

---

# Supplementary material for “Screening Sinkhorn Algorithm for Regularized Optimal Transport”

---

**Mokhtar Z. Alaya**  
LITIS EA4108  
University of Rouen  
mokhtarzahdi.alaya@gmail.com

**Maxime Bérar**  
LITIS EA4108  
University of Rouen  
maxime.berar@univ-rouen.fr

**Gilles Gasso**  
LITIS EA4108  
INSA, University of Rouen  
gilles.gasso@insa-rouen.fr

**Alain Rakotomamonjy**  
LITIS EA4108  
University of Rouen  
and Criteo AI Labs, Criteo Paris  
alain.rakoto@insa-rouen.fr

## 1 Proof of Lemma 1

Since the objective function  $\Psi_\kappa$  is convex with respect to  $(u, v)$ , the set of optima of problem (3) is non empty. Introducing two dual variables  $\lambda \in \mathbb{R}_+^n$  and  $\beta \in \mathbb{R}_+^m$  for each constraint, the Lagrangian of problem (3) reads as

$$\mathcal{L}(u, v, \lambda, \beta) = \frac{\varepsilon}{\kappa} \langle \lambda, \mathbf{1}_n \rangle + \varepsilon \kappa \langle \beta, \mathbf{1}_m \rangle + \mathbf{1}_n^\top B(u, v) \mathbf{1}_m - \langle \kappa u, \mu \rangle - \langle \frac{v}{\kappa}, \nu \rangle - \langle \lambda, e^u \rangle - \langle \beta, e^v \rangle$$

First order conditions then yield that the Lagrangian multipliers solutions  $\lambda^*$  and  $\beta^*$  satisfy

$$\begin{aligned} \nabla_u \mathcal{L}(u^*, v^*, \lambda^*, \beta^*) &= e^{u^*} \odot (K e^{v^*} - \lambda^*) - \kappa \mu = \mathbf{0}_n, \\ \text{and } \nabla_v \mathcal{L}(u^*, v^*, \lambda^*, \beta^*) &= e^{v^*} \odot (K^\top e^{u^*} - \beta) - \frac{\nu}{\kappa} = \mathbf{0}_m \end{aligned}$$

which leads to

$$\lambda^* = K e^{v^*} - \kappa \mu \odot e^{u^*} \text{ and } \beta^* = K^\top e^{u^*} - \nu \odot \kappa e^{v^*}$$

For all  $i = 1, \dots, n$  we have that  $e^{u_i^*} \geq \frac{\varepsilon}{\kappa}$ . Further, the condition on the dual variable  $\lambda_i^* > 0$  ensures that  $e^{u_i^*} = \frac{\varepsilon}{\kappa}$  and hence  $i \in I_{\varepsilon, \kappa}^0$ . We have that  $\lambda_i^* > 0$  is equivalent to  $e^{u_i^*} r_i(K) e^{v_j^*} > \kappa \mu_i$  which is satisfied when  $\varepsilon^2 r_i(K) > \kappa \mu_i$ . In a symmetric way we can prove the same statement for  $e^{v_j^*}$ .

## 2 Proof of Proposition 1

We prove only the first statement (7) and similarly we can prove the second one (8). For all  $i \in I_{\varepsilon, \kappa}$ , we have  $e^{u_i^{\text{sc}}} > \frac{\varepsilon}{\kappa}$  or  $e^{u_i^{\text{sc}}} = \frac{\varepsilon}{\kappa}$ . In one hand, if  $e^{u_i^{\text{sc}}} > \frac{\varepsilon}{\kappa}$  then according to the optimality conditions  $\lambda_i^{\text{sc}} = 0$ , which implies  $e^{u_i^{\text{sc}}} \sum_{j=1}^m K_{ij} e^{v_j^{\text{sc}}} = \kappa \mu_i$ . In another hand, we have

$$e^{u_i^{\text{sc}}} \min_{i,j} K_{ij} \sum_{j=1}^m e^{v_j^{\text{sc}}} \leq e^{u_i^{\text{sc}}} \sum_{j=1}^m K_{ij} e^{v_j^{\text{sc}}} = \kappa \mu_i.$$

We further observe that  $\sum_{j=1}^m e^{v_j^{\text{sc}}} = \sum_{j \in J_{\varepsilon, \kappa}} e^{v_j^{\text{sc}}} + \sum_{j \in J_{\varepsilon, \kappa}^0} e^{v_j^{\text{sc}}} \geq \varepsilon \kappa |J_{\varepsilon, \kappa}| + \varepsilon \kappa |J_{\varepsilon, \kappa}^0| = \varepsilon \kappa m$ . Then

$$\max_{i \in I_{\varepsilon, \kappa}} e^{u_i^{\text{sc}}} \leq \frac{\max_{i \in I_{\varepsilon, \kappa}} \mu_i}{m \varepsilon K_{\min}}.$$

Analogously, one can obtain for all  $j \in J_{\varepsilon, \kappa}$

$$\max_{j \in J_{\varepsilon, \kappa}} e^{v_j^{\text{sc}}} \leq \frac{\max_{j \in J_{\varepsilon, \kappa}} \nu_j}{n\varepsilon K_{\min}}. \quad (1)$$

Now, since  $K_{ij} \leq 1$ , we have

$$e^{u_i^{\text{sc}}} \sum_{j=1}^m e^{v_j^{\text{sc}}} \geq e^{u_i^{\text{sc}}} \sum_{j=1}^m K_{ij} e^{v_j^{\text{sc}}} = \kappa \mu_i.$$

Using (1), we get

$$\sum_{j=1}^m e^{v_j^{\text{sc}}} = \sum_{j \in J_{\varepsilon, \kappa}} e^{v_j^{\text{sc}}} + \sum_{j \in J_{\varepsilon, \kappa}^{\mathcal{C}}} e^{v_j^{\text{sc}}} \leq \varepsilon \kappa |J_{\varepsilon, \kappa}^{\mathcal{C}}| + \frac{\max_{j \in J_{\varepsilon, \kappa}} \nu_j}{n\varepsilon K_{\min}} |J_{\varepsilon, \kappa}|.$$

Therefore,

$$\min_{i \in I_{\varepsilon, \kappa}} e^{u_i^{\text{sc}}} \geq \frac{\varepsilon}{\kappa} \vee \frac{\kappa \min_{i \in I_{\varepsilon, \kappa}} \mu_i}{\varepsilon \kappa (m - m_b) + \frac{\max_{j \in J_{\varepsilon, \kappa}} \nu_j}{n\varepsilon K_{\min}} m_b}.$$

### 3 Proof of Proposition 2

We define the distance function  $\varrho : \mathbb{R}_+ \times \mathbb{R}_+ \mapsto [0, \infty]$  by  $\varrho(a, b) = b - a + a \log(\frac{a}{b})$ . While  $\varrho$  is not a metric, it is easy to see that  $\varrho$  is not nonnegative and satisfies  $\varrho(a, b) = 0$  iff  $a = b$ . The violations are computed through the following function:

$$d_{\varrho}(\gamma, \beta) = \sum_{i=1}^n \varrho(\gamma_i, \beta_i), \text{ for } \gamma, \beta \in \mathbb{R}_+^n.$$

Note that if  $\gamma, \beta$  are two vectors of positive entries,  $d_{\varrho}(\gamma, \beta)$  will return some measurement on how far they are from each other. The next Lemma is from [1] (see Lemma 7 herein).

**Lemma 1.** *For any  $\gamma, \beta \in \mathbb{R}_+^n$ , the following generalized Pinsker inequality holds*

$$\|\gamma - \beta\|_1 \leq \sqrt{7(\|\gamma\|_1 \wedge \|\beta\|_1) d_{\varrho}(\gamma, \beta)}.$$

The optimality conditions for  $(u^{\text{sc}}, v^{\text{sc}})$  entails

$$\mu_i^{\text{sc}} = \begin{cases} e^{u_i^{\text{sc}}} \sum_{j=1}^m K_{ij} e^{v_j^{\text{sc}}}, & \text{if } i \in I_{\varepsilon, \kappa}, \\ \frac{\varepsilon}{\kappa} \sum_{j=1}^m K_{ij} e^{v_j^{\text{sc}}}, & \text{if } i \in I_{\varepsilon, \kappa}^{\mathcal{C}}, \end{cases} = \begin{cases} \kappa \mu_i, & \text{if } i \in I_{\varepsilon, \kappa}, \\ \frac{\varepsilon}{\kappa} \sum_{j=1}^m K_{ij} e^{v_j^{\text{sc}}}, & \text{if } i \in I_{\varepsilon, \kappa}^{\mathcal{C}}, \end{cases} \quad (2)$$

and

$$\nu_j^{\text{sc}} = \begin{cases} e^{v_j^{\text{sc}}} \sum_{i=1}^n K_{ij} e^{u_i^{\text{sc}}}, & \text{if } j \in J_{\varepsilon, \kappa}, \\ \varepsilon \kappa \sum_{i=1}^n K_{ij} e^{u_i^{\text{sc}}}, & \text{if } j \in J_{\varepsilon, \kappa}^{\mathcal{C}}, \end{cases} = \begin{cases} \frac{\nu_j}{\kappa}, & \text{if } j \in J_{\varepsilon, \kappa}, \\ \varepsilon \kappa \sum_{i=1}^n K_{ij} e^{u_i^{\text{sc}}}, & \text{if } j \in J_{\varepsilon, \kappa}^{\mathcal{C}}. \end{cases} \quad (3)$$

By (2), we have

$$\begin{aligned} d_{\varrho}(\mu, \mu^{\text{sc}}) &= \sum_{i=1}^n \mu_i^{\text{sc}} - \mu_i + \mu_i \log\left(\frac{\mu_i}{\mu_i^{\text{sc}}}\right) \\ &= \sum_{i \in I_{\varepsilon, \kappa}} (\kappa - 1) \mu_i - \mu_i \log(\kappa) + \sum_{i \in I_{\varepsilon, \kappa}^{\mathcal{C}}} \frac{\varepsilon}{\kappa} \sum_{j=1}^m K_{ij} e^{v_j^{\text{sc}}} - \mu_i + \mu_i \log\left(\frac{\mu_i}{\frac{\varepsilon}{\kappa} \sum_{j=1}^m K_{ij} e^{v_j^{\text{sc}}}}\right) \\ &= \sum_{i \in I_{\varepsilon, \kappa}} (\kappa - \log(\kappa) - 1) \mu_i + \sum_{i \in I_{\varepsilon, \kappa}^{\mathcal{C}}} \frac{\varepsilon}{\kappa} \sum_{j=1}^m K_{ij} e^{v_j^{\text{sc}}} - \mu_i + \mu_i \log\left(\frac{\mu_i}{\frac{\varepsilon}{\kappa} \sum_{j=1}^m K_{ij} e^{v_j^{\text{sc}}}}\right). \end{aligned}$$

Now by (8), we have in one hand

$$\begin{aligned}
\sum_{i \in I_{\varepsilon, \kappa}^0} \frac{\varepsilon}{\kappa} \sum_{j=1}^m K_{ij} e^{v_j^{\text{sc}}} &= \sum_{i \in I_{\varepsilon, \kappa}^0} \frac{\varepsilon}{\kappa} \left( \sum_{j \in J_{\varepsilon, \kappa}} K_{ij} e^{v_j^{\text{sc}}} + \varepsilon \kappa \sum_{j \in J_{\varepsilon, \kappa}^0} K_{ij} \right) \\
&\leq \sum_{i \in I_{\varepsilon, \kappa}^0} \frac{\varepsilon}{\kappa} \left( m_b \max_{i,j} K_{ij} \frac{\max_{j \in J_{\varepsilon, \kappa}} \nu_j}{n \varepsilon K_{\min}} + (m - m_b) \varepsilon \kappa \max_{i,j} K_{ij} \right) \\
&\leq (n - n_b) \left( \frac{m_b \max_j \nu_j}{n \kappa K_{\min}} + (m - m_b) \varepsilon^2 \right).
\end{aligned}$$

On the other hand, we get

$$\begin{aligned}
\frac{\varepsilon}{\kappa} \sum_{j=1}^m K_{ij} e^{v_j^{\text{sc}}} &= \frac{\varepsilon}{\kappa} \left( \sum_{j \in J_{\varepsilon, \kappa}} K_{ij} e^{v_j^{\text{sc}}} + \varepsilon \kappa \sum_{j \in J_{\varepsilon, \kappa}^0} K_{ij} \right) \\
&\geq m_b K_{\min} \frac{m \varepsilon^2 K_{\min} \min_{j \in J_{\varepsilon, \kappa}} \nu_j}{\kappa((n - n_b) m \varepsilon^2 K_{\min} + m \varepsilon^2 K_{\min} + n_b \kappa \max_{i \in I_{\varepsilon, \kappa}} \mu_i)} \\
&\quad + \varepsilon^2 (m - m_b) K_{\min} \\
&\geq \frac{m m_b \varepsilon^2 (K_{\min})^2 \min_{j \in J_{\varepsilon, \kappa}} \nu_j}{\kappa((n - n_b) m \varepsilon^2 K_{\min} + m \varepsilon^2 K_{\min} + n_b \kappa \max_{i \in I_{\varepsilon, \kappa}} \mu_i)} \\
&\quad + \varepsilon^2 (m - m_b) K_{\min} \\
&\geq \frac{m m_b \varepsilon^2 K_{\min}^2 \min_{j \in J_{\varepsilon, \kappa}} \nu_j}{\kappa((n - n_b) m \varepsilon^2 K_{\min} + m \varepsilon^2 K_{\min} + n_b \kappa \max_{i \in I_{\varepsilon, \kappa}} \mu_i)}.
\end{aligned}$$

Then

$$\begin{aligned}
\frac{1}{\frac{\varepsilon}{\kappa} \sum_{j=1}^m K_{ij} e^{v_j^{\text{sc}}}} &\leq \frac{\kappa((n - n_b) m \varepsilon^2 K_{\min} + m \varepsilon^2 K_{\min} + n_b \kappa \max_{i \in I_{\varepsilon, \kappa}} \mu_i)}{m m_b \varepsilon^2 K_{\min}^2 \min_{j \in J_{\varepsilon, \kappa}} \nu_j} \\
&\leq \frac{\kappa(n - n_b + 1)}{m_b K_{\min} \min_{j \in J_{\varepsilon, \kappa}} \nu_j} + \frac{n_b \kappa^2 \max_{i \in I_{\varepsilon, \kappa}} \mu_i}{m m_b \varepsilon^2 K_{\min}^2 \min_{j \in J_{\varepsilon, \kappa}} \nu_j}.
\end{aligned}$$

It entails

$$\begin{aligned}
\sum_{i \in I_{\varepsilon, \kappa}^0} \frac{\varepsilon}{\kappa} \sum_{j=1}^m K_{ij} e^{v_j^{\text{sc}}} - \mu_i + \mu_i \log \left( \frac{\mu_i}{\frac{\varepsilon}{\kappa} \sum_{j=1}^m K_{ij} e^{v_j^{\text{sc}}}} \right) \\
\leq (n - n_b) \left( \frac{m_b}{n \kappa K_{\min}} + (m - m_b) \varepsilon^2 - \min_i \mu_i \right. \\
\left. + \max_i \mu_i \log \left( \frac{\kappa(n - n_b + 1) \max_i \mu_i}{m_b K_{\min} \min_{j \in J_{\varepsilon, \kappa}} \nu_j} + \frac{n_b \kappa^2 (\max_i \mu_i)^2}{m m_b \varepsilon^2 K_{\min}^2 \min_{j \in J_{\varepsilon, \kappa}} \nu_j} \right) \right).
\end{aligned}$$

Therefore

$$\begin{aligned}
d_\varrho(\mu, \mu^{\text{sc}}) &\leq n_b c_\kappa \max_i \mu_i + (n - n_b) \left( \frac{m_b \max_j \nu_j}{n \kappa K_{\min}} + (m - m_b) \varepsilon^2 - \min_i \mu_i \right. \\
&\quad \left. + \max_i \mu_i \log \left( \frac{\kappa(n - n_b + 1) \max_i \mu_i}{m_b K_{\min} \min_{j \in J_{\varepsilon, \kappa}} \nu_j} + \frac{n_b \kappa^2 (\max_i \mu_i)^2}{m m_b \varepsilon^2 K_{\min}^2 \min_{j \in J_{\varepsilon, \kappa}} \nu_j} \right) \right).
\end{aligned}$$

Finally, by Lemma 1 we obtain

$$\begin{aligned}
\|\mu - \mu^{\text{sc}}\|_1^2 &\leq n_b c_\kappa \max_i \mu_i + 7(n - n_b) \left( \frac{m_b \max_j \nu_j}{n \kappa K_{\min}} + (m - m_b) \varepsilon^2 - \min_i \mu_i \right. \\
&\quad \left. + \max_i \mu_i \log \left( \frac{\kappa(n - n_b + 1) \max_i \mu_i}{m_b K_{\min} \min_{j \in J_{\varepsilon, \kappa}} \nu_j} + \frac{n_b \kappa^2 (\max_i \mu_i)^2}{m m_b \varepsilon^2 K_{\min}^2 \min_{j \in J_{\varepsilon, \kappa}} \nu_j} \right) \right).
\end{aligned}$$

Following the same lines as above, we also have

$$\begin{aligned} \|\nu - \nu^{\text{sc}}\|_1^2 &\leq m_b c_{\frac{1}{\kappa}} \max_i \mu_i + 7(m - m_b) \left( \frac{n_b \kappa \max_i \mu_i}{m K_{\min}} + (n - n_b) \varepsilon^2 - \min_j \nu_j \right. \\ &\quad \left. + \max_j \nu_j \log \left( \frac{(m - m_b + 1) \max_j \nu_j}{n_b \kappa K_{\min} \min_{i \in I_{\varepsilon, \kappa}} \mu_i} + \frac{m_b (\max_j \nu_j)^2}{n n_b \varepsilon^2 \kappa^2 K_{\min}^2 \min_{i \in I_{\varepsilon, \kappa}} \mu_i} \right) \right). \end{aligned}$$

To get the closed forms (9) and (10), we used the following facts:

**Remark 1.** We have  $\log(1/K_{\min}^T) = r\|C\|_{\infty}/\eta$ , for every  $r \in \mathbb{N}$ . Using (5), we further derive:  $\varepsilon = \mathcal{O}((mnK_{\min}^2)^{-1/4})$ ,  $\kappa = \mathcal{O}(\sqrt{m/(nc_{\mu\nu}K_{\min})})$ ,  $\kappa^{-1} = \mathcal{O}(\sqrt{n/(mK_{\min}c_{\mu\nu})})$ ,  $(\kappa/\varepsilon)^2 = \mathcal{O}(m^{3/2}/\sqrt{nK_{\min}}(c_{\mu\nu})^{3/2})$ , and  $(\varepsilon\kappa)^{-2} = \mathcal{O}(n^{3/2}/\sqrt{mK_{\min}}c_{\mu\nu}^{3/2})$ .

## 4 Proof of Proposition 3

We first define  $\tilde{K}$  a rearrangement of  $K$  with respect to the active sets  $I_{\varepsilon, \kappa}$  and  $J_{\varepsilon, \kappa}$  as follows:

$$\tilde{K} = \begin{bmatrix} K_{(I_{\varepsilon, \kappa}, J_{\varepsilon, \kappa})} & K_{(I_{\varepsilon, \kappa}, J_{\varepsilon, \kappa}^c)} \\ K_{(I_{\varepsilon, \kappa}^c, J_{\varepsilon, \kappa})} & K_{(I_{\varepsilon, \kappa}^c, J_{\varepsilon, \kappa}^c)} \end{bmatrix}.$$

Setting  $\dot{\mu} = (\mu_{I_{\varepsilon, \kappa}}^\top, \mu_{I_{\varepsilon, \kappa}^c}^\top)^\top$ ,  $\dot{\nu} = (\nu_{J_{\varepsilon, \kappa}}^\top, \nu_{J_{\varepsilon, \kappa}^c}^\top)^\top$  and for each vectors  $u \in \mathbb{R}^n$  and  $v \in \mathbb{R}^m$  we set  $\dot{u} = (u_{I_{\varepsilon, \kappa}}^\top, u_{I_{\varepsilon, \kappa}^c}^\top)^\top$  and  $\dot{v} = (v_{J_{\varepsilon, \kappa}}^\top, v_{J_{\varepsilon, \kappa}^c}^\top)^\top$ . We then have

$$\Psi_{\varepsilon, \kappa}(u, v) = \mathbf{1}_n^\top \tilde{B}(\dot{u}, \dot{v}) \mathbf{1}_m - \kappa \dot{\mu}^\top \dot{u} - \kappa^{-1} \dot{\nu}^\top \dot{v},$$

and

$$\Psi(u, v) = \mathbf{1}_n^\top \tilde{B}(\dot{u}, \dot{v}) \mathbf{1}_m - \dot{\mu}^\top \dot{u} - \dot{\nu}^\top \dot{v},$$

where

$$\tilde{B}(\dot{u}, \dot{v}) = \Delta(e^{\dot{u}}) \tilde{K} \Delta(e^{\dot{v}}).$$

Let us consider the convex function

$$(\hat{u}, \hat{v}) \mapsto \langle \mathbf{1}_n, \tilde{B}(\hat{u}, \hat{v}) \mathbf{1}_m \rangle - \langle \kappa \hat{u}, \tilde{B}(\hat{u}^{\text{sc}}, \hat{v}^{\text{sc}}) \mathbf{1}_m \rangle - \langle \kappa^{-1} \hat{v}, \tilde{B}(\hat{u}^{\text{sc}}, \hat{v}^{\text{sc}})^\top \mathbf{1}_n \rangle.$$

Gradient inequality of any convex function  $g$  at point  $x_o$  reads as  $g(x_o) \geq g(x) + \langle \nabla g(x), x_o - x \rangle$ , for all  $x \in \text{dom}(g)$ . Applying the latter fact to the above function at point  $(u^*, v^*)$  we obtain

$$\begin{aligned} &\langle \mathbf{1}_n, \tilde{B}(\hat{u}^{\text{sc}}, \hat{v}^{\text{sc}}) \mathbf{1}_m \rangle - \langle \kappa \hat{u}^{\text{sc}}, \tilde{B}(\hat{u}^{\text{sc}}, \hat{v}^{\text{sc}}) \mathbf{1}_m \rangle - \langle \kappa^{-1} \hat{v}^{\text{sc}}, \tilde{B}(\hat{u}^{\text{sc}}, \hat{v}^{\text{sc}})^\top \mathbf{1}_n \rangle \\ &\quad - (\langle \mathbf{1}_n, \tilde{B}(\hat{u}^*, \hat{v}^*) \mathbf{1}_m \rangle - \langle \kappa \hat{u}^*, \tilde{B}(\hat{u}^{\text{sc}}, \hat{v}^{\text{sc}}) \mathbf{1}_m \rangle - \langle \kappa^{-1} \hat{v}^*, \tilde{B}(\hat{u}^{\text{sc}}, \hat{v}^{\text{sc}})^\top \mathbf{1}_n \rangle) \\ &\quad \leq \langle \hat{u}^{\text{sc}} - \hat{u}^*, (1 - \kappa) \tilde{B}(\hat{u}^{\text{sc}}, \hat{v}^{\text{sc}}) \mathbf{1}_m \rangle + \langle \hat{v}^{\text{sc}} - \hat{v}^*, (1 - \kappa^{-1}) \tilde{B}(\hat{u}^{\text{sc}}, \hat{v}^{\text{sc}})^\top \mathbf{1}_n \rangle. \end{aligned}$$

Moreover,

$$\begin{aligned} \Psi_{\varepsilon, \kappa}(u^{\text{sc}}, v^{\text{sc}}) - \Psi(u^*, v^*) &= \langle \mathbf{1}_n, \tilde{B}(\hat{u}^{\text{sc}}, \hat{v}^{\text{sc}}) \mathbf{1}_m \rangle - \langle \kappa \hat{u}^{\text{sc}}, \tilde{B}(\hat{u}^{\text{sc}}, \hat{v}^{\text{sc}}) \mathbf{1}_m \rangle - \langle \kappa^{-1} \hat{v}^{\text{sc}}, \tilde{B}(\hat{u}^{\text{sc}}, \hat{v}^{\text{sc}})^\top \mathbf{1}_n \rangle \\ &\quad - (\langle \mathbf{1}_n, \tilde{B}(\hat{u}^*, \hat{v}^*) \mathbf{1}_m \rangle - \langle \hat{u}^*, \tilde{B}(\hat{u}^{\text{sc}}, \hat{v}^{\text{sc}}) \mathbf{1}_m \rangle - \langle \hat{v}^*, \tilde{B}(\hat{u}^{\text{sc}}, \hat{v}^{\text{sc}})^\top \mathbf{1}_n \rangle) \\ &\quad + \langle \kappa \hat{u}^{\text{sc}} - \hat{u}^*, \tilde{B}(\hat{u}^{\text{sc}}, \hat{v}^{\text{sc}}) \mathbf{1}_m - \dot{\mu} \rangle + \langle \kappa^{-1} \hat{v}^{\text{sc}} - \hat{v}^*, \tilde{B}(\hat{u}^{\text{sc}}, \hat{v}^{\text{sc}})^\top \mathbf{1}_n - \dot{\nu} \rangle. \end{aligned}$$

Hence,

$$\begin{aligned} \Psi_{\varepsilon, \kappa}(u^{\text{sc}}, v^{\text{sc}}) - \Psi(u^*, v^*) &+ (\langle \mathbf{1}_n, \tilde{B}(\hat{u}^*, \hat{v}^*) \mathbf{1}_m \rangle \\ &\quad - \langle \hat{u}^*, \tilde{B}(\hat{u}^{\text{sc}}, \hat{v}^{\text{sc}}) \mathbf{1}_m \rangle - \langle \hat{v}^*, \tilde{B}(\hat{u}^{\text{sc}}, \hat{v}^{\text{sc}})^\top \mathbf{1}_n \rangle) \\ &\quad - \langle \kappa \hat{u}^{\text{sc}} - \hat{u}^*, \tilde{B}(\hat{u}^{\text{sc}}, \hat{v}^{\text{sc}}) \mathbf{1}_m - \dot{\mu} \rangle - \langle \kappa^{-1} \hat{v}^{\text{sc}} - \hat{v}^*, \tilde{B}(\hat{u}^{\text{sc}}, \hat{v}^{\text{sc}})^\top \mathbf{1}_n - \dot{\nu} \rangle \\ &\leq \langle \hat{u}^{\text{sc}} - \hat{u}^*, (1 - \kappa) \tilde{B}(\hat{u}^{\text{sc}}, \hat{v}^{\text{sc}}) \mathbf{1}_m \rangle + \langle \hat{v}^{\text{sc}} - \hat{v}^*, (1 - \kappa^{-1}) \tilde{B}(\hat{u}^{\text{sc}}, \hat{v}^{\text{sc}})^\top \mathbf{1}_n \rangle \\ &\quad + (\langle \mathbf{1}_n, \tilde{B}(\hat{u}^*, \hat{v}^*) \mathbf{1}_m \rangle - \langle \kappa \hat{u}^*, \tilde{B}(\hat{u}^{\text{sc}}, \hat{v}^{\text{sc}}) \mathbf{1}_m \rangle - \langle \kappa^{-1} \hat{v}^*, \tilde{B}(\hat{u}^{\text{sc}}, \hat{v}^{\text{sc}})^\top \mathbf{1}_n \rangle). \end{aligned}$$

Then,

$$\begin{aligned}\Psi_{\varepsilon, \kappa}(u^{\text{sc}}, v^{\text{sc}}) - \Psi(u^*, v^*) &\leq \langle \dot{u}^{\text{sc}} - \dot{u}^*, (1 - \kappa)\tilde{B}(\dot{u}^{\text{sc}}, \dot{v}^{\text{sc}})\mathbf{1}_m \rangle + \langle \dot{v}^{\text{sc}} - \dot{v}^*, (1 - \kappa^{-1})\tilde{B}(\dot{u}^{\text{sc}}, \dot{v}^{\text{sc}})^\top \mathbf{1}_n \rangle \\ &\quad + (\langle \mathbf{1}_n, \tilde{B}(\dot{u}^*, \dot{v}^*)\mathbf{1}_m \rangle - \langle \kappa \dot{u}^*, \tilde{B}(\dot{u}^{\text{sc}}, \dot{v}^{\text{sc}})\mathbf{1}_m \rangle - \langle \kappa^{-1} \dot{v}^*, \tilde{B}(\dot{u}^{\text{sc}}, \dot{v}^{\text{sc}})^\top \mathbf{1}_n \rangle) \\ &\quad + \langle \kappa \dot{u}^{\text{sc}} - \dot{u}^*, \tilde{B}(\dot{u}^{\text{sc}}, \dot{v}^{\text{sc}})\mathbf{1}_m - \dot{\mu} \rangle + \langle \kappa^{-1} \dot{v}^{\text{sc}} - \dot{v}^*, \tilde{B}(\dot{u}^{\text{sc}}, \dot{v}^{\text{sc}})^\top \mathbf{1}_n - \dot{\nu} \rangle \\ &\quad - (\langle \mathbf{1}_n, \tilde{B}(\dot{u}^*, \dot{v}^*)\mathbf{1}_m \rangle - \langle \dot{u}^*, \tilde{B}(\dot{u}^{\text{sc}}, \dot{v}^{\text{sc}})\mathbf{1}_m \rangle - \langle \dot{v}^*, \tilde{B}(\dot{u}^{\text{sc}}, \dot{v}^{\text{sc}})^\top \mathbf{1}_n \rangle),\end{aligned}$$

which yields

$$\begin{aligned}\Psi_{\varepsilon, \kappa}(u^{\text{sc}}, v^{\text{sc}}) - \Psi(u^*, v^*) &\leq \langle \kappa \dot{u}^{\text{sc}} - \dot{u}^*, \tilde{B}(\dot{u}^{\text{sc}}, \dot{v}^{\text{sc}})\mathbf{1}_m - \dot{\mu} \rangle + \langle \kappa^{-1} \dot{v}^{\text{sc}} - \dot{v}^*, \tilde{B}(\dot{u}^{\text{sc}}, \dot{v}^{\text{sc}})^\top \mathbf{1}_n - \dot{\nu} \rangle \\ &\quad + (1 - \kappa)\langle \dot{u}^{\text{sc}}, \tilde{B}(\dot{u}^{\text{sc}}, \dot{v}^{\text{sc}})\mathbf{1}_m \rangle + (1 - \kappa^{-1})\langle \dot{v}^{\text{sc}}, \tilde{B}(\dot{u}^{\text{sc}}, \dot{v}^{\text{sc}})^\top \mathbf{1}_n \rangle.\end{aligned}$$

Applying Holder's inequality gives

$$\begin{aligned}\Psi_{\varepsilon, \kappa}(u^{\text{sc}}, v^{\text{sc}}) - \Psi(u^*, v^*) &\leq \|\kappa \dot{u}^{\text{sc}} - \dot{u}^*\|_\infty \|\tilde{B}(\dot{u}^{\text{sc}}, \dot{v}^{\text{sc}})\mathbf{1}_m - \dot{\mu}\|_1 + \|\kappa^{-1} \dot{v}^{\text{sc}} - \dot{v}^*\|_\infty \|\tilde{B}(\dot{u}^{\text{sc}}, \dot{v}^{\text{sc}})^\top \mathbf{1}_n - \dot{\nu}\|_1 \\ &\quad + |1 - \kappa| \langle \dot{u}^{\text{sc}}, \tilde{B}(\dot{u}^{\text{sc}}, \dot{v}^{\text{sc}})\mathbf{1}_m \rangle + |1 - \kappa^{-1}| \langle \dot{v}^{\text{sc}}, \tilde{B}(\dot{u}^{\text{sc}}, \dot{v}^{\text{sc}})^\top \mathbf{1}_n \rangle \\ &\leq (\|\dot{u}^{\text{sc}} - \dot{u}^*\|_\infty + |1 - \kappa| \|\dot{u}^{\text{sc}}\|_\infty) \|\tilde{B}(\dot{u}^{\text{sc}}, \dot{v}^{\text{sc}})\mathbf{1}_m - \dot{\mu}\|_1 \\ &\quad + (\|\dot{v}^{\text{sc}} - \dot{v}^*\|_\infty + |1 - \kappa^{-1}| \|\dot{v}^{\text{sc}}\|_\infty) \|\tilde{B}(\dot{u}^{\text{sc}}, \dot{v}^{\text{sc}})^\top \mathbf{1}_n - \dot{\nu}\|_1 \\ &\quad + |1 - \kappa| \langle \dot{u}^{\text{sc}}, \tilde{B}(\dot{u}^{\text{sc}}, \dot{v}^{\text{sc}})\mathbf{1}_m \rangle + |1 - \kappa^{-1}| \langle \dot{v}^{\text{sc}}, \tilde{B}(\dot{u}^{\text{sc}}, \dot{v}^{\text{sc}})^\top \mathbf{1}_n \rangle\end{aligned}$$

where, in the last inequality, we use the facts that  $\|\kappa \dot{u}^{\text{sc}} - \dot{u}^*\|_\infty \leq \|\dot{u}^{\text{sc}} - \dot{u}^*\|_\infty + |1 - \kappa| \|\dot{u}^{\text{sc}}\|_\infty$  and  $\|\kappa^{-1} \dot{v}^{\text{sc}} - \dot{v}^*\|_\infty \leq \|\dot{v}^{\text{sc}} - \dot{v}^*\|_\infty + |1 - \kappa^{-1}| \|\dot{v}^{\text{sc}}\|_\infty$ . Moreover, note that

$$\begin{cases} \|\dot{u}^{\text{sc}} - \dot{u}^*\|_\infty = \|u^{\text{sc}} - u^*\|_\infty, \\ \|\dot{v}^{\text{sc}} - \dot{v}^*\|_\infty = \|v^{\text{sc}} - v^*\|_\infty, \end{cases} \quad \text{and} \quad \begin{cases} \|\tilde{B}(\dot{u}^{\text{sc}}, \dot{v}^{\text{sc}})\mathbf{1}_m - \dot{\mu}\|_1 = \|B(u^{\text{sc}}, v^{\text{sc}})\mathbf{1}_m - \mu\|_1 = \|\mu^{\text{sc}} - \mu\|_1, \\ \|\tilde{B}(\dot{u}^{\text{sc}}, \dot{v}^{\text{sc}})^\top \mathbf{1}_n - \dot{\nu}\|_1 = \|B(u^{\text{sc}}, v^{\text{sc}})^\top \mathbf{1}_n - \nu\|_1 = \|\nu^{\text{sc}} - \nu\|_1. \end{cases}$$

Then

$$\begin{aligned}\Psi_{\varepsilon, \kappa}(u^{\text{sc}}, v^{\text{sc}}) - \Psi(u^*, v^*) &\leq (\|u^{\text{sc}} - u^*\|_\infty + |1 - \kappa| \|u^{\text{sc}}\|_\infty) \|\mu^{\text{sc}} - \mu\|_1 \\ &\quad + (\|v^{\text{sc}} - v^*\|_\infty + |1 - \kappa^{-1}| \|v^{\text{sc}}\|_\infty) \|\nu^{\text{sc}} - \nu\|_1 \\ &\quad + |1 - \kappa| \langle u^{\text{sc}}, \mu^{\text{sc}} \rangle + |1 - \kappa^{-1}| \langle v^{\text{sc}}, \nu^{\text{sc}} \rangle \\ &\leq (\|u^{\text{sc}} - u^*\|_\infty + |1 - \kappa| \|u^{\text{sc}}\|_\infty) \|\mu^{\text{sc}} - \mu\|_1 \\ &\quad + (\|v^{\text{sc}} - v^*\|_\infty + |1 - \kappa^{-1}| \|v^{\text{sc}}\|_\infty) \|\nu^{\text{sc}} - \nu\|_1 \\ &\quad + |1 - \kappa| \|u^{\text{sc}}\|_\infty \|\mu^{\text{sc}}\|_1 + |1 - \kappa^{-1}| \|v^{\text{sc}}\|_\infty \|\nu^{\text{sc}}\|_1.\end{aligned} \tag{4}$$

Next, we bound the two terms  $\|u^{\text{sc}} - u^*\|_\infty$  and  $\|v^{\text{sc}} - v^*\|_\infty$ . If  $r \in I_{\varepsilon, \kappa}^{\mathbf{C}}$ , then we have

$$\begin{aligned}|(u^{\text{sc}})_r - u_r^*| &= \left| \log \left( \frac{\sum_{j=1}^m K_{rj} e^{v_j^*}}{\sum_{j=1}^m \frac{\kappa \mu_r}{m\varepsilon}} \right) \right| \\ &\stackrel{(*)}{\leq} \left| \log \left( \max_{1 \leq i \leq m} \frac{K_{rj} e^{v_j^*}}{\frac{\kappa \mu_r}{m\varepsilon}} \right) \right| \\ &\leq \left| \max_{1 \leq j \leq m} (v_j^* - \log(\frac{\kappa \mu_r}{m\varepsilon})) \right| \\ &\leq \|v^* - \log(\frac{\kappa \mu_r}{m\varepsilon})\|_\infty \\ &\leq \|v^* - v^{\text{sc}}\|_\infty + \log(\frac{m\varepsilon^2}{c_{\mu\nu}}).\end{aligned}$$

where the inequality  $(*)$  comes from the fact that  $\frac{\sum_{j=1}^n a_j}{\sum_{j=1}^n b_j} \leq \max_{1 \leq j \leq n} \frac{a_j}{b_j}, \forall a_j, b_j > 0$ . Now, if  $r \in I_{\varepsilon, \kappa}$ , we get

$$|u_r^{\text{sc}} - u_r^*| = \left| \log \left( \frac{\kappa \sum_{j=1}^m K_{rj} e^{v_j^*}}{\sum_{j=1}^m K_{rj} e^{(v^{\text{sc}})_j}} \right) \right| \leq \left| \log \left( \frac{\sum_{j=1}^m K_{rj} e^{v_j^*}}{\sum_{j=1}^m K_{rj} e^{(v^{\text{sc}})_j}} \right) \right| \stackrel{(*)}{\leq} \|v^{\text{sc}} - v^*\|_\infty.$$

If  $s \in J_{\varepsilon, \kappa}^{\mathcal{C}}$  then

$$\begin{aligned}
|v_s^{\text{sc}} - v_s^*| &= \left| \log(\varepsilon \kappa) - \log\left(\frac{\nu_s}{\sum_{i=1}^n K_{is} e^{u_i^*}}\right) \right| \\
&\leq \left| \log\left(\max_{1 \leq i \leq n} \frac{K_{is} e^{u_i^*}}{\frac{\nu_s}{n \kappa \varepsilon}}\right) \right| \\
&\stackrel{(*)}{\leq} \left| \max_{1 \leq i \leq n} (u_i^* - \log(\frac{\nu_s}{n \kappa \varepsilon})) \right| \\
&\leq \|u^* - \log(\frac{\nu_s}{n \kappa \varepsilon})\|_{\infty} \\
&\leq \|u^* - u^{\text{sc}}\|_{\infty} + \log(\frac{n \varepsilon^2}{c_{\mu\nu}}).
\end{aligned}$$

If  $s \in J_{\varepsilon, \kappa}$  then

$$|v_s^{\text{sc}} - v_s^*| = \left| \log\left(\frac{\kappa \sum_{i=1}^m K_{ri} e^{u_i^*}}{\sum_{i=1}^m K_{ri} e^{(u^{\text{sc}})_i}}\right) \right| \leq \left| \log\left(\frac{\kappa \sum_{i=1}^m K_{ri} e^{v_j^*}}{\sum_{i=1}^m K_{ri} e^{(u^{\text{sc}})_i}}\right) \right| \stackrel{(*)}{\leq} \|u^{\text{sc}} - u^*\|_{\infty}.$$

Therefore, we obtain the following bound:

$$\begin{aligned}
\max\{\|u^* - u^{\text{sc}}\|_{\infty}, \|v^* - v^{\text{sc}}\|_{\infty}\} &\leq \max\left\{\|u^*\|_{\infty} + \|u^{\text{sc}}\|_{\infty} + \log(\frac{n \varepsilon^2}{c_{\mu\nu}}), \|v^*\|_{\infty} + \|v^{\text{sc}}\|_{\infty} + \log(\frac{m \varepsilon^2}{c_{\mu\nu}})\right\} \\
&\leq 2\left(\|u^*\|_{\infty} + \|v^*\|_{\infty} + \|u^{\text{sc}}\|_{\infty} + \|v^{\text{sc}}\|_{\infty} + \log\left(\frac{(n \vee m) \varepsilon^2}{c_{\mu\nu}}\right)\right). \tag{5}
\end{aligned}$$

Now, Lemma 3.2 in [4] provides an upper bound for the  $\ell_{\infty}$  of the optimal solution pair  $(u^*, v^*)$  of problem (2) as follows:  $\|u^*\|_{\infty} \leq A$  and  $\|v^*\|_{\infty} \leq A$ , where

$$A = \frac{\|C\|_{\infty}}{\eta} + \log\left(\frac{n \vee m}{c_{\mu\nu}^2}\right). \tag{6}$$

Plugging (5) and (6) in (4), we obtain

$$\begin{aligned}
\Psi_{\varepsilon, \kappa}(u^{\text{sc}}, v^{\text{sc}}) - \Psi(u^*, v^*) &\leq 2\left(A + \|u^{\text{sc}}\|_{\infty} + \|v^{\text{sc}}\|_{\infty} + \log\left(\frac{(n \vee m) \varepsilon^2}{c_{\mu\nu}}\right)\right)(\|\mu^{\text{sc}} - \mu\|_1 + \|\nu^{\text{sc}} - \nu\|_1) \\
&\quad + |1 - \kappa|(\|u^{\text{sc}}\|_{\infty} \|\mu^{\text{sc}}\|_1 + \|\mu^{\text{sc}} - \mu\|_1) \\
&\quad + |1 - \kappa^{-1}|(\|v^{\text{sc}}\|_{\infty} \|\nu^{\text{sc}}\|_1 + \|\nu^{\text{sc}} - \nu\|_1). \tag{7}
\end{aligned}$$

By Proposition 1, we have

$$\|u^{\text{sc}}\|_{\infty} \leq \log\left(\frac{\varepsilon}{\kappa} \vee \frac{1}{m \varepsilon K_{\min}}\right) \text{ and } \|v^{\text{sc}}\|_{\infty} \leq \log\left(\varepsilon \kappa \vee \frac{1}{n \varepsilon K_{\min}}\right)$$

and hence by Remark 1,

$$\|u^{\text{sc}}\|_{\infty} = \mathcal{O}(\log(n^{1/4}/(m K_{\min})^{3/4} c_{\mu\nu}^{1/4})) \text{ and } \|u^{\text{sc}}\|_{\infty} = \mathcal{O}(\log(m^{1/4}/(n K_{\min})^{3/4} c_{\mu\nu}^{1/4})).$$

Acknowledging that  $\log(1/K_{\min}^2) = 2\|C\|_{\infty}/\eta$ , we have

$$A + \|u^{\text{sc}}\|_{\infty} + \|v^{\text{sc}}\|_{\infty} + \log\left(\frac{(n \vee m) \varepsilon^2}{c_{\mu\nu}}\right) = \mathcal{O}\left(\frac{\|C\|_{\infty}}{\eta} + \log\left(\frac{(n \vee m)^2}{n m c_{\mu\nu}^{7/2}}\right)\right).$$

Letting  $\Omega_{\kappa} := |1 - \kappa|(\|u^{\text{sc}}\|_{\infty} \|\mu^{\text{sc}}\|_1 + \|\mu^{\text{sc}} - \mu\|_1) + |1 - \kappa^{-1}|(\|v^{\text{sc}}\|_{\infty} \|\nu^{\text{sc}}\|_1 + \|\nu^{\text{sc}} - \nu\|_1)$ . We have that

$$\begin{aligned}
\Omega_{\kappa} &= \mathcal{O}\left(\left(\frac{\|C\|_{\infty}}{\eta} + \log\left(\frac{1}{(nm)^{3/4} c_{\mu\nu}^{1/2}}\right)\right)\left(|1 - \kappa|(\|\mu^{\text{sc}}\|_1 + \|\mu^{\text{sc}} - \mu\|_1) + |1 - \kappa^{-1}|(\|\nu^{\text{sc}}\|_1 + \|\nu^{\text{sc}} - \nu\|_1)\right)\right) \\
&= \mathcal{O}\left(\left(\frac{\|C\|_{\infty}}{\eta} + \log\left(\frac{(n \vee m)^2}{n m c_{\mu\nu}^{7/2}}\right)\right)(|1 - \kappa| \|\mu^{\text{sc}}\|_1 + |1 - \kappa^{-1}| \|\nu^{\text{sc}}\|_1 + |1 - \kappa| + |1 - \kappa^{-1}|)\right).
\end{aligned}$$

Hence, we arrive at

$$\Psi_{\varepsilon, \kappa}(u^{\text{sc}}, v^{\text{sc}}) - \Psi(u^*, v^*) = \mathcal{O}(R(\|\mu - \mu^{\text{sc}}\|_1 + \|\nu - \nu^{\text{sc}}\|_1 + \omega_\kappa)).$$

□

To more characterize  $\omega_\kappa$ , the following lemma expresses an upper bound with respect to  $\ell_1$ -norm of  $\mu^{\text{sc}}$  and  $\nu^{\text{sc}}$ .

**Lemma 2.** *Let  $(u^{\text{sc}}, v^{\text{sc}})$  be an optimal solution of problem (6). Then one has*

$$\|\mu^{\text{sc}}\|_1 = \mathcal{O}\left(\frac{n_b \sqrt{m}}{\sqrt{n K_{\min} c_{\mu\nu}}} + (n - n_b) \left( \frac{m_b}{\sqrt{n m c_{\mu\nu} K_{\min}^{3/2}}} + \frac{m - m_b}{\sqrt{n m K_{\min}}} \right)\right), \quad (8)$$

and

$$\|\nu^{\text{sc}}\|_1 = \mathcal{O}\left(\frac{m_b \sqrt{n}}{\sqrt{m K_{\min} c_{\mu\nu}}} + (m - m_b) \left( \frac{n_b}{\sqrt{n m c_{\mu\nu} K_{\min}^{3/2}}} + \frac{n - n_b}{\sqrt{n m K_{\min}}} \right)\right). \quad (9)$$

*Proof.* Using inequality (8), we obtain

$$\begin{aligned} \|\mu^{\text{sc}}\|_1 &= \sum_{i \in I_{\varepsilon, \kappa}} \mu_i^{\text{sc}} + \sum_{i \in I_{\varepsilon, \kappa}^{\mathfrak{G}}} \mu_i^{\text{sc}} \\ &\stackrel{(2)}{=} \kappa \|\mu_{I_{\varepsilon, \kappa}}^{\text{sc}}\|_1 + \frac{\varepsilon}{\kappa} \sum_{i \in I^{\mathfrak{G}}} \left( \sum_{j \in J_{\varepsilon, \kappa}} K_{ij} e^{v_j^{\text{sc}}} + \varepsilon \kappa \sum_{j \in J_{\varepsilon, \kappa}^{\mathfrak{G}}} K_{ij} \right) \\ &\stackrel{(8)}{\leq} \kappa \|\mu_{I_{\varepsilon, \kappa}}^{\text{sc}}\|_1 + (n - n_b) \left( \frac{m_b \max_{j \in J_{\varepsilon, \kappa}} \nu_j}{n \kappa K_{\min}} + (m - m_b) \varepsilon^2 \right). \end{aligned}$$

Using Remark 1, we get the desired closed form in (8). Similarly, we can prove the same statement for  $\|\nu^{\text{sc}}\|_1$ . □

## 5 Additional experimental results

**Experimental setup.** All computations have been run on each single core of an Intel Xeon E5-2630 processor clocked at 2.4 GHz in a Linux machine with 144 Gb of memory.

**On the use of a constrained L-BFGS-B solver.** It is worth to note that standard Sinkhorn's alternating minimization cannot be applied for the constrained screened dual problem (6). This appears more clearly while writing its optimality conditions (see Equations (2) and (3)). We resort to a L-BFGS-B algorithm to solve the constrained convex optimization problem on the screened variables (6), but any other efficient solver (e.g., proximal based method or Newton method) could be used. The choice of the starting point for the L-BFGS-B algorithm is given by the solution of the RESTRICTED SINKHORN method (see Algorithm 1), which is a Sinkhorn-like algorithm applied to the active dual variables. While simple and efficient the solution of this RESTRICTED SINKHORN algorithm does not satisfy the lower bound constraints of Problem (6). We further note that, as for the SINKHORN algorithm, our SCREENKHORN algorithm can be accelerated using a GPU implementation<sup>1</sup> of the L-BFGS-B algorithm [3].

**Comparison with other solvers.** We have considered experiments with GREENKHORN algorithm [2] but the implementation in POT library and our Python version of Matlab Altschuler's Greenkhorn code<sup>2</sup> were not competitive with SINKHORN. Hence, for both versions, SCREENKHORN is more competitive than GREENKHORN. The computation time gain reaches an order of 30 when comparing our method with GREENKHORN while SCREENKHORN is almost 2 times faster than SINKHORN.

<sup>1</sup><https://github.com/nepluno/lbfgsb-gpu>

<sup>2</sup><https://github.com/JasonAltschuler/OptimalTransportNIPS17>

---

**Algorithm 1: RESTRICTED SINKHORN**


---

1. **set:**  $\bar{f}_u = \varepsilon \kappa \mathcal{C}(K_{I_{\varepsilon, \kappa}, J_{\varepsilon, \kappa}^0}), \bar{f}_v = \varepsilon \kappa^{-1} r(K_{I_{\varepsilon, \kappa}, J_{\varepsilon, \kappa}^0});$
  2. **for**  $t = 1, 2, 3$  **do**

$f_v^{(t)} \leftarrow K_{I_{\varepsilon, \kappa}, J_{\varepsilon, \kappa}}^\top u + \bar{f}_v;$   
 $v^{(t)} \leftarrow \frac{\nu_{J_{\varepsilon, \kappa}}}{\kappa f_v^{(t)}};$   
 $f_u^{(t)} \leftarrow K_{I_{\varepsilon, \kappa}, J_{\varepsilon, \kappa}} v + \bar{f}_u;$   
 $u^{(t)} \leftarrow \frac{\kappa \mu_{I_{\varepsilon, \kappa}}}{f_u^{(t)}};$   
 $u \leftarrow u^{(t)}, v \leftarrow v^{(t)};$
  3. **return**  $(u^{(t)}, v^{(t)})$
- 

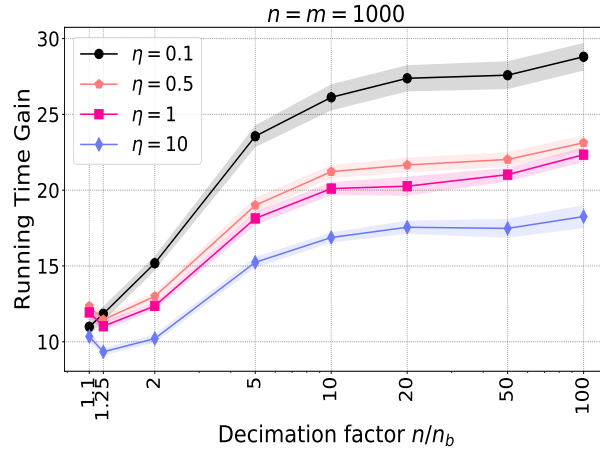


Figure 1:  $\frac{T_{\text{GREENKHORN}}}{T_{\text{SCREENKHORN}}}$ : Running time gain for the toy problem (see Section 5.2) as a function of the data decimation factor in SCREENKHORN, for different settings of the regularization parameter  $\eta$ .

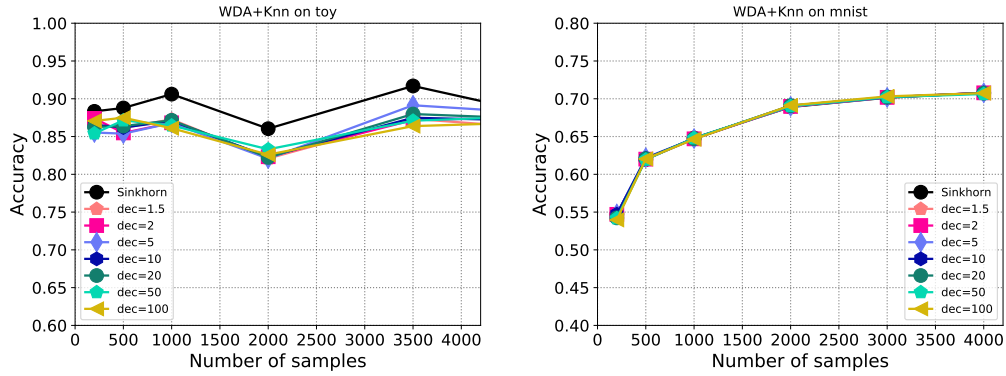


Figure 2: Accuracy of a 1-nearest-neighbour after WDA for the (left) toy problem and, (right) MNIST). We note a slight loss of performance for the toy problem, whereas for MNIST, all approaches yield the same performance.



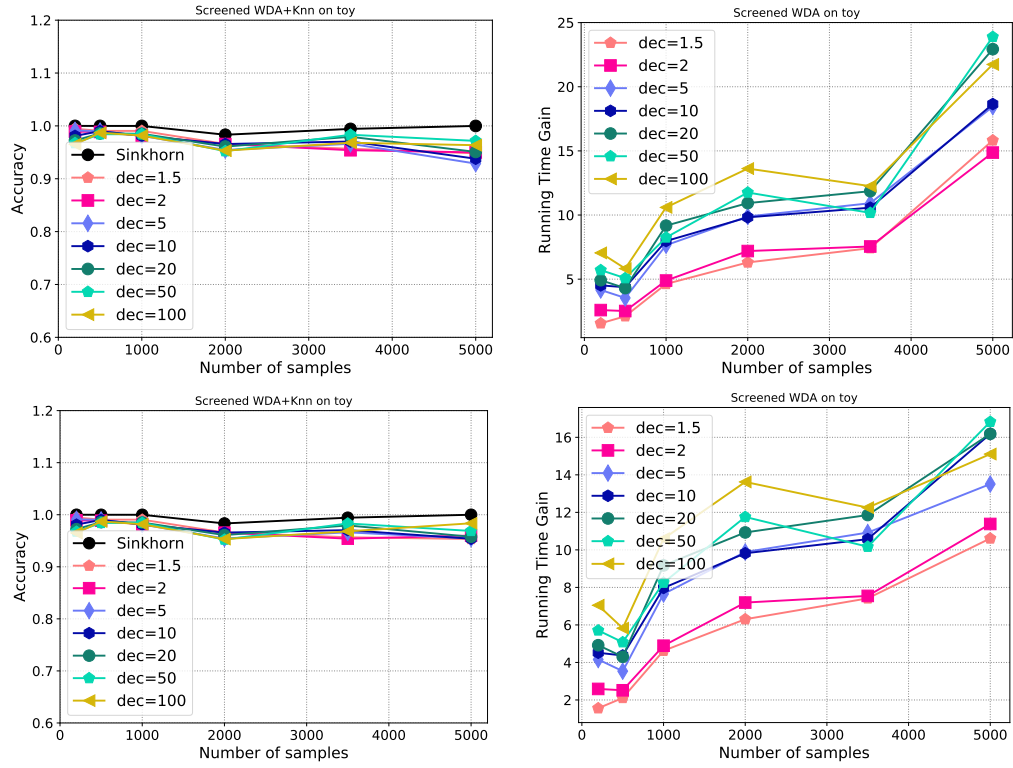


Figure 3: (top-left) Accuracy and (bottom-right) computational time gain on the toy dataset for  $\eta = 0.1$  and 1-nearest-neighbour. (bottom) accuracy and gain but for a 5-nearest-neighbour. We can note that a slight loss of performances occur for larger training set sizes especially for 1-nearest-neighbour. Computational gains increase with the dataset size and are on average of the order of magnitude.

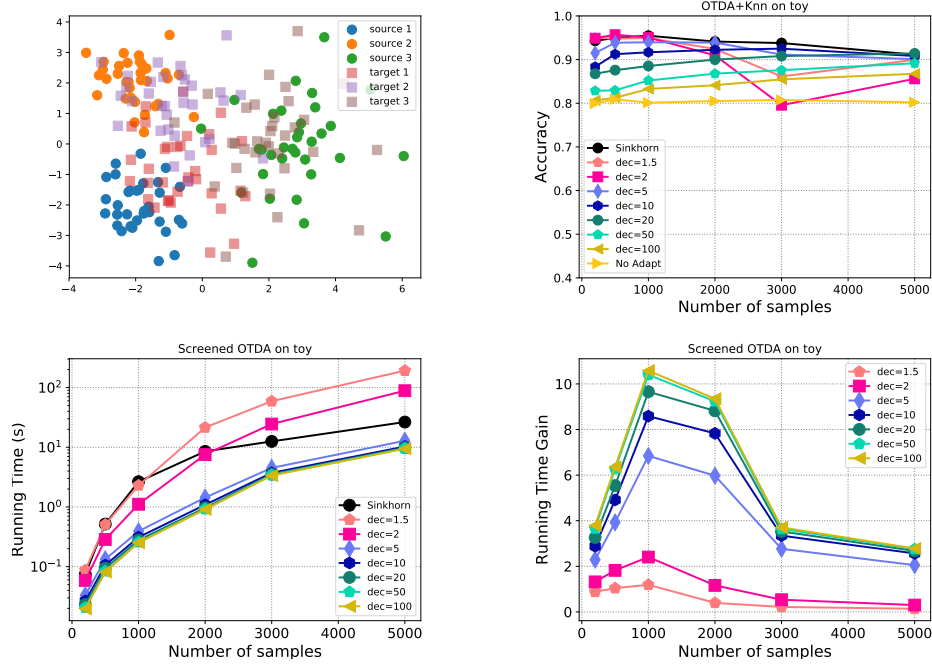


Figure 4: OT Domain Adaptation on a 3-class Gaussian toy problem. (top-left) Examples of source and target samples. (top-right) Evolution of the accuracy of a 1-nearest-neighbour classifier with respect to the number of samples. (bottom-left) Running time of the SINKHORN and SCREENKHORN for different decimation factors. (bottom-right). Gain in computation time. This toy problem is a problem in which classes are overlapping and distance between samples are rather limited. According to our analysis, this may be a situation in which SCREENKHORN may result in smaller computational gain. We can remark that with respect to the accuracy SCREENKHORN with decimation factors up to 10 are competitive with SINKHORN, although a slight loss of performance. Regarding computational time, for this example, small decimation factors does not result in gain. However for above 5-factor decimation, the gain goes from 2 to 10 depending on the number of samples.

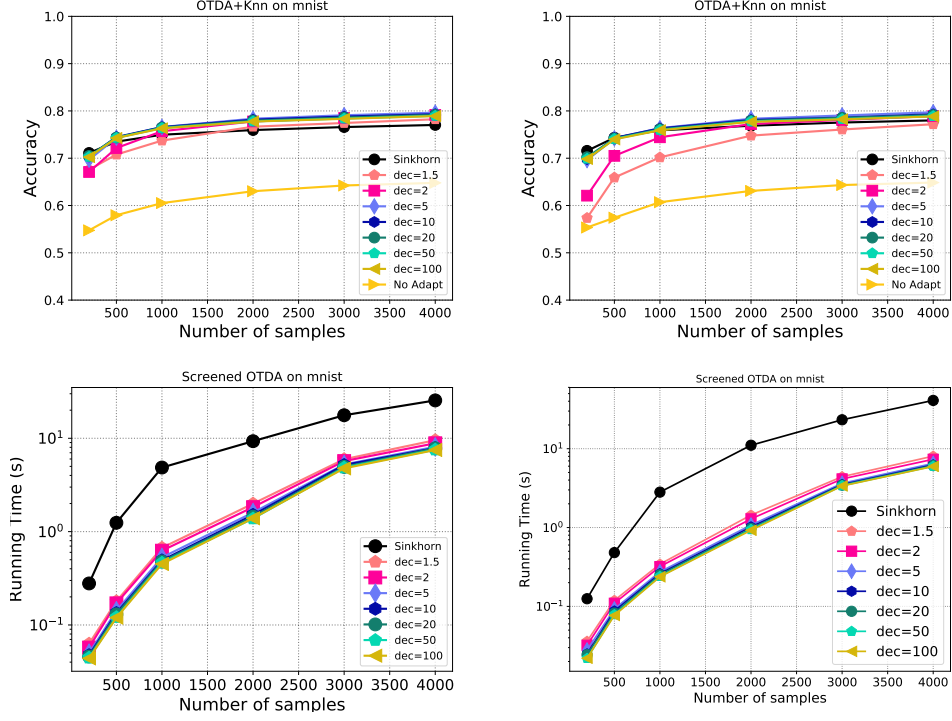


Figure 5: OT Domain adaptation MNIST to USPS : (top) Accuracy and (bottom) running time of SINKHORN and SCREENKHORN for hyperparameter of the  $\ell_{p,1}$  regularizer (left)  $\lambda = 1$  and (right)  $\lambda = 10$ . Note that this value impacts the ground cost of each Sinkhorn problem involved in the iterative algorithm. We remark that the strenght of the class-based regularization has influence on the performance of SCREENKHORN given a decimation factor. For small value on the left, SCREENKHORN slightly performs better than SINKHORN, while for large value, some decimation factors leads to loss of performances. Regarding, running time, we can note that SINKHORN is far less efficient than SCREENKHORN with an order of magnitude for intermediate number of samples.

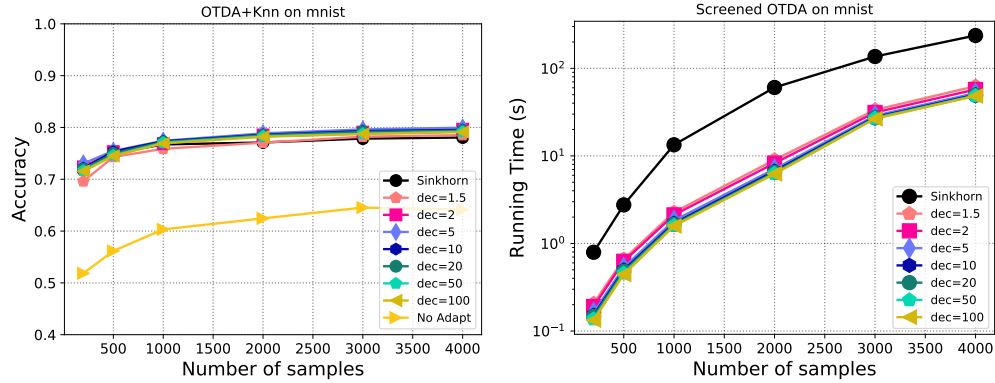


Figure 6: OT Domain adaptation MNIST to USPS : (left) Accuracy and (right) running time of SINKHORN and SCREENKHORN for the best performing (on average of 10 trials) hyperparameter  $\ell_{p,1}$  chosen among the set  $\{0.1, 1, 5, 10\}$ . We can note that in this situation, there is not loss of accuracy while our SCREENKHORN is still about an order of magnitude more efficient than Sinkhorn.

## References

- [1] B. K. Abid and R. Gower. Stochastic algorithms for entropy-regularized optimal transport problems. In Amos Storkey and Fernando Perez-Cruz, editors, *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pages 1505–1512, Playa Blanca, Lanzarote, Canary Islands, 2018. PMLR.
- [2] J. Altschuler, J. Weed, and P. Rigollet. Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 1964–1974. Curran Associates, Inc., 2017.
- [3] Y. Fei, G. Rong, B. Wang, and W. Wang. Parallel L-BFGS-B algorithm on GPU. *Computers & Graphics*, 40:1 – 9, 2014.
- [4] T. Lin, N. Ho, and M. I. Jordan. On efficient optimal transport: An analysis of greedy and accelerated mirror descent algorithms. *CoRR*, abs/1901.06482, 2019.