A. Proofs

A.1. Proof for Proposition 1

Proof. Let $\pi_\star$ be a minimizer of (3). Then using the optimality condition for $\sup_{c\in\mathcal{C}(\mathcal{X}^2)}\int c\,d\pi - F^*(c)$, any c such that $\pi_\star \in \partial F^*(c)$ is a best response to $\pi_\star$. But by Fenchel-Young inequality, such c are exactly those in $\partial F(\pi_\star) = \{\nabla F(\pi_\star)\}$. Since $\nabla F(\pi_\star)$ is the unique best response to $\pi_\star$, it is necessarily optimal in (4). Conversely, if there is a unique maximizer $c_\star$, then as a result of the above, $c_\star = \nabla F(\pi_\star)$ for some minimizer $\pi_\star$ of the primal. Then $\nabla F^*(c_\star)$ is optimal in the primal. □

A.2. Proof for Remark 2

Proof. As in the proof of Theorem 1:

$$\begin{aligned}\inf_{\pi\in\Pi(\mu,\nu)} F(\pi) &= \inf_{\pi\in\Pi(\mu,\nu)} -(-F)^{**}(\pi)\\ &= \inf_{\pi\in\Pi(\mu,\nu)} -\sup_{c\in\mathcal{C}(\mathcal{X}^2)}\int c\,d\pi - (-F)^*(c)\\ &= \inf_{\pi\in\Pi(\mu,\nu)}\inf_{c\in\mathcal{C}(\mathcal{X}^2)}\int -c\,d\pi + (-F)^*(c)\\ &= \inf_{\pi\in\Pi(\mu,\nu)}\inf_{c\in\mathcal{C}(\mathcal{X}^2)}\int c\,d\pi + (-F)^*(-c)\\ &= \inf_{c\in\mathcal{C}(\mathcal{X}^2)} \mathscr{T}_c(\mu,\nu) + (-F)^*(-c).\end{aligned}$$

□

A.3. Proof for Proposition 2

Proof. As in the proof of Theorem 1, we use Sion's minimax theorem to get

$$\begin{aligned}&\sup_{\mathbf{c}\in\mathbb{R}_+^{n\times n}}\min_{\boldsymbol{\pi}\in\Pi(\boldsymbol{\mu},\boldsymbol{\nu})}\langle\mathbf{c},\boldsymbol{\pi}\rangle - \varepsilon\sum_{ij} R_{ij}^*\left(\frac{\mathbf{c}_{ij}-\mathbf{c}_{0ij}}{\varepsilon}\right)\\ &= \min_{\boldsymbol{\pi}\in\Pi(\boldsymbol{\mu},\boldsymbol{\nu})}\sup_{\mathbf{c}\in\mathbb{R}_+^{n\times n}}\langle\mathbf{c},\boldsymbol{\pi}\rangle - \varepsilon\sum_{ij} R_{ij}^*\left(\frac{\mathbf{c}_{ij}-\mathbf{c}_{0ij}}{\varepsilon}\right).\end{aligned}$$

Since the optimization in $\mathbf{c}\in\mathbb{R}_+^{n\times n}$ is separable, we only need to consider this optimization coordinate by coordinate, *i.e.* we only need to compute $\sup_{\mathbf{c}_{ij}\in\mathbb{R}_+}\boldsymbol{\pi}_{ij}\mathbf{c}_{ij} - f_{ij}^*(\mathbf{c}_{ij})$ for all $i,j\in[\![n]\!]$, where $f_{ij}^*(\mathbf{c}_{ij}) = \varepsilon R_{ij}^*\left(\frac{\mathbf{c}_{ij}-\mathbf{c}_{0ij}}{\varepsilon}\right)$.

Fix $\boldsymbol{\pi}\in\Pi(\boldsymbol{\mu},\boldsymbol{\nu})$ and $i,j\in[\![n]\!]$, and define $g_{ij}:\mathbb{R}\ni\mathbf{c}_{ij}\mapsto\boldsymbol{\pi}_{ij}\mathbf{c}_{ij} - f_{ij}^*(\mathbf{c}_{ij})$.

Suppose that $z_{ij} = f_{ij}'(\boldsymbol{\pi}_{ij}) \geq 0$. Then

$$f_{ij}(\boldsymbol{\pi}_{ij}) = f_{ij}^{**}(\boldsymbol{\pi}_{ij}) = g_{ij}(z_{ij}) = \sup_{\mathbf{c}_{ij}\in\mathbb{R}} g_{ij}(\mathbf{c}_{ij}),$$

and since $z_{ij}\geq 0$, $\sup_{\mathbf{c}_{ij}\in\mathbb{R}_+} g_{ij}(\mathbf{c}_{ij}) = f_{ij}(\boldsymbol{\pi}_{ij})$. This means that $\widehat{R}_{ij}(\boldsymbol{\pi}_{ij}) = R_{ij}(\boldsymbol{\pi}_{ij})$.

Suppose now that $z_{ij} = f_{ij}'(\boldsymbol{\pi}_{ij}) < 0$. This means that

$$\sup_{\mathbf{c}_{ij}\in\mathbb{R}_+} g_{ij}(\mathbf{c}_{ij}) < \sup_{\mathbf{c}_{ij}\in\mathbb{R}} g_{ij}(\mathbf{c}_{ij}).$$

Since g_{ij} is concave, this shows that $\sup_{\mathbf{c}_{ij}\in\mathbb{R}_+} g_{ij}(\mathbf{c}_{ij}) = g_{ij}(0) = -f_{ij}^*(0)$, *i.e.* $\widehat{R}_{ij}(\boldsymbol{\pi}_{ij}) = \frac{-\mathbf{c}_{0ij}}{\varepsilon}\boldsymbol{\pi}_{ij} - R_{ij}^*\left(\frac{-\mathbf{c}_{0ij}}{\varepsilon}\right)$.

Since R_{ij} is convex, R_{ij}' is increasing with pseudo-inverse $R_{ij}^{*\prime}$. Furthermore, the optimality condition in the convex conjugate problem gives, for any $\alpha\in\mathbb{R}$:

$$R_{ij}^*(\alpha) = \alpha\times R_{ij}^{*\prime}(\alpha) - R_{ij}\circ R_{ij}^{*\prime}(\alpha).$$

So if R_{ij} is of class C^1, taking $\alpha = \frac{-\mathbf{c}_{0ij}}{\varepsilon}$, as x increases to $R_{ij}^{*\prime}\left(-\frac{\mathbf{c}_{0ij}}{\varepsilon}\right)$:

$$\widehat{R}_{ij}(x) \longrightarrow R_{ij}\circ R_{ij}^{*\prime}\left(-\frac{\mathbf{c}_{0ij}}{\varepsilon}\right) = \widehat{R}_{ij}\circ R_{ij}^{*\prime}\left(-\frac{\mathbf{c}_{0ij}}{\varepsilon}\right),$$

meaning that $\widehat{R}_{ij}$ is of class C^1. □

A.4. Proof for Example 3

Proof. We denote by $\operatorname{sgn}(x)$ the set $\{+1\}$ if $x>0$, $\{-1\}$ if $x<0$ and $[-1,1]$ if $x=0$. We apply Corollary 1 with $R:\mathbb{R}^{n\times n}\to\mathbb{R}$ defined as $R(\boldsymbol{\pi}) = \frac{1}{p}\|\boldsymbol{\pi}\|_{\mathbf{w},p}^p$, for which we need to compute its convex conjugate:

$$R^*(\mathbf{c}) = \sup_{\boldsymbol{\pi}\in\mathbb{R}^{n\times n}}\langle\boldsymbol{\pi},\mathbf{c}\rangle - \frac{1}{p}\sum_{ij}\mathbf{w}_{ij}|\boldsymbol{\pi}_{ij}|^p.$$

Subdifferentiating with respect to $\boldsymbol{\pi}_{ij}$:

$$\begin{aligned}\mathbf{c}_{ij} &\in \frac{1}{p}\mathbf{w}_{ij}\frac{\partial}{\partial\boldsymbol{\pi}_{ij}}|\boldsymbol{\pi}_{ij}|^p\\ &= \mathbf{w}_{ij}\operatorname{sgn}(\boldsymbol{\pi}_{ij})|\boldsymbol{\pi}_{ij}|^{p-1}\end{aligned}$$

This implies that $\operatorname{sgn}(\boldsymbol{\pi}_{ij}) = \operatorname{sgn}(\mathbf{c}_{ij})$, so:

$$\boldsymbol{\pi}_{ij} = \operatorname{sgn}(\mathbf{c}_{ij})\left|\frac{\mathbf{c}_{ij}}{\mathbf{w}_{ij}}\right|^{q-1}.$$

Finally,

$$\begin{aligned}R^*(\mathbf{c}) &= \sum_{ij}\mathbf{c}_{ij}\operatorname{sgn}(\mathbf{c}_{ij})\left|\frac{\mathbf{c}_{ij}}{\mathbf{w}_{ij}}\right|^{q-1} - \frac{1}{p}\mathbf{w}_{ij}\left|\frac{\mathbf{c}_{ij}}{\mathbf{w}_{ij}}\right|^q\\ &= \frac{1}{q}\sum_{ij}\frac{1}{\mathbf{w}_{ij}^{q-1}}|\mathbf{c}_{ij}|^q\\ &= \frac{1}{q}\|\mathbf{c}\|_{1/\mathbf{w}^{q-1},q}^q.\end{aligned}$$

□

A.5. Proof for Example 6

Proof. Since $\boldsymbol{\pi} \in \Pi(\boldsymbol{\mu}, \boldsymbol{\nu})$, $\sum_{ij} \boldsymbol{\pi}_{ij} = 1$ so we can drop it for now and only consider the term $R(\boldsymbol{\pi}) = \frac{1}{q-1}\|\boldsymbol{\pi}\|_q^q$ which is separable in the coordinates of $\boldsymbol{\pi}$:

$$R(\boldsymbol{\pi}) = \sum_{ij} f(\boldsymbol{\pi}_{ij})$$

where we have defined the convex function

$$f(x) = \begin{cases} \frac{1}{q-1}x^q & \text{if } x \geq 0 \\ +\infty & \text{otherwise.} \end{cases}$$

We compute its convex conjugate:

$$\begin{aligned} f^*(y) &= \sup_{x \geq 0} \left\{ xy - \frac{1}{q-1}x^q \right\} \\ &= \begin{cases} \left(\frac{y}{p}\right)^p & \text{if } y \leq 0 \\ +\infty & \text{if } y > 0 \end{cases} \end{aligned}$$

where $p = \frac{q}{q-1} \leq 0$ is such that $1/p + 1/q = 1$. Then $R^*(\mathbf{c}) = +\infty$ if $\mathbf{c}$ has a positive entry, and over $\mathbb{R}_-^{n \times n}$:

$$\begin{aligned} R^*(\mathbf{c}) &= \sum_{ij} f^*(\mathbf{c}_{ij}) = \sum_{ij} \left(\frac{\mathbf{c}_{ij}}{p}\right)^p \\ &= \sum_{ij} \left(\frac{-\mathbf{c}_{ij}}{-p}\right)^p \\ &= (-p)^{-p} \sum_{ij} \left(\frac{1}{-\mathbf{c}_{ij}}\right)^{-p}. \end{aligned}$$

Adding the term $\frac{\varepsilon}{1-q}$ we left aside to the result of Corollary 1, we find that Tsallis regularized OT is equal to:

$$\begin{aligned} &\sup_{\mathbf{c} \in \mathbb{R}^{n \times n}} \mathscr{T}_{\mathbf{c}}(\boldsymbol{\mu}, \boldsymbol{\nu}) - \varepsilon R^* \left(\frac{\mathbf{c} - \mathbf{c}_0}{\varepsilon}\right) + \frac{\varepsilon}{1-q} \\ &= \sup_{\substack{\mathbf{c} \in \mathbb{R}^{n \times n} \\ \mathbf{c} \leq \mathbf{c}_0}} \mathscr{T}_{\mathbf{c}}(\boldsymbol{\mu}, \boldsymbol{\nu}) - \varepsilon(-p)^{-p} \sum_{ij} \left[\frac{\varepsilon}{\mathbf{c}_{0ij} - \mathbf{c}_{ij}}\right]^{-p} \\ &\qquad + \frac{\varepsilon}{1-q} \\ &= \sup_{\substack{\mathbf{c} \in \mathbb{R}^{n \times n} \\ \mathbf{c} \leq \mathbf{c}_0}} \mathscr{T}_{\mathbf{c}}(\boldsymbol{\mu}, \boldsymbol{\nu}) - \varepsilon^{\frac{1}{1-q}}(-p)^{-p} \sum_{ij} \left[\frac{1}{\mathbf{c}_{0ij} - \mathbf{c}_{ij}}\right]^{-p} \\ &\qquad + \frac{\varepsilon}{1-q} \\ &= \sup_{\substack{\mathbf{c} \in \mathbb{R}^{n \times n} \\ \mathbf{c} \leq \mathbf{c}_0}} \mathscr{T}_{\mathbf{c}}(\boldsymbol{\mu}, \boldsymbol{\nu}) - \varepsilon^{\frac{1}{1-q}}(-p)^{-p} \left\|\frac{1}{\mathbf{c}_0 - \mathbf{c}}\right\|_{-p}^{-p} \\ &\qquad + \frac{\varepsilon}{1-q}. \end{aligned}$$

□

A.6. Proof for Subsection 7.2

Entropic OT In the case of entropic OT,

$$F(\boldsymbol{\pi}) = \langle \boldsymbol{\pi}, \mathbf{c}_0 \rangle + \varepsilon \sum_{ij} \boldsymbol{\pi}_{ij} \left[\log \boldsymbol{\pi}_{ij} - 1\right],$$

so

$$F^*(\mathbf{c}) = \varepsilon \sum_{ij} \exp\left(\frac{\mathbf{c}_{ij} - \mathbf{c}_{0ij}}{\varepsilon}\right)$$

and

$$\nabla F^*(\mathbf{c}) = \left[\exp\left(\frac{\mathbf{c}_{ij} - \mathbf{c}_{0ij}}{\varepsilon}\right)\right]_{ij}.$$

Then the system of equations (13) (14) is:

$$\begin{aligned} \forall i,\ \boldsymbol{\mu}_i &= \sum_j \exp\left(\frac{\boldsymbol{\phi}_{\star i} + \boldsymbol{\psi}_{\star j} - \mathbf{c}_{0ij}}{\varepsilon}\right) \\ &= \exp(\boldsymbol{\phi}_{\star i}/\varepsilon)\left[K \exp(\boldsymbol{\psi}_\star/\varepsilon)\right]_i \\ \forall j,\ \boldsymbol{\nu}_j &= \sum_i \exp\left(\frac{\boldsymbol{\phi}_{\star i} + \boldsymbol{\psi}_{\star j} - \mathbf{c}_{0ij}}{\varepsilon}\right) \\ &= \exp(\boldsymbol{\psi}_{\star j}/\varepsilon)\left[K^\top \exp(\boldsymbol{\phi}_\star/\varepsilon)\right]_j \end{aligned}$$

where $K = \exp(-\mathbf{c}_0/\varepsilon) \in \mathbb{R}^{n \times n}$ and exp is taken elementwise. Then solving alternatively for $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ is exactly Sinkhorn algorithm.

Quadratic OT In the case of quadratic OT, using the notations and results from example 8:

$$F(\boldsymbol{\pi}) = \langle \boldsymbol{\pi}, \mathbf{c}_0 \rangle + \varepsilon \varphi_2(\boldsymbol{\pi}_{ij}),$$

and

$$F^*(\mathbf{c}) = \frac{1}{2\varepsilon} \sum_{ij} \left[\left(\mathbf{c}_{ij} - \mathbf{c}_{0ij}\right)_+\right]^2.$$

Then:

$$\nabla F^*(\mathbf{c}) = \frac{1}{\varepsilon}\left(\mathbf{c} - \mathbf{c}_0\right)_+.$$

The system of equations (13) (14) is:

$$\begin{aligned} \forall i,\ \varepsilon \boldsymbol{\mu}_i &= \sum_j \left(\boldsymbol{\phi}_{\star i} + \boldsymbol{\psi}_{\star j} - \mathbf{c}_{0ij}\right)_+ \\ \forall j,\ \varepsilon \boldsymbol{\nu}_j &= \sum_i \left(\boldsymbol{\phi}_{\star i} + \boldsymbol{\psi}_{\star j} - \mathbf{c}_{0ij}\right)_+ \end{aligned}$$

which is what (Blondel et al., 2018) solve in their appendix B.