Appendix

A  Proofs

A.1  Derivation of the smooth relaxed dual

Recall that

\[ \text{OT}_\Omega(a,b) = \min_{T \in \mathcal{U}(a,b)} \sum_{j=1}^{n} t_j^T c_j + \Omega(t_j). \tag{16} \]

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

\[ \text{OT}_\Omega(a,b) = \min_{T \geq 0} \max_{\alpha \in \mathbb{R}^m, \beta \in \mathbb{R}^n} \sum_{j=1}^{n} t_j^T c_j + \Omega(t_j) + \alpha^T (T1_n - a) + \beta^T (T^T 1_m - b). \]

Since \( 16 \) is a convex optimization problem with only linear equality and inequality constraints, Slater's conditions reduce to feasibility \cite{BoydVandenberghe2004} \S 5.2.3] and hence strong duality holds:

\[ \text{OT}_\Omega(a,b) = \max_{\alpha \in \mathbb{R}^m, \beta \in \mathbb{R}^n} \min_{T \geq 0} \sum_{j=1}^{n} t_j^T c_j + \Omega(t_j) + \alpha^T (T1_n - a) + \beta^T (T^T 1_m - b) \]

\[ \quad = \max_{\alpha \in \mathbb{R}^m, \beta \in \mathbb{R}^n} \sum_{j=1}^{n} \min_{t_j \geq 0} t_j^T (c_j + \alpha + \beta 1_m) + \Omega(t_j) - \alpha^T a - \beta^T b \]

\[ \quad = \max_{\alpha \in \mathbb{R}^m, \beta \in \mathbb{R}^n} \alpha^T a + \beta^T b - \sum_{j=1}^{n} \max_{t_j \geq 0} t_j^T (\alpha + \beta 1_m - c_j) - \Omega(t_j). \]

Finally, plugging the expression of \( 16 \) gives the claimed result.

A.2  Derivation of the convex conjugate

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

\[ \text{OT}_\Omega^*(g,b) = \sup_{a \in \Delta^m} g^T a - \text{OT}_\Omega(a,b). \]

Following a similar argument as \cite{CuturiPeyre2016} Theorem 2.4], we have

\[ \text{OT}_\Omega^*(g,b) = \max_{T \geq 0 \atop T 1_m = b} \langle T, g1_n^T - C \rangle - \sum_{j=1}^{n} \Omega(t_j). \]

Notice that this is an easier optimization problem than \( 16 \), since there are equality constraints only in one direction. \cite{CuturiPeyre2016} showed that this optimization problem admits a closed form in the case of entropic regularization. Here, we show how to compute \( \text{OT}_\Omega^* \) for any strongly-convex regularization.

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

\[ \text{OT}_\Omega^*(g,b) = \sum_{j=1}^{n} \max_{t_j \geq 0, t_j 1_m = b_j} t_j^T (g - c_j) - \Omega(t_j) \]

\[ \quad = \sum_{j=1}^{n} b_j \max_{T_j \in \Delta^m} T_j^T (g - c_j) - \frac{1}{b_j} \Omega(b_j) \]

\[ \quad = \sum_{j=1}^{n} b_j \max_{\Omega_j} (g - c_j), \]
where we defined $\Omega_j(y) := \frac{1}{b_j} \Omega(b_jy)$ 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)**

\[
\text{ROT}_\Phi(a, b) = \max_{\alpha, \beta \in P(C)} -\frac{1}{2} \Phi^*(-2\alpha, a) - \frac{1}{2} \Phi^*(-2\beta, b)
\]

\[
\text{ROT}_\Phi(a, b) = \max_{\alpha, \beta \in P(C)} -\Phi^*(-\alpha, a) + \beta^T b
\]

\[
= \max_{\alpha \in \mathbb{R}^m} -\Phi^*(-\alpha, a) - \sum_{j=1}^n b_j \max_{i \in [m]} (\alpha_i - c_{i,j}),
\]

where $\Phi^*$ is the conjugate of $\Phi$ in the first argument.

The duals are strongly convex if $\Phi$ is smooth.

When $\Phi(x, y) = \frac{1}{2\gamma} \|x - y\|^2$, $\Phi^*(-\alpha, a) = \frac{1}{2}\|\alpha\|^2 - \alpha^T a$. Plugging that expression in the above, we get

\[
\text{ROT}_\Phi(a, b) = \max_{\alpha, \beta \in P(C)} \alpha^T a + \beta^T b - \gamma (\|\alpha\|^2 + \|\beta\|^2)
\]

and

\[
\text{ROT}_\Phi(a, b) = \max_{\alpha, \beta \in P(C)} \alpha^T a + \beta^T b - \frac{\gamma}{2}\|\alpha\|^2
\]

\[
= \max_{\alpha \in \mathbb{R}^m} \sum_{i=1}^m \alpha_i \max_{j \in [n]} (\alpha_i - c_{i,j}) - \frac{\gamma}{2}\|\alpha\|^2.
\]

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 $H(a) := -i \log a_i$ and $H(T) := -\sum_{i,j} t_{i,j} \log t_{i,j}$ be the joint entropy.

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

\[
\max\{H(a), H(b)\} \leq H(T) \leq H(a) + H(b).
\]

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

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

\[
\max_{T \in \mathcal{U}(a, b)} H(T) = H(ab^T) = H(a) + H(b).
\]

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

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

\[
\sum_{i=1}^m \sum_{j=1}^n \left( \frac{a_i}{n} + \frac{b_j}{m} - \frac{1}{mn} \right)^2 \leq \|T\|^2 \leq \min\{\|a\|^2, \|b\|^2\}.
\]
Proof. The tightest lower bound is given by \( \min_{T \in \mathcal{U}(a,b)} \|T\|^2 \). 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 \( \min_{T_{1a}=a, T_{1m}=b} \|T\|^2 \) (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

\[
\|T\|^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} t_{i,j}^2
\]

\[
= \sum_{i=1}^{m} \sum_{j=1}^{n} \left( \frac{t_{i,j}}{a_i} \right)^2
\]

\[
= \sum_{i=1}^{m} a_i^2 \sum_{j=1}^{n} \left( \frac{t_{i,j}}{a_i} \right)^2
\]

\[
\leq \sum_{i=1}^{m} a_i^2 \sum_{j=1}^{n} \left( \frac{t_{i,j}}{a_i} \right)^2
\]

\[
= \|a\|^2.
\]

We can do the same with \( b \in \Delta^n \) to obtain \( \|T\|^2 \leq \|b\|^2 \), yielding the claimed result. \( \square \)

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

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

\[
\text{OT}(a, b) + \Omega(T^*_\Omega) = \langle T^*, C \rangle + \Omega(T^*_\Omega) \leq \langle T^*_\Omega, C \rangle + \Omega(T^*_\Omega) = \text{OT}_\Omega(a, b).
\]

Likewise,

\[
\text{OT}_\Omega(a, b) = \langle T^*_\Omega, C \rangle + \Omega(T^*_\Omega) \leq \langle T^*, C \rangle + \Omega(T^*) = \text{OT}(a, b) + \Omega(T^*).
\]

Combining the two, we obtain

\[
\text{OT}(a, b) + \Omega(T^*_\Omega) \leq \text{OT}_\Omega(a, b) \leq \text{OT}(a, b) + \Omega(T^*).
\]

Using \( T^*, T^*_\Omega \in \mathcal{U}(a, b) \) 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 \( \alpha \) and \( \beta \) for \( (\alpha, \beta) \in \mathcal{P}(C) \)

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

\[
0 \leq \|\alpha\|_1 + \|\beta\|_1 \leq \|C\|_\infty (\nu + n)
\]

where

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

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

The 1-norm can be rewritten as

\[
|\alpha|_1 + |\beta|_1 = \max_{r \in (-1,1)^m} \max_{s \in (-1,1)^n} r^\top \alpha + s^\top \beta.
\]
Our goal is to upper bound the following objective

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

with a constant that does not depend on \( r \) and \( s \). We call the above the dual problem. Its Lagrangian is

\[
L(\alpha, \beta, \mu, \nu, T) = r^\top \alpha + s^\top \beta + \mu \alpha^\top 1_m + \nu(\alpha^\top a + \beta^\top b) + \sum_{i,j=1}^{m,n} t_{i,j} (c_{i,j} - \alpha_i - \beta_j)
\]

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

\[
\nu = \max \left\{ \frac{2 + n/m}{a_i}, \frac{1}{b_j} \right\}
\]

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

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

is feasible. We finally bound the objective, \( (T, C) \leq \|C\|_\infty \sum_{i,j} t_{i,j} \leq \|C\|_\infty (\nu + n) \). \( \square \)

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

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

\[
0 \leq \|\alpha\|_1 \leq 2\|C\|_\infty \|a^{-1}\|_\infty.
\]

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

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

with a constant which does not depend on \( r \). The corresponding primal is

\[
\min_{T \geq 0, \mu \in \mathbb{R}, \nu \geq 0} (T, C) \quad \text{s.t. } T1_n = \nu a + r + \mu 1_m, \\
T^\top 1_m = \nu b.
\]

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 \( \sum_{i,j} t_{i,j} \) provides the same values w.r.t. the last two constraints. Next, we choose \( \nu = \max \frac{1}{r_i} \), which ensures the non-negativity of \( \nu a + r + \mu 1_m \) \( (\nu b \geq 0 \) is also satisfied since \( \nu \geq 0 \) which appears in the r.h.s. of the second constraint, independently of \( r \). It follows that the transportation plan \( T \) defined by

\[
T = \frac{1}{\nu b^\top 1_n} (\nu a + r + \mu 1_m)(\nu b)^\top = (\nu a + r + \mu 1_m)b^\top
\]
holds
\[ \forall (\alpha^*, \beta^*) \text{ and } (\alpha^*_p, \beta^*_p) \] respectively. Since \((\alpha^*_p)^\top a + (\beta^*_p)^\top b \leq (\alpha^*)^\top a + (\beta^*)^\top b, we have
\[ \text{ROT}_\phi(a, b) \leq \text{OT}(a, b) - \frac{\gamma}{2} (\|\alpha\|^2 + \|\beta\|^2). \]

Likewise,
\[ \text{OT}(a, b) - \frac{\gamma}{2} (\|\alpha\|^2 + \|\beta\|^2) \leq \text{ROT}_\phi(a, b). \]

Combining the two, we get
\[ \text{OT}(a, b) - \frac{\gamma}{2} (\|\alpha\|^2 + \|\beta\|^2) \leq \text{ROT}_\phi(a, b) \leq \text{OT}(a, b) - \frac{\gamma}{2} (\|\alpha\|^2 + \|\beta\|^2). \tag{18} \]

Hence we need to bound variables \(\alpha, \beta \in \mathcal{P}(C).\) Since \(\|\cdot\|_2 \leq \|\cdot\|_1, we can upper bound \(\|\alpha\|_1 + \|\beta\|_1.\) In addition, we can always add the additional constraint that \(\alpha^\top a + \beta^\top b \geq 0, a + 0^\top b = 0 \) since \((0, 0)\) is dual feasible for \((9).\) Since for any optimal pair \(\alpha^*, \beta^*,\) the pair \(\alpha^* - \sigma \mathbf{1}, \beta^* + \sigma \mathbf{1}\) is also feasible and optimal for any \(\sigma \in \mathbb{R},\) we can also add the constraint \(\alpha^\top \mathbf{1}_m = 0.\) The obtained bound will obviously hold for any optimal pair \(\alpha^*, \beta^*.\) Hence, we can apply Lemma [4]. By the same reasoning but using the constraint \(\beta^\top \mathbf{1}_m = 0\) in place of \(\alpha^\top \mathbf{1}_m = 0,\) we can obtain a similar bound. By combining these two bounds, we obtain our final bound:
\[ \|\alpha\|_1 + \|\beta\|_1 \leq \|C\|_\infty \min\{\nu_1 + n, \nu_2 + m\} \]

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

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 \(\beta(\alpha)\) be an optimal solution of (7) given \(\alpha\) fixed, and similarly for \(\alpha(\beta).\) From the first-order optimality conditions,
\[ \nabla \delta_{\Omega} (\alpha + \beta_j (\alpha) \mathbf{1}_m - c_j)^\top \mathbf{1}_m = b_j \quad \forall j \in [n] \tag{19} \]
and similarly for \(\alpha\) given \(\beta\) fixed. Solving these equations is non-trivial in general. However, because
\[ \nabla \delta_{\Omega} (\alpha + \beta_j (\alpha) \mathbf{1}_m - c_j) = b_j \nabla \max_{\Omega_j} (\alpha - c_j) \]
holds \(\forall \alpha \in \mathbb{R}^m, j \in [n],\) we can retrieve \(\beta_j(\alpha)\) if we know how to compute \(\nabla \max_{\Omega}(x)\) and the inverse map \((\nabla \delta_{\Omega})^{-1}(y)\) exists. That map exists and equals \(\nabla \Omega(y)\) provided that \(\Omega\) is differentiable and \(y > 0.\)

Entropic regularization. It is easy to verify that (19) is satisfied with
\[ \beta(\alpha) = \gamma \log \left( \frac{b}{K^\top e^{\frac{b}{K}} \mathbf{1}_m} \right) \quad \text{where} \quad K := e^{\frac{c}{\gamma}} \]
and similarly for \(\alpha(\beta).\) These updates recover the iterates of the Sinkhorn algorithm \cite{cuturi}. Squared 2-norm regularization. Plugging the expression of \(\nabla \delta_{\Omega}\) in (19), we get that \(\beta(\alpha)\) must satisfy
\[ [\alpha + \beta_j (\alpha) \mathbf{1}_m - c_j]^\top \mathbf{1}_m = \gamma b_j \quad \forall j \in [n]. \]
Close inspection shows that it is exactly the same optimality condition as the Euclidean projection onto the simplex \( \text{argmin}_{y \in \Delta^m} \| y - x \|_2^2 \) must satisfy, with \( x = \frac{\alpha - c_j}{\gamma b_j} \). Let \( x_{[1]} \geq \cdots \geq x_{[m]} \) be the values of \( x \) in sorted order. Following [Michelot 1986, Duchi et al. 2008], if we let

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

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

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

The expression for \( \alpha(\beta) \) 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 \( \beta(\alpha) \) and \( \alpha(\beta) \), there are a number of ways we can build a proper algorithm to solve the smoothed dual. Perhaps the simplest is to alternate between \( \beta \leftarrow \beta(\alpha) \) and \( \alpha \leftarrow \alpha(\beta) \). 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](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.