j=1}^n \min_{\boldsymbol{t}_j \geq 0} \boldsymbol{t}_j^\top (\boldsymbol{c}_j + \boldsymbol{\alpha} + \beta_j \mathbf{1}_m) + \Omega(\boldsymbol{t}_j) - \boldsymbol{\alpha}^\top \boldsymbol{a} - \boldsymbol{\beta}^\top \boldsymbol{b} \\
&= \max_{\boldsymbol{\alpha} \in \mathbb{R}^m, \boldsymbol{\beta} \in \mathbb{R}^n} - \sum_{j=1}^n \max_{\boldsymbol{t}_j \geq 0} \boldsymbol{t}_j^\top (-\boldsymbol{c}_j - \boldsymbol{\alpha} - \beta_j \mathbf{1}_m) - \Omega(\boldsymbol{t}_j) - \boldsymbol{\alpha}^\top \boldsymbol{a} - \boldsymbol{\beta}^\top \boldsymbol{b} \\
&= \max_{\boldsymbol{\alpha} \in \mathbb{R}^m, \boldsymbol{\beta} \in \mathbb{R}^n} \boldsymbol{\alpha}^\top \boldsymbol{a} + \boldsymbol{\beta}^\top \boldsymbol{b} - \sum_{j=1}^n \max_{\boldsymbol{t}_j \geq 0} \boldsymbol{t}_j^\top (\boldsymbol{\alpha} + \beta_j \mathbf{1}_m - \boldsymbol{c}_j) - \Omega(\boldsymbol{t}_j).
\end{aligned}$$

Finally, plugging the expression of (6) gives the claimed result.

A.2 Derivation of the convex conjugate

The convex conjugate of $\mathrm{OT}_\Omega(\boldsymbol{a}, \boldsymbol{b})$ w.r.t. the first argument is

$$\mathrm{OT}_\Omega^*(\boldsymbol{g}, \boldsymbol{b}) = \sup_{\boldsymbol{a} \in \triangle^m} \boldsymbol{g}^\top \boldsymbol{a} - \mathrm{OT}_\Omega(\boldsymbol{a}, \boldsymbol{b}).$$

Following a similar argument as [Cuturi and Peyré, 2016, Theorem 2.4], we have

$$\mathrm{OT}_\Omega^*(\boldsymbol{g}, \boldsymbol{b}) = \max_{\substack{T \geq 0 \\ T^\top \mathbf{1}_m = \boldsymbol{b}}} \langle T, \boldsymbol{g} \mathbf{1}_n^\top - C \rangle - \sum_{j=1}^n \Omega(\boldsymbol{t}_j).$$

Notice that this is an easier optimization problem than (5), since there are equality constraints only in one direction. Cuturi and Peyré [2016] showed that this optimization problem admits a closed form in the case of entropic regularization. Here, we show how to compute OT_Ω^* for any strongly-convex regularization.

The problem clearly decomposes over columns and we can rewrite it as

$$\begin{aligned}
\mathrm{OT}_\Omega^*(\boldsymbol{g}, \boldsymbol{b}) &= \sum_{j=1}^n \max_{\substack{\boldsymbol{t}_j \geq 0 \\ \boldsymbol{t}_j^\top \mathbf{1}_m = b_j}} \boldsymbol{t}_j^\top (\boldsymbol{g} - \boldsymbol{c}_j) - \Omega(\boldsymbol{t}_j) \\
&= \sum_{j=1}^n b_j \max_{\boldsymbol{\tau}_j \in \triangle^m} \boldsymbol{\tau}_j^\top (\boldsymbol{g} - \boldsymbol{c}_j) - \frac{1}{b_j} \Omega(b_j \boldsymbol{\tau}_j) \\
&= \sum_{j=1}^n b_j \max\nolimits_{\Omega_j} (\boldsymbol{g} - \boldsymbol{c}_j),
\end{aligned}$$

where we defined $\Omega_j(\boldsymbol{y}) \coloneqq \frac{1}{b_j}\Omega(b_j \boldsymbol{y})$ and where $\max_\Omega$ is defined in (8).

A.3 Expression of the strongly-convex duals

Using a similar derivation as before, we obtain the duals of (13) and (14).

Proposition 3 *Duals of* (13) *and* (14)

$$\begin{aligned}
\mathrm{ROT}_\Phi(\boldsymbol{a}, \boldsymbol{b}) &= \max_{\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathcal{P}(C)} -\frac{1}{2}\Phi^*(-2\boldsymbol{\alpha}, \boldsymbol{a}) - \frac{1}{2}\Phi^*(-2\boldsymbol{\beta}, \boldsymbol{b}) \\
\widetilde{\mathrm{ROT}}_\Phi(\boldsymbol{a}, \boldsymbol{b}) &= \max_{\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathcal{P}(C)} -\Phi^*(-\boldsymbol{\alpha}, \boldsymbol{a}) + \boldsymbol{\beta}^\top \boldsymbol{b} \\
&= \max_{\boldsymbol{\alpha} \in \mathbb{R}^m} -\Phi^*(-\boldsymbol{\alpha}, \boldsymbol{a}) - \sum_{j=1}^n b_j \max_{i \in [m]} (\alpha_i - c_{i,j}),
\end{aligned}$$

where Φ^* *is the conjugate of* Φ *in the first argument.*

The duals are strongly convex if Φ is smooth.
When $\Phi(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{2\gamma}\|\boldsymbol{x} - \boldsymbol{y}\|^2$, $\Phi^*(-\boldsymbol{\alpha}, \boldsymbol{a}) = \frac{\gamma}{2}\|\boldsymbol{\alpha}\|^2 - \boldsymbol{\alpha}^\top \boldsymbol{a}$. Plugging that expression in the above, we get

$$\mathrm{ROT}_\Phi(\boldsymbol{a}, \boldsymbol{b}) = \max_{\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathcal{P}(C)} \boldsymbol{\alpha}^\top \boldsymbol{a} + \boldsymbol{\beta}^\top \boldsymbol{b} - \gamma \left(\|\boldsymbol{\alpha}\|^2 + \|\boldsymbol{\beta}\|^2\right) \tag{17}$$

and

$$\begin{aligned}
\widetilde{\mathrm{ROT}}_\Phi(\boldsymbol{a}, \boldsymbol{b}) &= \max_{\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathcal{P}(C)} \boldsymbol{\alpha}^\top \boldsymbol{a} + \boldsymbol{\beta}^\top \boldsymbol{b} - \frac{\gamma}{2}\|\boldsymbol{\alpha}\|^2 \\
&= \max_{\boldsymbol{\alpha} \in \mathbb{R}^m} \boldsymbol{\alpha}^\top \boldsymbol{a} - \sum_{j=1}^n b_j \max_{i \in [m]} (\alpha_i - c_{i,j}) - \frac{\gamma}{2}\|\boldsymbol{\alpha}\|^2.
\end{aligned}$$

This corresponds to the original dual and semi-dual with squared 2-norm regularization on the variables.

A.4 Proof of Theorem 1

Before proving the theorem, we introduce the next two lemmas, which bound the regularization value achieved by any transportation plan.

Lemma 2 *Bounding the entropy of a transportation plan*

Let $H(\boldsymbol{a}) \coloneqq -\sum_i a_i \log a_i$ *and* $H(T) \coloneqq -\sum_{i,j} t_{i,j} \log t_{i,j}$ *be the joint entropy.*
Let $\boldsymbol{a} \in \triangle^m$, $\boldsymbol{b} \in \triangle^n$ *and* $T \in \mathcal{U}(\boldsymbol{a}, \boldsymbol{b})$. *Then,*

$$\max\{H(\boldsymbol{a}), H(\boldsymbol{b})\} \leq H(T) \leq H(\boldsymbol{a}) + H(\boldsymbol{b}).$$

Proof. See, for instance, [Cover and Thomas, 2006].

Together with $0 \leq H(\boldsymbol{a}) \leq \log m$ and $0 \leq H(\boldsymbol{b}) \leq \log n$, this provides lower and upper bounds for the entropy of a transportation plan. As noted in [Cuturi, 2013], the upper bound is tight since

$$\max_{T \in \mathcal{U}(\boldsymbol{a}, \boldsymbol{b})} H(T) = H(\boldsymbol{a}\boldsymbol{b}^\top) = H(\boldsymbol{a}) + H(\boldsymbol{b}).$$

Lemma 3 *Bounding the squared* 2*-norm of a transportation plan*

Let $\boldsymbol{a} \in \triangle^m$, $\boldsymbol{b} \in \triangle^n$ *and* $T \in \mathcal{U}(\boldsymbol{a}, \boldsymbol{b})$. *Then,*

$$\sum_{i=1}^m \sum_{j=1}^n \left(\frac{a_i}{n} + \frac{b_j}{m} - \frac{1}{mn}\right)^2 \leq \|T\|^2 \leq \min\left\{\|\boldsymbol{a}\|^2, \|\boldsymbol{b}\|^2\right\}.$$

Proof. The tightest lower bound is given by $\min_{T \in \mathcal{U}(\boldsymbol{a},\boldsymbol{b})} \|T\|^2$. An exact iterative algorithm was proposed in [Calvillo and Romero, 2016] to solve this problem. However, since we are interested in an explicit formula, we consider instead the lower bound $\min_{\substack{T\mathbf{1}_n=\boldsymbol{a} \\ T^\top\mathbf{1}_m=\boldsymbol{b}}} \|T\|^2$ (*i.e.*, we ignore the non-negativity constraint). It is known [Romero, 1990] that the minimum is achieved at $t_{i,j} = \frac{a_i}{n} + \frac{b_j}{m} - \frac{1}{mn}$, hence our lower bound. For the upper bound, we have

$$
\begin{aligned}
\|T\|^2 &= \sum_{i=1}^m \sum_{j=1}^n t_{i,j}^2 \\
&= \sum_{i=1}^m \sum_{j=1}^n \left(a_i \frac{t_{i,j}}{a_i}\right)^2 \\
&= \sum_{i=1}^m a_i^2 \sum_{j=1}^n \left(\frac{t_{i,j}}{a_i}\right)^2 \\
&\leq \sum_{i=1}^m a_i^2 \sum_{j=1}^n \left(\frac{t_{i,j}}{a_i}\right) \\
&= \|\boldsymbol{a}\|^2.
\end{aligned}
$$

We can do the same with $\boldsymbol{b} \in \triangle^n$ to obtain $\|T\|^2 \leq \|\boldsymbol{b}\|^2$, yielding the claimed result. □

Together with $0 \leq \|\boldsymbol{a}\|^2 \leq 1$ and $0 \leq \|\boldsymbol{b}\|^2 \leq 1$, this provides lower and upper bounds for the squared 2-norm of a transportation plan.

Proof of the theorem. Let $T^\star$ and $T^\star_\Omega$ be optimal solutions of (2) and (5), respectively. Then,

$$\mathrm{OT}(\boldsymbol{a},\boldsymbol{b}) + \Omega(T^\star_\Omega) = \langle T^\star, C\rangle + \Omega(T^\star_\Omega) \leq \langle T^\star_\Omega, C\rangle + \Omega(T^\star_\Omega) = \mathrm{OT}_\Omega(\boldsymbol{a},\boldsymbol{b}).$$

Likewise,

$$\mathrm{OT}_\Omega(\boldsymbol{a},\boldsymbol{b}) = \langle T^\star_\Omega, C\rangle + \Omega(T^\star_\Omega) \leq \langle T^\star, C\rangle + \Omega(T^\star) = \mathrm{OT}(\boldsymbol{a},\boldsymbol{b}) + \Omega(T^\star).$$

Combining the two, we obtain

$$\mathrm{OT}(\boldsymbol{a},\boldsymbol{b}) + \Omega(T^\star_\Omega) \leq \mathrm{OT}_\Omega(\boldsymbol{a},\boldsymbol{b}) \leq \mathrm{OT}(\boldsymbol{a},\boldsymbol{b}) + \Omega(T^\star).$$

Using $T^\star, T^\star_\Omega \in \mathcal{U}(\boldsymbol{a},\boldsymbol{b})$ together with Lemma 2 and Lemma 3 gives the claimed results.

A.5 Proof of Theorem 2

To prove the theorem, we first need the following two lemmas.

Lemma 4 *Bounding the* 1*-norm of* $\boldsymbol{\alpha}$ *and* $\boldsymbol{\beta}$ *for* $(\boldsymbol{\alpha},\boldsymbol{\beta}) \in \mathcal{P}(C)$

Let $\boldsymbol{\alpha},\boldsymbol{\beta} \in \mathcal{P}(C)$ *with extra constraints* $\boldsymbol{\alpha}^\top\mathbf{1}_m = 0$ *and* $\boldsymbol{\alpha}^\top\boldsymbol{a} + \boldsymbol{\beta}^\top\boldsymbol{b} \geq 0$*, where* $\boldsymbol{a} \in \triangle^m$ *and* $\boldsymbol{b} \in \triangle^n$*. Then,*

$$0 \leq \|\boldsymbol{\alpha}\|_1 + \|\boldsymbol{\beta}\|_1 \leq \|C\|_\infty(\nu + n)$$

where

$$\nu = \max\left\{(2 + n/m)\left\|\boldsymbol{a}^{-1}\right\|_\infty, \ \left\|\boldsymbol{b}^{-1}\right\|_\infty\right\}.$$

Proof. The proof technique is inspired by [Meshi et al., 2012, Supplementary material Lemma 1.2].

The 1-norm can be rewritten as

$$\|\boldsymbol{\alpha}\|_1 + \|\boldsymbol{\beta}\|_1 = \max_{\substack{\boldsymbol{r}\in\{-1,1\}^m \\ \boldsymbol{s}\in\{-1,1\}^n}} \quad \boldsymbol{r}^\top\boldsymbol{\alpha} + \boldsymbol{s}^\top\boldsymbol{\beta}.$$

Our goal is to upper bound the following objective

$$\max_{\boldsymbol{\alpha}\in\mathbb{R}^m,\boldsymbol{\beta}\in\mathbb{R}^n} \quad \boldsymbol{r}^\top\boldsymbol{\alpha} + \boldsymbol{s}^\top\boldsymbol{\beta} \quad \text{s.t.} \quad 0 \le \boldsymbol{\alpha}^\top\boldsymbol{a} + \boldsymbol{\beta}^\top\boldsymbol{b}, \quad \alpha_i + \beta_j \le c_{i,j}, \quad \boldsymbol{\alpha}^\top\mathbf{1}_m = 0,$$

with a constant that does not depend on $\boldsymbol{r}$ and $\boldsymbol{s}$. We call the above the dual problem. Its Lagrangian is

$$\begin{aligned} L(\boldsymbol{\alpha},\boldsymbol{\beta},\mu,\nu,T) &= \boldsymbol{r}^\top\boldsymbol{\alpha} + \boldsymbol{s}^\top\boldsymbol{\beta} + \mu\boldsymbol{\alpha}^\top\mathbf{1}_m + \nu(\boldsymbol{\alpha}^\top\boldsymbol{a} + \boldsymbol{\beta}^\top\boldsymbol{b}) + \sum_{i,j=1}^{m,n} t_{i,j}\left(c_{i,j} - \alpha_i - \beta_j\right) \\ &= (\boldsymbol{r} + \mu\mathbf{1}_m + \nu\boldsymbol{a} - T\mathbf{1}_n)^\top\boldsymbol{\alpha} + (\boldsymbol{s} + \nu\boldsymbol{b} - T^\top\mathbf{1}_m)^\top\boldsymbol{\beta} + \langle T, C\rangle \end{aligned}$$

with $\mu \in \mathbb{R}$, $\nu \ge 0$, $T \ge \mathbf{0}$. Maximizing the Lagrangian w.r.t. $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ gives the corresponding primal problem

$$\min_{T\ge 0,\ \mu\in\mathbb{R},\ \nu\ge 0} \langle T, C\rangle \quad \text{s.t.} \quad T\mathbf{1}_n = \nu\boldsymbol{a} + \boldsymbol{r} + \mu\mathbf{1}_m, \quad T^\top\mathbf{1}_m = \nu\boldsymbol{b} + \boldsymbol{s}.$$

By weak duality, any feasible primal point provides an upper bound of the dual problem. We start by choosing $\mu = \frac{1}{m}(\sum_j s_j - \sum_i r_i)$ so that $\sum_{i,j} t_{i,j}$ provides the same values w.r.t. the last two constraints. Next, we choose

$$\nu = \max\left\{\max_i \frac{2 + n/m}{a_i}, \max_j \frac{1}{b_j}\right\}$$

which ensures the non-negativity of $\nu\boldsymbol{a} + \boldsymbol{r} + \mu\mathbf{1}_m$ and $\nu\boldsymbol{b} + \boldsymbol{s}$ regardless of $\boldsymbol{r}$ and $\boldsymbol{s}$. It follows that the transportation plan T defined by

$$T = \frac{1}{(\nu\boldsymbol{b} + \boldsymbol{s})^T\mathbf{1}_n}(\nu\boldsymbol{a} + \boldsymbol{r} + \mu\mathbf{1}_m)(\nu\boldsymbol{b} + \boldsymbol{s})^\top$$

is feasible. We finally bound the objective, $\langle T, C\rangle \le \|C\|_\infty \sum_{i,j} t_{i,j} \le \|C\|_\infty (\nu + n)$. □

Lemma 5 *Bounding the* 1*-norm of* $\boldsymbol{\alpha}$ *for* $(\boldsymbol{\alpha}, \cdot) \in \mathcal{P}(C)$

Let $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathcal{P}(C)$ *with extra constraints* $\sum_{i=1}^m \alpha_i = 0$ *and* $\alpha^\top\boldsymbol{a} + \boldsymbol{\beta}^\top\boldsymbol{b} \ge 0$, *where* $\boldsymbol{a} \in \triangle^m$ *and* $\boldsymbol{b} \in \triangle^n$. *Then,*

$$0 \le \|\boldsymbol{\alpha}\|_1 \le 2\|C\|_\infty \left\|\boldsymbol{a}^{-1}\right\|_\infty.$$

Proof. Similarly as before, our goal is to upper bound

$$\max_{\boldsymbol{\alpha}\in\mathbb{R}^m,\boldsymbol{\beta}\in\mathbb{R}^n} \boldsymbol{r}^\top\boldsymbol{\alpha} \quad \text{s.t.} \quad 0 \le \boldsymbol{\alpha}^\top\boldsymbol{a} + \boldsymbol{\beta}^\top\boldsymbol{b}, \quad \alpha_i + \beta_j \le c_{i,j}, \quad \boldsymbol{\alpha}^\top\mathbf{1}_m = 0,$$

with a constant which does not depend on $\boldsymbol{r}$. The corresponding primal is

$$\min_{T\ge 0,\ \mu\in\mathbb{R},\ \nu\ge 0} \langle T, C\rangle \quad \text{s.t.} \quad T\mathbf{1}_n = \nu\boldsymbol{a} + \boldsymbol{r} + \mu\mathbf{1}_m, \quad T^\top\mathbf{1}_m = \nu\boldsymbol{b}.$$

By weak duality, any feasible primal point gives us an upper bound. We start by choosing $\mu = \frac{1}{m}\sum_i r_i$ so that $\sum_{ij} t_{i,j}$ provides the same values w.r.t. the last two constraints. Next, we choose, $\nu = \max_i \frac{2}{a_i}$, which ensures the non-negativity of $\nu\boldsymbol{a} + \boldsymbol{r} + \mu\mathbf{1}_m$ ($\nu\boldsymbol{b} \ge 0$ is also satisfied since $\nu \ge 0$) which appears in the r.h.s. of the second constraint, independently of $\boldsymbol{r}$. It follows that the transportation plan T defined by

$$T = \frac{1}{\nu\boldsymbol{b}^\top\mathbf{1}_n}(\nu\boldsymbol{a} + \boldsymbol{r} + \mu\mathbf{1}_m)(\nu\boldsymbol{b})^\top = (\nu\boldsymbol{a} + \boldsymbol{r} + \mu\mathbf{1}_m)\boldsymbol{b}^\top$$

is feasible. We finally bound the objective

$$\langle T, C\rangle \leq \|C\|_\infty \sum_{i,j} t_{i,j} \leq \nu \|C\|_\infty = 2\|C\|_\infty \left\|\boldsymbol{a}^{-1}\right\|_\infty,$$

which concludes the proof. □

Proof of the theorem. We begin by deriving the bound for the relaxed primal. Let $(\boldsymbol{\alpha}^\star, \boldsymbol{\beta}^\star)$ and $(\boldsymbol{\alpha}^\star_\Phi, \boldsymbol{\beta}^\star_\Phi)$ be optimal solutions of (3) and (17), respectively. Since $(\boldsymbol{\alpha}^\star_\Phi)^\top \boldsymbol{a} + (\boldsymbol{\beta}^\star_\Phi)^\top \boldsymbol{b} \leq (\boldsymbol{\alpha}^\star)^\top \boldsymbol{a} + (\boldsymbol{\beta}^\star)^\top \boldsymbol{b}$, we have

$$\mathrm{ROT}_\Phi(\boldsymbol{a}, \boldsymbol{b}) \leq \mathrm{OT}(\boldsymbol{a}, \boldsymbol{b}) - \frac{\gamma}{2}(\|\boldsymbol{\alpha}_\Phi\|^2 + \|\boldsymbol{\beta}_\Phi\|^2).$$

Likewise,

$$\mathrm{OT}(\boldsymbol{a}, \boldsymbol{b}) - \frac{\gamma}{2}(\|\boldsymbol{\alpha}^\star\|^2 + \|\boldsymbol{\beta}^\star\|^2) \leq \mathrm{ROT}_\Phi(\boldsymbol{a}, \boldsymbol{b}).$$

Combining the two, we get

$$\mathrm{OT}(\boldsymbol{a}, \boldsymbol{b}) - \frac{\gamma}{2}(\|\boldsymbol{\alpha}^\star\|^2 + \|\boldsymbol{\beta}^\star\|^2) \leq \mathrm{ROT}_\Phi(\boldsymbol{a}, \boldsymbol{b}) \leq \mathrm{OT}(\boldsymbol{a}, \boldsymbol{b}) - \frac{\gamma}{2}(\|\boldsymbol{\alpha}_\Phi\|^2 + \|\boldsymbol{\beta}_\Phi\|^2). \tag{18}$$

Hence we need to bound variables $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathcal{P}(C)$. Since $\|\cdot\|_2 \leq \|\cdot\|_1$, we can upper bound $\|\boldsymbol{\alpha}^\star\|_1 + \|\boldsymbol{\beta}^\star\|_1$. In addition, we can always add the additional constraint that $\boldsymbol{\alpha}^\top \boldsymbol{a} + \beta^\top \boldsymbol{b} \geq \mathbf{0}^\top \boldsymbol{a} + \mathbf{0}^\top \boldsymbol{b} = 0$ since $(\mathbf{0}, \mathbf{0})$ is dual feasible for (3). Since for any optimal pair $\boldsymbol{\alpha}^\star, \boldsymbol{\beta}^\star$, the pair $\boldsymbol{\alpha}^\star - \sigma\mathbf{1},\ \boldsymbol{\beta}^\star + \sigma\mathbf{1}$ is also feasible and optimal for any $\sigma \in \mathbb{R}$, we can also add the constraint $\boldsymbol{\alpha}^\top \mathbf{1}_m = 0$. The obtained bound will obviously hold for any optimal pair $\boldsymbol{\alpha}^\star, \boldsymbol{\beta}^\star$. Hence, we can apply Lemma 4. By the same reasoning but using the constraint $\boldsymbol{\beta}^\top \mathbf{1}_n = 0$ in place of $\boldsymbol{\alpha}^\top \mathbf{1}_m = 0$, we can obtain a similar bound. By combining these two bounds, we obtain our final bound:

$$\|\boldsymbol{\alpha}\|_1 + \|\boldsymbol{\beta}\|_1 \leq \|C\|_\infty \min\{\nu_1 + n, \nu_2 + m\}$$

where

$$\nu_1 = \max\left\{(2 + n/m)\left\|\boldsymbol{a}^{-1}\right\|_\infty, \left\|\boldsymbol{b}^{-1}\right\|_\infty\right\}$$
$$\nu_2 = \max\left\{\left\|\boldsymbol{a}^{-1}\right\|_\infty, (2 + m/n)\left\|\boldsymbol{b}^{-1}\right\|_\infty\right\}.$$

Taking the square of this bound and plugging the result in (18) gives the claimed result. Applying the same reasoning with Lemma 5 gives the claimed result for the semi-relaxed primal.

B Alternating minimization with exact block updates

General case. Let $\boldsymbol{\beta}(\boldsymbol{\alpha})$ be an optimal solution of (7) given $\boldsymbol{\alpha}$ fixed, and similarly for $\boldsymbol{\alpha}(\boldsymbol{\beta})$. From the first-order optimality conditions,

$$\nabla\delta_\Omega\left(\boldsymbol{\alpha} + \beta_j(\boldsymbol{\alpha})\mathbf{1}_m - \boldsymbol{c}_j\right)^\top \mathbf{1}_m = b_j \quad \forall j \in [n] \tag{19}$$

and similarly for $\boldsymbol{\alpha}$ given $\boldsymbol{\beta}$ fixed. Solving these equations is non-trivial in general. However, because

$$\nabla\delta_\Omega\left(\boldsymbol{\alpha} + \beta_j(\boldsymbol{\alpha})\mathbf{1}_m - \boldsymbol{c}_j\right) = b_j \nabla\mathrm{max}_{\Omega_j}(\boldsymbol{\alpha} - \boldsymbol{c}_j)$$

holds $\forall \boldsymbol{\alpha} \in \mathbb{R}^m,\ j \in [n]$, we can retrieve $\beta_j(\boldsymbol{\alpha})$ if we know how to compute $\nabla\mathrm{max}_\Omega(\boldsymbol{x})$ and the inverse map $(\nabla\delta_\Omega)^{-1}(\boldsymbol{y})$ exists. That map exists and equals $\nabla\Omega(\boldsymbol{y})$ provided that Ω is differentiable and $\boldsymbol{y} > \mathbf{0}$.

Entropic regularization. It is easy to verify that (19) is satisfied with

$$\boldsymbol{\beta}(\boldsymbol{\alpha}) = \gamma \log\left(\frac{\boldsymbol{b}}{K^\top e^{\frac{\boldsymbol{\alpha}}{\gamma} - \mathbf{1}_m}}\right) \quad \text{where} \quad K := e^{\frac{-C}{\gamma}}$$

and similarly for $\boldsymbol{\alpha}(\boldsymbol{\beta})$. These updates recover the iterates of the Sinkhorn algorithm [Cuturi, 2013].

Squared 2-norm regularization. Plugging the expression of $\nabla\delta_\Omega$ in (19), we get that $\boldsymbol{\beta}(\boldsymbol{\alpha})$ must satisfy

$$[\boldsymbol{\alpha} + \beta_j(\boldsymbol{\alpha})\mathbf{1}_m - \boldsymbol{c}_j]_+^\top \mathbf{1}_m = \gamma b_j \quad \forall j \in [n].$$

Close inspection shows that it is exactly the same optimality condition as the Euclidean projection onto the simplex $\underset{\boldsymbol{y} \in \triangle^m}{\operatorname{argmin}} \|\boldsymbol{y} - \boldsymbol{x}\|^2$ must satisfy, with $\boldsymbol{x} = \frac{\boldsymbol{\alpha} - \boldsymbol{c}_j}{\gamma b_j}$. Let $x_{[1]} \geq \cdots \geq x_{[m]}$ be the values of $\boldsymbol{x}$ in sorted order. Following [Michelot, 1986, Duchi et al., 2008], if we let

$$\rho \coloneqq \max \left\{ i \in [m] \colon x_{[i]} - \frac{1}{i} \left(\sum_{r=1}^{i} x_{[r]} - 1 \right) > 0 \right\}$$

then $\boldsymbol{y}^\star$ is *exactly* achieved at $[\boldsymbol{x} + \frac{\beta_j(\boldsymbol{\alpha})}{\gamma b_j} \mathbf{1}_m]_+$, where

$$\beta_j(\boldsymbol{\alpha}) = -\frac{\gamma b_j}{\rho} \left(\sum_{r=1}^{\rho} x_{[r]} - 1 \right).$$

The expression for $\boldsymbol{\alpha}(\boldsymbol{\beta})$ is completely symmetrical. While a projection onto the simplex is required for each coordinate, as discussed in §3.3, this can be done in expected linear time. In addition, each coordinate-wise solution can be computed in parallel.

Alternating minimization. Once we know how to compute $\boldsymbol{\beta}(\boldsymbol{\alpha})$ and $\boldsymbol{\alpha}(\boldsymbol{\beta})$, there are a number of ways we can build a proper algorithm to solve the smoothed dual. Perhaps the simplest is to alternate between $\boldsymbol{\beta} \leftarrow \boldsymbol{\beta}(\boldsymbol{\alpha})$ and $\boldsymbol{\alpha} \leftarrow \boldsymbol{\alpha}(\boldsymbol{\beta})$. For entropic regularization, this two-block coordinate descent (CD) scheme is known as the Sinkhorn algorithm and was recently popularized in the context of optimal transport by Cuturi [2013]. A disadvantage of this approach, however, is that computational effort is spent updating coordinates that may already be near-optimal. To address this issue, we can instead adopt a greedy CD scheme as recently proposed for entropic regularization by Altschuler et al. [2017].

C Additional experiments

We ran the same experiments as Figure 2 and Figure 3 on one more image pair: "Grafiti" by Jon Ander and "Rainbow Bridge National Monument Utah", by Bernard Spragg. Both images are in the public domain. The results, presented in Figure 5 and Figure 6 below, confirm the empirical findings described in §6.1 and §6.2. The images are available at `https://github.com/mblondel/smooth-ot/tree/master/data`.

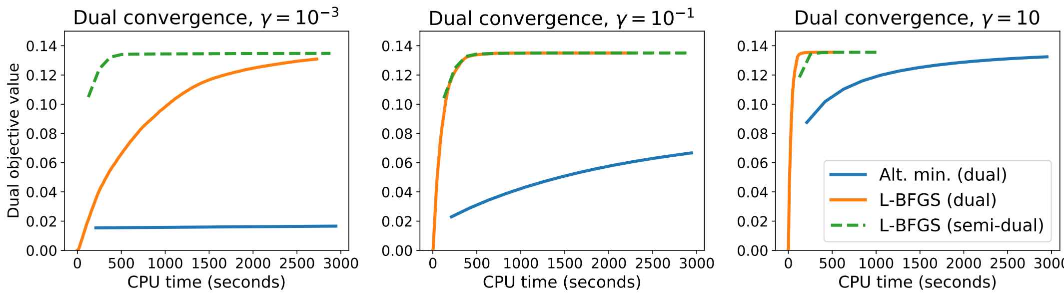

Figure 5: Same experiment as Figure 3 on one more image pair.

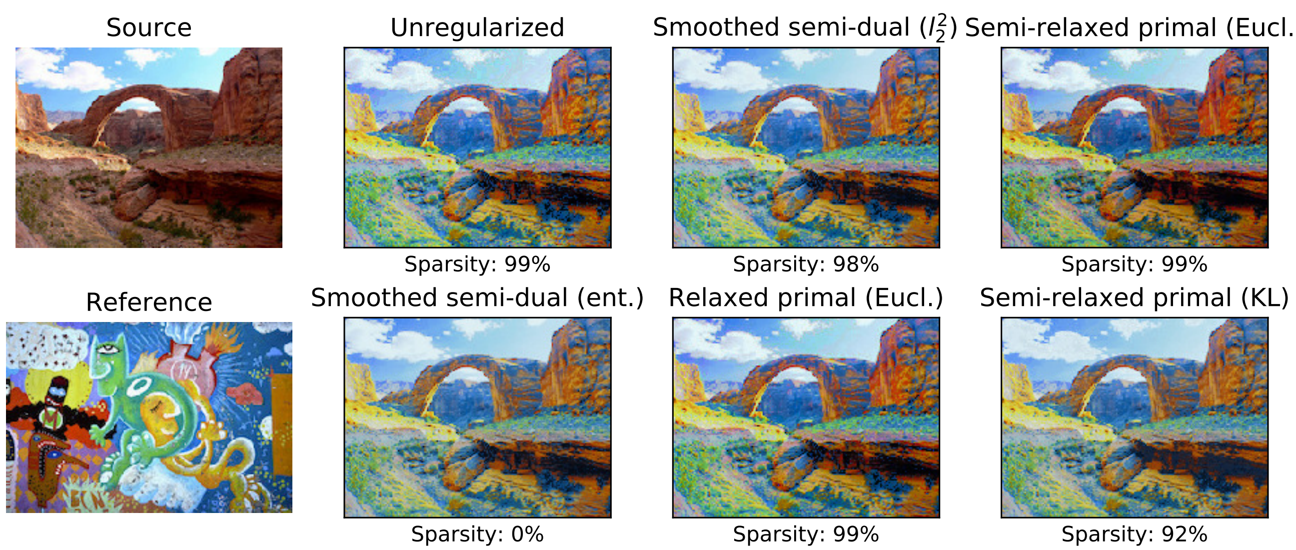

Figure 6: Same experiment as Figure 2 on one more image pair.