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

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

First order conditions then yield that the Lagrangian multiplicators solutions λ^* and β^* satisfy

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

which leads to

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

For all $i = 1, \ldots, n$ 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 $i \in I^{\complement}_{\varepsilon,\kappa}$. 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 $e^{v_j^*}$.

$\square$

2 Proof of Lemma 2

We prove only the first statement (9) and similarly 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 $\sum_{j=1}^m e^{v_j^{\text{sc}}} = \sum_{j \in J_{\varepsilon,\kappa}} e^{v_j^{\text{sc}}} + \sum_{j \in J^{\complement}_{\varepsilon,\kappa}} e^{v_j^{\text{sc}}} \geq \varepsilon\kappa|J_{\varepsilon,\kappa}| + \varepsilon\kappa|J^{\complement}_{\varepsilon,\kappa}| = \varepsilon\kappa m$. 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 $j \in J_{\varepsilon,\kappa}$

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

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

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

Therefore,

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

□

3 Proof of Lemma 3

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

$$\mu_i^{\text{sc}} = \begin{cases} e^{u_i^{\text{sc}}} \sum_{j=1}^m K_{ij} e^{v_j^{\text{sc}}}, \text{ if } i \in I_{\varepsilon,\kappa}, \\ \frac{\varepsilon}{\kappa} \sum_{j=1}^m K_{ij} e^{v_j^{\text{sc}}}, \text{ if } i \in I_{\varepsilon,\kappa}^{\complement} \end{cases} = \begin{cases} \kappa\mu_i, \text{ if } i \in I_{\varepsilon,\kappa}, \\ \frac{\varepsilon}{\kappa} \sum_{j=1}^m K_{ij} e^{v_j^{\text{sc}}}, \text{ if } i \in I_{\varepsilon,\kappa}^{\complement}. \end{cases} \tag{2}$$

Using Inequality (10), we obtain

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

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

□

4 Proof of Proposition 1

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

Straighforward,

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

Using (10), we have in one hand

$$\begin{aligned}
\sum_{i\in I^{\complement}_{\varepsilon,\kappa}} \frac{\varepsilon}{\kappa}\sum_{j=1}^{m} K_{ij}e^{v_j^{\text{sc}}} &= \sum_{i\in I^{\complement}_{\varepsilon,\kappa}} \frac{\varepsilon}{\kappa}\Big(\sum_{j\in J_{\varepsilon,\kappa}} K_{ij}e^{v_j^{\text{sc}}} + \varepsilon\kappa \sum_{j\in J^{\complement}_{\varepsilon,\kappa}} K_{ij}\Big)\\
&\le \sum_{i\in I^{\complement}_{\varepsilon,\kappa}} \frac{\varepsilon}{\kappa}\Big(m_b \max_{i,j} K_{ij}\frac{\max_{j\in J_{\varepsilon,\kappa}}\nu_j}{n\varepsilon\min_{i,j}K_{ij}} + (m-m_b)\varepsilon\kappa\max_{i,j}K_{ij}\Big)\\
&\le (n-n_b)\Big(\frac{m_b\max_j \nu_j}{n\kappa\min_{i,j}K_{ij}} + (m-m_b)\varepsilon^2\Big).
\end{aligned}$$

On the other hand

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

Then

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

Therefore

$$\begin{aligned}
&\sum_{i\in I^{\complement}_{\varepsilon,\kappa}} \frac{\varepsilon}{\kappa}\sum_{j=1}^{m}K_{ij}e^{v_j^{\text{sc}}} - \mu_i + \mu_i\log\Big(\frac{\mu_i}{\frac{\varepsilon}{\kappa}\sum_{j=1}^{m}K_{ij}e^{v_j^{\text{sc}}}}\Big)\\
&\le (n-n_b)\Bigg(\frac{m_b}{n\kappa\min_{i,j}K_{ij}} + (m-m_b)\varepsilon^2 - \min_i\mu_i\\
&\quad + \max_i\mu_i\log\Big(\frac{\kappa(n-n_b+1)\max_i\mu_i}{m_b\min_{i,j}K_{ij}\min_{j\in J_{\varepsilon,\kappa}}\nu_j} + \frac{\kappa^2\max_i\mu_i}{mm_b\varepsilon^2(\min_{i,j}K_{ij})^2\min_{j\in J_{\varepsilon,\kappa}}\nu_j}\Big)\Bigg)
\end{aligned}$$

Therefore

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

Finally, by Lemma 6 in [1] we get

$$\begin{aligned}
\|\mu^\star - \mu^{\text{sc}}\|_1^2 \le & n_b(\kappa-\log(\kappa)-1)\max_i\mu_i + 7(n-n_b)\Bigg(\frac{m_b\max_j\nu_j}{n\kappa\min_{i,j}K_{ij}} + (m-m_b)\varepsilon^2 - \min_i\mu_i\\
&+ \max_i\mu_i\log\Big(\frac{\kappa(n-n_b+1)\max_i\mu_i}{m_b\min_{i,j}K_{ij}\min_{j\in J_{\varepsilon,\kappa}}\nu_j} + \frac{\kappa^2\max_i\mu_i}{mm_b\varepsilon^2(\min_{i,j}K_{ij})^2\min_{j\in J_{\varepsilon,\kappa}}\nu_j}\Big).
\end{aligned}$$

Proof of the upper bound for $\|\nu - \nu^{\text{sc}}\|_1^2$ follows the same lines as above.

$\square$

References

[1] 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.