

---

# Supplement to “Screening Sinkhorn Algorithm via Dual Projections”

---

Anonymous Author(s)

Affiliation

Address

email

## 1 Proof of Lemma 1

Introducing two dual variables  $\lambda \in \mathbb{R}_+^n$  and  $\beta \in \mathbb{R}_+^m$  for each constraint, the Lagrangian of problem (5) 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  $e^{u_i^*} \geq \varepsilon \kappa^{-1}$ . Further, the condition on the dual variable  $\lambda_i^* > 0$  ensures that  $e^{u_i^*} = \varepsilon \kappa^{-1}$  and hence  $i \in I_{\varepsilon, \kappa}^{\mathbf{C}}$ . 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 Lemma 2

We prove only the first statement (9) and symmetrically we can prove the second statement (10). 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  $\lambda_i^{\text{sc}} = 0$ . Then  $e^{u_i^{\text{sc}}} \sum_{j=1}^m K_{ij} e^{v_j^{\text{sc}}} = \kappa \mu_i$ . 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  $\sum_{j=1}^m e^{v_j^{\text{sc}}} = \sum_{j \in J_{\varepsilon, \kappa}} e^{v_j^{\text{sc}}} + \sum_{j \in J_{\varepsilon, \kappa}^{\mathbf{C}}} e^{v_j^{\text{sc}}} \geq \varepsilon \kappa |J_{\varepsilon, \kappa}| + \varepsilon \kappa |J_{\varepsilon, \kappa}^{\mathbf{C}}| = \varepsilon \kappa m$ . Then

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

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

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

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

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

19 Using (1), we get

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

20 Therefore,

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

21

□

### 22 3 Proof of Lemma 3

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

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

24 Using Inequality (10), we obtain

$$\begin{aligned} \|\mu^{\text{sc}}\|_1 &= \sum_{i \in I_{\varepsilon, \kappa}} \mu_i^{\text{sc}} + \sum_{i \in I_{\varepsilon, \kappa}^{\mathcal{G}}} \mu_i^{\text{sc}} \\ &\stackrel{(2)}{=} \kappa \|\mu_{I_{\varepsilon, \kappa}}^{\text{sc}}\|_1 + \frac{\varepsilon}{\kappa} \sum_{i \in I_{\varepsilon, \kappa}^{\mathcal{G}}} \left( \sum_{j \in J_{\varepsilon, \kappa}} K_{ij} e^{v_j^{\text{sc}}} + \varepsilon \kappa \sum_{j \in J_{\varepsilon, \kappa}^{\mathcal{G}}} K_{ij} \right) \\ &\stackrel{(10)}{\leq} \kappa \|\mu_{I_{\varepsilon, \kappa}}^{\text{sc}}\|_1 + (n - m_b) \left( \frac{m_b \max_{j \in J_{\varepsilon, \kappa}} \nu_j}{n \kappa \min_{i,j} K_{ij}} + (m - m_b) \varepsilon^2 \right) \end{aligned}$$

25 Similarly, we can prove the same statement for  $\|\nu^{\text{sc}}\|_1$ .

26

□

### 27 4 Proof of Proposition 1

28 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  
29 not a metric, it is easy to see that  $\varrho$  is not nonnegative and satisfies  $\varrho(a, b) = 0$  iff  $a = b$ .

30 Straightforward,

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

31 Using (10), we have in one hand

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

32 On the other hand

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

33 Then

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

34 Therefore

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

35 Therefore

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

36 Finally, by Lemma 6 in [1] we get

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

## 37 **References**

- 38 [1] J. Altschuler, J. Weed, and P. Rigollet. Near-linear time approximation algorithms for optimal transport  
39 via sinkhorn iteration. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan,  
40 and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 1964–1974. Curran  
41 Associates, Inc., 2017.