A Proofs

A.1 Euclidean Projection onto $\mathcal{S}$

It amounts to the following problem.

$$\underset{\mathbf{P}\in\mathcal{S}}{\arg\min}\ \|\mathbf{P}-\mathbf{K}\|_2^2\,. \tag{3}$$

With $\mathbf{W}\in\mathbb{R}^{n\times n}$, the Lagrangian takes the form:

$$\mathcal{L}(\mathbf{P},\mathbf{W}) = \|\mathbf{P}-\mathbf{K}\|_2^2 + \langle \mathbf{W}, \mathbf{P}-\mathbf{P}^\top\rangle\,. \tag{4}$$

Cancelling the gradient of $\mathcal{L}$ with respect to $\mathbf{P}$ gives $2(\mathbf{P}^\star-\mathbf{K})+\mathbf{W}-\mathbf{W}^\top=\mathbf{0}$. Thus $\mathbf{P}^\star = \mathbf{K}+\frac{1}{2}\left(\mathbf{W}^\top-\mathbf{W}\right)$. Using the symmetry constraint on $\mathbf{P}^\star$ yields $\mathbf{P}^\star=\frac{1}{2}\left(\mathbf{K}+\mathbf{K}^\top\right)$. Hence we have:

$$\underset{\mathbf{P}\in\mathcal{S}}{\arg\min}\ \|\mathbf{P}-\mathbf{K}\|_2^2 = \frac{1}{2}\left(\mathbf{K}+\mathbf{K}^\top\right)\,. \tag{5}$$

A.2 From Symmetric Entropy-Constrained OT to Sinkhorn Iterations

In this section, we derive Sinkhorn iterations from the problem (EOT). Let $\mathbf{C}\in\mathcal{D}$. We start by making the constraints explicit.

$$\min_{\mathbf{P}\in\mathbb{R}_+^{n\times n}} \quad \langle \mathbf{P},\mathbf{C}\rangle \tag{6}$$

$$\text{s.t.} \quad \sum_{i\in[\![n]\!]} \mathrm{H}(\mathbf{P}_{i:}) \geq \eta \tag{7}$$

$$\mathbf{P}\mathbf{1}=\mathbf{1}, \quad \mathbf{P}=\mathbf{P}^\top\,. \tag{8}$$

For the above convex problem the Lagrangian writes, where $\nu\in\mathbb{R}_+$, $\mathbf{f}\in\mathbb{R}^n$ and $\mathbf{\Gamma}\in\mathbb{R}^{n\times n}$:

$$\mathcal{L}(\mathbf{P},\mathbf{f},\nu,\mathbf{\Gamma}) = \langle \mathbf{P},\mathbf{C}\rangle + \Big\langle \nu, \eta - \sum_{i\in[\![n]\!]}\mathrm{H}(\mathbf{P}_i)\Big\rangle + 2\langle \mathbf{f}, \mathbf{1}-\mathbf{P}\mathbf{1}\rangle + \left\langle \mathbf{\Gamma}, \mathbf{P}-\mathbf{P}^\top\right\rangle\,. \tag{9}$$

Strong duality holds and the first order KKT condition gives for the optimal primal $\mathbf{P}^\star$ and dual $(\nu^\star,\mathbf{f}^\star,\mathbf{\Gamma}^\star)$ variables:

$$\nabla_{\mathbf{P}}\mathcal{L}(\mathbf{P}^\star,\mathbf{f}^\star,\nu^\star,\mathbf{\Gamma}^\star) = \mathbf{C}+\nu^\star\log\mathbf{P}^\star - 2\mathbf{f}^\star\mathbf{1}^\top + \mathbf{\Gamma}^\star - \mathbf{\Gamma}^{\star\top} = \mathbf{0}\,. \tag{10}$$

Since $\mathbf{P}^\star,\mathbf{C}\in\mathcal{S}$ we have $\mathbf{\Gamma}^\star-\mathbf{\Gamma}^{\star\top} = \mathbf{f}^\star\mathbf{1}^\top - \mathbf{1}\mathbf{f}^{\star\top}$. Hence $\mathbf{C}+\nu^\star\log\mathbf{P}^\star - \mathbf{f}^\star\oplus\mathbf{f}^\star = \mathbf{0}$. Suppose that $\nu^\star=0$ then the previous reasoning implies that $\forall(i,j), C_{ij}=f_i^\star+f_j^\star$. Using that $\mathbf{C}\in\mathcal{D}$ we have $C_{ii}=C_{jj}=0$ thus $\forall i, f_i^\star=0$ and thus this would imply that $\mathbf{C}=0$ which is not allowed by hypothesis. Therefore $\nu^\star\neq 0$ and the entropy constraint is saturated at the optimum by complementary slackness. Isolating $\mathbf{P}^\star$ then yields:

$$\mathbf{P}^\star = \exp\left((\mathbf{f}^\star\oplus\mathbf{f}^\star-\mathbf{C})/\nu^\star\right)\,. \tag{11}$$

$\mathbf{P}^\star$ must be primal feasible in particular $\mathbf{P}^\star\mathbf{1}=\mathbf{1}$. This constraint gives us the Sinkhorn fixed point relation for $\mathbf{f}^\star$:

$$\forall i\in[\![n]\!], \quad [\mathbf{f}^\star]_i = -\nu^\star\,\mathrm{LSE}\left((\mathbf{f}^\star-\mathbf{C}_{:i})/\nu^\star\right), \tag{12}$$

where for a vector $\boldsymbol{\alpha}$, we use the notation $\mathrm{LSE}(\boldsymbol{\alpha})=\log\sum_k\exp(\alpha_k)$.

A.3 Proof of Proposition 1

We recall the result

Proposition 1. *Let $\mathbf{C}\in\mathbb{R}^{n\times n}$ without constant rows. Then $\mathbf{P}^{\mathrm{e}}$ solves the entropic affinity problem (EA) with cost $\mathbf{C}$ if and only if $\mathbf{P}^{\mathrm{e}}$ is the unique solution of the convex problem*

$$\min_{\mathbf{P}\in\mathcal{H}_\xi}\ \langle\mathbf{P},\mathbf{C}\rangle. \tag{EA as OT}$$

Proof. We begin by rewriting the above problem to make the constraints more explicit.

$$\begin{aligned} \min_{\mathbf{P}\in\mathbb{R}_+^{n\times n}} \quad & \langle \mathbf{P}, \mathbf{C}\rangle \\ \text{s.t.} \quad & \forall i,\ \mathrm{H}(\mathbf{P}_{i:}) \geq \log \xi + 1 \\ & \mathbf{P}\mathbf{1} = \mathbf{1}\,. \end{aligned}$$

By concavity of entropy, one has that the entropy constraint is convex thus the above primal problem is a convex optimization problem. Moreover, the latter is strictly feasible for any $\xi \in [\![n-1]\!]$. Therefore Slater's condition is satisfied and strong duality holds.

Introducing the dual variables $\boldsymbol{\lambda} \in \mathbb{R}^n$ and $\boldsymbol{\varepsilon} \in \mathbb{R}_+^n$, the Lagrangian of the above problem writes:

$$\mathcal{L}(\mathbf{P}, \boldsymbol{\lambda}, \boldsymbol{\varepsilon}) = \langle \mathbf{P}, \mathbf{C}\rangle + \langle \boldsymbol{\varepsilon}, (\log\xi + 1)\mathbf{1} - \mathrm{H_r}(\mathbf{P})\rangle + \langle \boldsymbol{\lambda}, \mathbf{1} - \mathbf{P}\mathbf{1}\rangle\,, \tag{13}$$

where we recall that $\mathrm{H_r}(\mathbf{P}) = (\mathrm{H}(\mathbf{P}_{i:}))_i$. Note that we will deal with the constraint $\mathbf{P} \in \mathbb{R}_+^{n\times n}$ directly, hence there is no associated dual variable. Since strong duality holds, for any solution $\mathbf{P}^\star$ to the primal problem and any solution $(\boldsymbol{\varepsilon}^\star, \boldsymbol{\lambda}^\star)$ to the dual problem, the pair $\mathbf{P}^\star, (\boldsymbol{\varepsilon}^\star, \boldsymbol{\lambda}^\star)$ must satisfy the Karush-Kuhn-Tucker (KKT) conditions. The first-order optimality condition gives:

$$\nabla_{\mathbf{P}}\mathcal{L}(\mathbf{P}^\star, \boldsymbol{\varepsilon}^\star, \boldsymbol{\lambda}^\star) = \mathbf{C} + \operatorname{diag}(\boldsymbol{\varepsilon}^\star)\log\mathbf{P}^\star - \boldsymbol{\lambda}^\star\mathbf{1}^\top = \mathbf{0}\,. \tag{first-order}$$

Assume that there exists $\ell \in [\![n]\!]$ such that $\varepsilon^\star_\ell = 0$. Then (first-order) gives that the ℓ^{th} row of $\mathbf{C}$ is constant which is not allowed by hypothesis. Therefore $\boldsymbol{\varepsilon}^\star > \mathbf{0}$ (*i.e.*, $\boldsymbol{\varepsilon}^\star$ has positive entries). Thus isolating $\mathbf{P}^\star$ in the first order condition results in:

$$\mathbf{P}^\star = \operatorname{diag}(\mathbf{u})\exp\left(-\operatorname{diag}(\boldsymbol{\varepsilon}^\star)^{-1}\mathbf{C}\right) \tag{14}$$

where $\mathbf{u} = \exp\left(\boldsymbol{\lambda}^\star \oslash \boldsymbol{\varepsilon}^\star\right)$. This matrix must satisfy the stochasticity constraint $\mathbf{P}\mathbf{1} = \mathbf{1}$. Hence one has $\mathbf{u} = \mathbf{1} \oslash \left(\exp\left(\operatorname{diag}(\boldsymbol{\varepsilon}^\star)^{-1}\mathbf{C}\right)\mathbf{1}\right)$ and $\mathbf{P}^\star$ has the form

$$\forall (i,j) \in [\![n]\!]^2, \quad P^\star_{ij} = \frac{\exp\left(-C_{ij}/\varepsilon^\star_i\right)}{\sum_\ell \exp\left(-C_{i\ell}/\varepsilon^\star_i\right)}\,. \tag{15}$$

As a consequence of $\boldsymbol{\varepsilon}^\star > \mathbf{0}$, complementary slackness in the KKT conditions gives us that for all i, the entropy constraint is saturated *i.e.*, $\mathrm{H}(\mathbf{P}^\star_{i:}) = \log\xi + 1$. Therefore $\mathbf{P}^\star$ solves the problem (EA). Conversely any solution of (EA) $P^\star_{ij} = \frac{\exp(-C_{ij}/\varepsilon^\star_i)}{\sum_\ell \exp(-C_{i\ell}/\varepsilon^\star_i)}$ with $(\varepsilon^\star_i)$ such that $\mathrm{H}(\mathbf{P}^\star_{i:}) = \log\xi + 1$ gives an admissible matrix for $\min_{\mathbf{P}\in\mathcal{H}_\xi}\langle \mathbf{P}, \mathbf{C}\rangle$ and the associated variables satisfy the KKT conditions which are sufficient conditions for optimality since the problem is convex. □

A.4 Proof of Proposition 4 and Proposition 5

The goal of this section is to prove the following results:

Proposition 4 (Saturation of the entropies)**.** *Let $\mathbf{C} \in \mathcal{S}$ with zero diagonal, then (SEA) with cost $\mathbf{C}$ has a* unique solution *that we denote by $\mathbf{P}^{\mathrm{se}}$. If moreover $\mathbf{C} \in \mathcal{D}$, then for at least $n-1$ indices $i \in [\![n]\!]$ the solution satisfies $\mathrm{H}(\mathbf{P}^{\mathrm{se}}_{i:}) = \log\xi + 1$.*

Proposition 5 (Solving for SEA)**.** *Let $\mathbf{C} \in \mathcal{D}$, $\mathcal{L}(\mathbf{P}, \boldsymbol{\gamma}, \boldsymbol{\lambda}) = \langle \mathbf{P}, \mathbf{C}\rangle + \langle \boldsymbol{\gamma}, (\log\xi + 1)\mathbf{1} - \mathrm{H_r}(\mathbf{P})\rangle + \langle \boldsymbol{\lambda}, \mathbf{1} - \mathbf{P}\mathbf{1}\rangle$ and $q(\boldsymbol{\gamma}, \boldsymbol{\lambda}) = \min_{\mathbf{P}\in\mathbb{R}_+^{n\times n}\cap\mathcal{S}} \mathcal{L}(\mathbf{P}, \boldsymbol{\gamma}, \boldsymbol{\lambda})$. Strong duality holds for (SEA). Moreover, let $\boldsymbol{\gamma}^\star, \boldsymbol{\lambda}^\star \in \operatorname{argmax}_{\boldsymbol{\gamma}\geq 0, \boldsymbol{\lambda}} q(\boldsymbol{\gamma}, \boldsymbol{\lambda})$ be the optimal dual variables respectively associated with the entropy and marginal constraints. Then, for at least $n-1$ indices $i \in [\![n]\!], \gamma^\star_i > 0$. When $\forall i \in [\![n]\!]$, $\gamma^\star_i > 0$ then $\mathrm{H_r}(\mathbf{P}^{\mathrm{se}}) = (\log\xi + 1)\mathbf{1}$ and $\mathbf{P}^{\mathrm{se}}$ has the form*

$$\mathbf{P}^{\mathrm{se}} = \exp\left((\boldsymbol{\lambda}^\star \oplus \boldsymbol{\lambda}^\star - 2\mathbf{C}) \oslash (\boldsymbol{\gamma}^\star \oplus \boldsymbol{\gamma}^\star)\right)\,. \tag{2}$$

The unicity of the solution in Proposition 4 is a consequence of the following lemma

Lemma 7. *Let $\mathbf{C} \neq 0 \in \mathcal{S}$ with zero diagonal. Then the problem $\min_{\mathbf{P}\in\mathcal{H}_\xi\cap\mathcal{S}} \langle \mathbf{P}, \mathbf{C}\rangle$ has a unique solution.*

Proof. Making the constraints explicit, the primal problem of symmetric entropic affinity takes the following form

$$\begin{aligned} \min_{\mathbf{P}\in\mathbb{R}_+^{n\times n}} \quad & \langle \mathbf{P}, \mathbf{C}\rangle \\ \text{s.t.} \quad & \forall i,\ \mathrm{H}(\mathbf{P}_{i:}) \geq \log \xi + 1 \\ & \mathbf{P}\mathbf{1} = \mathbf{1}, \quad \mathbf{P} = \mathbf{P}^\top\,. \end{aligned} \tag{SEA}$$

Suppose that the solution is not unique *i.e.*, there exists a couple of optimal solutions $(\mathbf{P}_1, \mathbf{P}_2)$ that satisfy the constraints of (SEA) and such that $\langle \mathbf{P}_1, \mathbf{C}\rangle = \langle \mathbf{P}_2, \mathbf{C}\rangle$. For $i \in [\![n]\!]$, we denote the function $f_i : \mathbf{P} \to (\log\xi + 1) - \mathrm{H}(\mathbf{P}_{i:})$. Then f_i is continuous, strictly convex and the entropy conditions of (SEA) can be written as $\forall i \in [\![n]\!], f_i(\mathbf{P}) \le 0$.

Now consider $\mathbf{Q} = \frac{1}{2}(\mathbf{P}_1 + \mathbf{P}_2)$. Then clearly $\mathbf{Q1} = \mathbf{1}, \mathbf{Q} = \mathbf{Q}^\top$. Since f_i is strictly convex we have $f_i(\mathbf{Q}) = f_i(\frac{1}{2}\mathbf{P}_1 + \frac{1}{2}\mathbf{P}_2) < \frac{1}{2}f_i(\mathbf{P}_1) + \frac{1}{2}f(\mathbf{P}_2) \le 0$. Thus $f_i(\mathbf{Q}) < 0$ for any $i \in [\![n]\!]$. Take any $\varepsilon > 0$ and $i \in [\![n]\!]$. By continuity of f_i there exists $\delta_i > 0$ such that, for any $\mathbf{H}$ with $\|\mathbf{H}\|_F \le \delta_i$, we have $f_i(\mathbf{Q}+\mathbf{H}) < f_i(\mathbf{Q}) + \varepsilon$. Take $\varepsilon > 0$ such that $\forall i \in [\![n]\!], 0 < \varepsilon < -\frac{1}{2}f_i(\mathbf{Q})$ (this is possible since for any $i \in [\![n]\!], f_i(\mathbf{Q}) < 0$) and $\mathbf{H}$ with $\|\mathbf{H}\|_F \le \min_{i\in[\![n]\!]} \delta_i$. Then for any $i \in [\![n]\!], f_i(\mathbf{Q}+\mathbf{H}) < 0$. In other words, we have proven that there exists $\eta > 0$ such that for any $\mathbf{H}$ such that $\|\mathbf{H}\|_F \le \eta$, it holds: $\forall i \in [\![n]\!], f_i(\mathbf{Q}+\mathbf{H}) < 0$.

Now let us take $\mathbf{H}$ as the Laplacian matrix associated to $\mathbf{C}$ *i.e.*, for any $(i,j) \in [\![n]\!]^2$, $H_{ij} = -C_{ij}$ if $i \neq j$ and $\sum_l C_{il}$ otherwise. Then we have $\langle \mathbf{H}, \mathbf{C}\rangle = -\sum_{i\neq j} C_{ij}^2 + 0 = -\sum_{i\neq j} C_{ij}^2 \neq 0$ since $\mathbf{C}$ has zero diagonal (and is nonzero). Moreover, $\mathbf{H} = \mathbf{H}^\top$ since $\mathbf{C}$ is symmetric and $\mathbf{H1} = \mathbf{0}$ by construction. Consider for $0 < \beta \le \frac{\eta}{\|\mathbf{H}\|_F}$, the matrix $\mathbf{H}_\beta(\mathbf{C}) := -\beta \operatorname{sign}(\langle \mathbf{H}, \mathbf{C}\rangle)\mathbf{H}$. Then $\|\mathbf{H}_\beta(\mathbf{C})\|_F = \beta\|\mathbf{H}\|_F \le \eta$. By the previous reasoning one has: $\forall i \in [\![n]\!], f_i(\mathbf{Q}+\mathbf{H}_\beta(\mathbf{C})) < 0$. Moreover, $(\mathbf{Q}+\mathbf{H}_\beta(\mathbf{C}))^\top = \mathbf{Q}+\mathbf{H}_\beta(\mathbf{C})$ and $(\mathbf{Q}+\mathbf{H}_\beta(\mathbf{C}))\mathbf{1} = \mathbf{1}$. For β small enough we have $\mathbf{Q}+\mathbf{H}_\beta(\mathbf{C}) \in \mathbb{R}_+^{n\times n}$ and thus there is a β (that depends on $\mathbf{P}_1$ and $\mathbf{P}_2$) such that $\mathbf{Q}+\mathbf{H}_\beta(\mathbf{C})$ is admissible *i.e.*, satisfies the constraints of (SEA). Then, for such β,

$$\begin{aligned}\langle \mathbf{C}, \mathbf{Q}+\mathbf{H}_\beta(\mathbf{C})\rangle - \langle \mathbf{C}, \mathbf{P}_1\rangle &= \frac{1}{2}\langle \mathbf{C}, \mathbf{P}_1 + \mathbf{P}_2\rangle + \langle \mathbf{C}, \mathbf{H}_\beta(\mathbf{C})\rangle - \langle \mathbf{C}, \mathbf{P}_1\rangle \\ &= \langle \mathbf{C}, \mathbf{H}_\beta(\mathbf{C})\rangle = -\beta \operatorname{sign}(\langle \mathbf{H}, \mathbf{C}\rangle)\langle \mathbf{H}, \mathbf{C}\rangle < 0\,.\end{aligned} \tag{16}$$

Thus $\langle \mathbf{C}, \mathbf{Q}+\mathbf{H}_\beta(\mathbf{C})\rangle < \langle \mathbf{C}, \mathbf{P}_1\rangle$ which leads to a contradiction. □

We can now prove the rest of the claims of Proposition 4 and Proposition 5.

Proof. Let $\mathbf{C} \in \mathcal{D}$. We first prove Proposition 4. The unicity is a consequence of Lemma 7. For the saturation of the entropies we consider the Lagrangian of the problem (SEA) that writes

$$\mathcal{L}(\mathbf{P}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\Gamma}) = \langle \mathbf{P}, \mathbf{C}\rangle + \langle \boldsymbol{\gamma}, (\log\xi + 1)\mathbf{1} - \mathrm{H}_r(\mathbf{P})\rangle + \langle \boldsymbol{\lambda}, \mathbf{1} - \mathbf{P1}\rangle + \langle \boldsymbol{\Gamma}, \mathbf{P} - \mathbf{P}^\top\rangle$$

for dual variables $\boldsymbol{\gamma} \in \mathbb{R}_+^n$, $\boldsymbol{\lambda} \in \mathbb{R}^n$ and $\boldsymbol{\Gamma} \in \mathbb{R}^{n\times n}$. Strong duality holds by Slater's conditions because $\frac{1}{n}\mathbf{1}\mathbf{1}^\top$ is stricly feasible for $\xi \le n-1$. Since strong duality holds, for any solution $\mathbf{P}^\star$ to the primal problem and any solution $(\boldsymbol{\gamma}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\Gamma}^\star)$ to the dual problem, the pair $\mathbf{P}^\star, (\boldsymbol{\gamma}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\Gamma}^\star)$ must satisfy the KKT conditions. They can be stated as follows:

$$\begin{aligned}&\mathbf{C} + \operatorname{diag}(\boldsymbol{\gamma}^\star)\log \mathbf{P}^\star - \boldsymbol{\lambda}^\star\mathbf{1}^\top + \boldsymbol{\Gamma}^\star - \boldsymbol{\Gamma}^{\star\top} = \mathbf{0}\\ &\mathbf{P}^\star\mathbf{1} = \mathbf{1},\ \mathrm{H}_r(\mathbf{P}^\star) \ge (\log\xi+1)\mathbf{1},\ \mathbf{P}^\star = \mathbf{P}^{\star\top}\\ &\boldsymbol{\gamma}^\star \ge \mathbf{0}\\ &\forall i, \gamma_i^\star(\mathrm{H}(\mathbf{P}_{i:}^\star) - (\log\xi + 1)) = 0\,.\end{aligned} \tag{KKT-SEA}$$

Let us denote $I = \{\ell \in [\![n]\!] \text{ s.t. } \gamma_\ell^\star = 0\}$. For $\ell \in I$, using the first-order condition, one has for $i \in [\![n]\!], C_{\ell i} = \lambda_\ell^\star - \Gamma_{\ell i}^\star + \Gamma_{i\ell}^\star$. Since $\mathbf{C} \in \mathcal{D}$, we have $C_{\ell\ell} = 0$ thus $\lambda_\ell^\star = 0$ and $C_{\ell i} = \Gamma_{i\ell}^\star - \Gamma_{\ell i}^\star$. For $(\ell, \ell') \in I^2$, one has $C_{\ell\ell'} = \Gamma_{\ell'\ell}^\star - \Gamma_{\ell\ell'}^\star = -(\Gamma_{\ell\ell'}^\star - \Gamma_{\ell'\ell}^\star) = -C_{\ell'\ell}$. $\mathbf{C}$ is symmetric thus $C_{\ell\ell'} = 0$. Since $\mathbf{C}$ only has null entries on the diagonal, this shows that $\ell = \ell'$ and therefore I has at most one element. By complementary slackness condition (last row of the KKT-SEA conditions) it holds that $\forall i \neq \ell, \mathrm{H}(\mathbf{P}_{i:}^\star) = \log\xi + 1$. Since the solution of (SEA) is unique $\mathbf{P}^\star = \mathbf{P}^{\mathrm{se}}$ and thus $\forall i \neq \ell, \mathrm{H}(\mathbf{P}_{i:}^{\mathrm{se}}) = \log\xi + 1$ which proves Proposition 4 but also that for at least $n-1$ indices $\gamma_i^\star > 0$. Moreover, from the KKT conditions we have

$$\forall (i,j) \in [\![n]\!]^2,\ \Gamma_{ji}^\star - \Gamma_{ij}^\star = C_{ij} + \gamma_i^\star \log P_{ij}^\star - \lambda_i^\star\,. \tag{17}$$

Now take $(i,j) \in [\![n]\!]^2$ fixed. From the previous equality $\Gamma_{ji}^\star - \Gamma_{ij}^\star = C_{ij} + \gamma_i^\star \log P_{ij}^\star - \lambda_i^\star$ but also $\Gamma_{ij}^\star - \Gamma_{ji}^\star = C_{ji} + \gamma_j^\star \log P_{ji}^\star - \lambda_j^\star$. Using that $\mathbf{P}^\star = (\mathbf{P}^\star)^\top$ and $\mathbf{C} \in \mathcal{S}$ we get $\Gamma_{ij}^\star - \Gamma_{ji}^\star = C_{ij} + \gamma_j^\star \log P_{ij}^\star - \lambda_j^\star$. But $\Gamma_{ij}^\star - \Gamma_{ji}^\star = -(\Gamma_{ji}^\star - \Gamma_{ij}^\star)$ which gives

$$C_{ij} + \gamma_j^\star \log P_{ij}^\star - \lambda_j^\star = -(C_{ij} + \gamma_i^\star \log P_{ij}^\star - \lambda_i^\star)\,. \tag{18}$$

This implies

$$\forall (i,j) \in [\![n]\!]^2,\ 2C_{ij} + (\gamma_i^\star + \gamma_j^\star) \log P_{ij}^\star - (\lambda_i^\star + \lambda_j^\star) = 0\,. \tag{19}$$

Consequently, if $\boldsymbol{\gamma}^\star > 0$ we have the desired form from the above equation and by complementary slackness $\mathrm{H_r}(\mathbf{P}^{\mathrm{se}}) = (\log \xi + 1)\mathbf{1}$ which proves Proposition 5. Note that otherwise, it holds

$$\forall (i,j) \neq (\ell,\ell),\ P_{ij}^\star = \exp\left(\frac{\lambda_i^\star + \lambda_j^\star - 2C_{ij}}{\gamma_i^\star + \gamma_j^\star}\right)\,. \tag{20}$$

□

A.5 EA and SEA as a KL projection

We prove the characterization as a projection of (EA) in Lemma 8 and of (SEA) in Lemma 9.

Lemma 8. *Let* $\mathbf{C} \in \mathcal{D}, \sigma > 0$ *and* $\mathbf{K}_\sigma = \exp(-\mathbf{C}/\sigma)$*. Then for any* $\sigma \leq \min_i \varepsilon_i^\star$*, it holds* $\mathbf{P}^{\mathrm{se}} = \mathrm{Proj}^{\mathrm{KL}}_{\mathcal{H}_\xi}(\mathbf{K}_\sigma) = \arg\min_{\mathbf{P} \in \mathcal{H}_\xi} \mathrm{KL}(\mathbf{P}|\mathbf{K}_\sigma)$.

Proof. The KL projection of $\mathbf{K}$ onto $\mathcal{H}_\xi$ reads

$$\min_{\mathbf{P} \in \mathbb{R}_+^{n\times n}} \quad \mathrm{KL}(\mathbf{P}|\mathbf{K}) \tag{21}$$

$$\text{s.t.} \quad \forall i,\ \mathrm{H}(\mathbf{P}_{i:}) \geq \log \xi + 1 \tag{22}$$

$$\mathbf{P}\mathbf{1} = \mathbf{1}\,. \tag{23}$$

Introducing the dual variables $\boldsymbol{\lambda} \in \mathbb{R}^n$ and $\boldsymbol{\kappa} \in \mathbb{R}_+^n$, the Lagrangian of this problem reads:

$$\mathcal{L}(\mathbf{P}, \boldsymbol{\lambda}, \boldsymbol{\kappa}) = \mathrm{KL}(\mathbf{P}|\mathbf{K}) + \langle \boldsymbol{\kappa}, (\log \xi + 1)\mathbf{1} - \mathrm{H}(\mathbf{P})\rangle + \langle \boldsymbol{\lambda}, \mathbf{1} - \mathbf{P}\mathbf{1}\rangle \tag{24}$$

Strong duality holds hence for any solution $\mathbf{P}^\star$ to the above primal problem and any solution $(\boldsymbol{\kappa}^\star, \boldsymbol{\lambda}^\star)$ to the dual problem, the pair $\mathbf{P}^\star, (\boldsymbol{\kappa}^\star, \boldsymbol{\lambda}^\star)$ must satisfy the KKT conditions. The first-order optimality condition gives:

$$\nabla_{\mathbf{P}} \mathcal{L}(\mathbf{P}^\star, \boldsymbol{\kappa}^\star, \boldsymbol{\lambda}^\star) = \log\left(\mathbf{P}^\star \oslash \mathbf{K}\right) + \mathrm{diag}(\boldsymbol{\kappa}^\star) \log \mathbf{P}^\star - \boldsymbol{\lambda}^\star \mathbf{1}^\top = \mathbf{0}\,. \tag{25}$$

Solving for $\boldsymbol{\lambda}^\star$ given the stochasticity constraint and isolating $\mathbf{P}^\star$ gives

$$\forall (i,j) \in [\![n]\!]^2, \quad P_{ij}^\star = \frac{\exp\left((\log K_{ij})/(1+\kappa_i^\star)\right)}{\sum_\ell \exp\left((\log K_{i\ell})/(1+\kappa_i^\star)\right)}\,. \tag{26}$$

We now consider $\mathbf{P}^\star$ as a function of $\boldsymbol{\kappa}$. Plugging this expression back in $\mathcal{L}$ yields the dual function $\boldsymbol{\kappa} \mapsto \mathcal{G}(\boldsymbol{\kappa})$. The latter is concave as any dual function and its gradient reads:

$$\nabla_{\boldsymbol{\kappa}} \mathcal{G}(\boldsymbol{\kappa}) = (\log \xi + 1)\mathbf{1} - \mathrm{H}(\mathbf{P}^\star(\boldsymbol{\kappa}))\,. \tag{27}$$

Denoting by $\boldsymbol{\rho} = \mathbf{1} + \boldsymbol{\kappa}$ and taking the dual feasibility constraint $\boldsymbol{\kappa} \geq \mathbf{0}$ into account gives the solution: for any i, $\rho_i^\star = \max(\varepsilon_i^\star, 1)$ where $\boldsymbol{\varepsilon}^\star$ solves (EA) with cost $\mathbf{C} = -\log \mathbf{K}$. Moreover we have that $\sigma \leq \min(\boldsymbol{\varepsilon}^\star)$ where $\boldsymbol{\varepsilon}^\star \in (\mathbb{R}_+^*)^n$ solves (EA). Therefore for any $i \in [\![n]\!]$, one has $\varepsilon_i^\star/\sigma \geq 1$. Thus there exists $\kappa_i^\star \in \mathbb{R}_+$ such that $\sigma(1+\kappa_i^\star) = \varepsilon_i^\star$.

This $\boldsymbol{\kappa}^\star$ cancels the above gradient *i.e.*, $(\log \xi + 1)\mathbf{1} = \mathrm{H}(\mathbf{P}^\star(\boldsymbol{\kappa}^\star))$ thus solves the dual problem. Therefore given the expression of $\mathbf{P}^\star$ we have that $\mathrm{Proj}^{\mathrm{KL}}_{\mathcal{H}_\xi}(\mathbf{K}) = \mathbf{P}^{\mathrm{e}}$. □

Lemma 9. *Let* $\mathbf{C} \in \mathcal{D}, \sigma > 0$ *and* $\mathbf{K}_\sigma = \exp(-\mathbf{C}/\sigma)$*. Suppose that the optimal dual variable* $\gamma^\star$ *associated with the entropy constraint of (SEA) is positive. Then for any* $\sigma \leq \min_i \gamma_i^\star$*, it holds* $\mathbf{P}^{\mathrm{se}} = \mathrm{Proj}^{\mathrm{KL}}_{\mathcal{H}_\xi \cap \mathcal{S}}(\mathbf{K}_\sigma)$.

Proof. Let $\sigma > 0$. The KL projection of $\mathbf{K}$ onto $\mathcal{H}_\xi \cap \mathcal{S}$ boils down to the following optimization problem.

$$\begin{aligned} \min_{\mathbf{P} \in \mathbb{R}_+^{n\times n}} \quad & \mathrm{KL}(\mathbf{P}|\mathbf{K}_\sigma) \\ \text{s.t.} \quad & \forall i,\ \mathrm{H}(\mathbf{P}_{i:}) \geq \log \xi + 1 \\ & \mathbf{P}\mathbf{1} = \mathbf{1}, \quad \mathbf{P}^\top = \mathbf{P}\,. \end{aligned} \tag{SEA-Proj}$$

By strong convexity of $\mathbf{P} \to \mathrm{KL}(\mathbf{P}|\mathbf{K}_\sigma)$ and convexity of the constraints the problem (SEA-Proj) admits a unique solution. Moreover, the Lagrangian of this problem takes the following form, where $\boldsymbol{\omega} \in \mathbb{R}_+^n$, $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\boldsymbol{\Gamma} \in \mathbb{R}^{n \times n}$:

$$\mathcal{L}(\mathbf{P}, \boldsymbol{\mu}, \boldsymbol{\omega}, \boldsymbol{\Gamma}) = \mathrm{KL}(\mathbf{P}|\mathbf{K}_\sigma) + \langle \boldsymbol{\omega}, (\log \xi + 1)\mathbf{1} - \mathrm{H}_r(\mathbf{P}) \rangle + \langle \boldsymbol{\mu}, \mathbf{1} - \mathbf{P}\mathbf{1} \rangle + \langle \boldsymbol{\beta}, \mathbf{P} - \mathbf{P}^\top \rangle \,.$$

Strong duality holds by Slater's conditions thus the KKT conditions are necessary and sufficient. In particular if $\mathbf{P}^\star$ and $(\boldsymbol{\omega}^\star, \boldsymbol{\mu}^\star, \boldsymbol{\beta}^\star)$ satisfy

$$\begin{aligned}
&\nabla_{\mathbf{P}} \mathcal{L}(\mathbf{P}^\star, \boldsymbol{\mu}^\star, \boldsymbol{\omega}^\star, \boldsymbol{\Gamma}^\star) = \log(\mathbf{P}^\star \oslash \mathbf{K}) + \mathrm{diag}(\boldsymbol{\omega}^\star) \log \mathbf{P}^\star - \boldsymbol{\mu}^\star \mathbf{1}^\top + \boldsymbol{\beta}^\star - \boldsymbol{\beta}^{\star\top} = \mathbf{0} \\
&\mathbf{P}^\star \mathbf{1} = \mathbf{1},\ \mathrm{H}_r(\mathbf{P}^\star) \geq (\log \xi + 1)\mathbf{1},\ \mathbf{P}^\star = \mathbf{P}^{\star\top} \\
&\boldsymbol{\omega}^\star \geq \mathbf{0} \\
&\forall i, \omega_i^\star (\mathrm{H}(\mathbf{P}_{i:}^\star) - (\log \xi + 1)) = 0 \,.
\end{aligned} \tag{KKT-Proj}$$

then $\mathbf{P}^\star$ is a solution to (SEA-Proj) and $(\boldsymbol{\omega}^\star, \boldsymbol{\mu}^\star, \boldsymbol{\beta}^\star)$ are optimal dual variables. The first condition rewrites

$$\forall (i,j),\ \log(P_{ij}^\star) + \frac{1}{\sigma} C_{ij} + \omega_i^\star \log(P_{ij}^\star) - \mu_i^\star + \beta_{ij}^\star - \beta_{ji}^\star = 0 \,, \tag{28}$$

which is equivalent to

$$\forall (i,j),\ \sigma(1 + \omega_i^\star) \log(P_{ij}^\star) + C_{ij} - \sigma \mu_i^\star + \sigma(\beta_{ij}^\star - \beta_{ji}^\star) = 0 \,. \tag{29}$$

Now take $\mathbf{P}^{\mathrm{se}}$ the optimal solution of (SEA). As written in the proof Proposition 5 of $\mathbf{P}^{\mathrm{se}}$ and the optimal dual variables $(\boldsymbol{\gamma}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\Gamma}^\star)$ satisfy the KKT conditions:

$$\begin{aligned}
&\forall (i,j),\ C_{ij} + \gamma_i^\star \log P_{ij}^{\mathrm{se}} - \lambda_i^\star + \Gamma_{ij}^\star - \Gamma_{ji}^\star = \mathbf{0} \\
&\mathbf{P}^{\mathrm{se}} \mathbf{1} = \mathbf{1},\ \mathrm{H}_r(\mathbf{P}^{\mathrm{se}}) \geq (\log \xi + 1)\mathbf{1},\ \mathbf{P}^{\mathrm{se}} = (\mathbf{P}^{\mathrm{se}})^\top \\
&\boldsymbol{\gamma}^\star \geq \mathbf{0} \\
&\forall i, \gamma_i^\star (\mathrm{H}(\mathbf{P}_{i:}^{\mathrm{se}}) - (\log \xi + 1)) = 0 \,.
\end{aligned} \tag{KKT-SEA}$$

By hypothesis $\boldsymbol{\gamma}^\star > 0$ which gives $\forall i, \mathrm{H}(\mathbf{P}_{i:}^{\mathrm{se}}) - (\log \xi + 1) = 0$. Now take $0 < \sigma \leq \min_i \gamma_i^\star$ and define $\forall i, \omega_i^\star = \frac{\gamma_i^\star}{\sigma} - 1$. Using the hypothesis on σ we have $\forall i, \omega_i^\star \geq 0$ and $\boldsymbol{\omega}^\star$ satisfies $\forall i,\ \sigma(1 + \omega_i^\star) = \gamma_i^\star$. Moreover for any $i \in [\![n]\!]$

$$\omega_i^\star (\mathrm{H}(\mathbf{P}_{i:}^{\mathrm{se}}) - (\log \xi + 1)) = 0 \,. \tag{30}$$

Define also $\forall i, \mu_i^\star = \lambda_i^\star / \sigma$ and $\forall (i,j), \beta_{ij}^\star = \Gamma_{ij}^\star / \sigma$. Since $\mathbf{P}^{\mathrm{se}}, (\boldsymbol{\gamma}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\Gamma}^\star)$ satisfies the KKT conditions (KKT-SEA) then by the previous reasoning $\mathbf{P}^{\mathrm{se}}, (\boldsymbol{\omega}^\star, \boldsymbol{\mu}^\star, \boldsymbol{\beta}^\star)$ satisfy the KKT conditions (KKT-Proj) and in particular $\mathbf{P}^{\mathrm{se}}$ is an optimal solution of (SEA-Proj) since KKT conditions are sufficient. Thus we have proven that $\mathbf{P}^{\mathrm{se}} \in \arg\min_{\mathbf{P} \in \mathcal{H}_\xi \cap \mathcal{S}} \mathrm{KL}(\mathbf{P}|\mathbf{K}_\sigma)$ and by the uniqueness of the solution this is in fact an equality. □

B Alternating Bregman Projections for Solving (SEA)

For $\sigma > 0$ and $\mathbf{K}_\sigma = \exp(-\mathbf{C}/\sigma)$, we introduce $\mathbf{P}_\sigma^{\mathrm{se}} = \mathrm{Proj}_{\mathcal{H}_\xi \cap \mathcal{S}}^{\mathrm{KL}}(\mathbf{K}_\sigma)$. Note that Lemma 9 gives us that when $\sigma \leq \min_i \gamma_i^\star$ ($\boldsymbol{\gamma}^\star$ is defined as the solution in $\boldsymbol{\gamma}$ of the Dual-SEA problem), we get $\mathbf{P}_\sigma^{\mathrm{se}} = \mathbf{P}^{\mathrm{se}}$. In Section 3.2, we have seen a dual ascent algorithm to compute $\mathbf{P}^{\mathrm{se}}$. We now provide an alternative computational approach to compute $\mathbf{P}_\sigma^{\mathrm{se}}$ for any σ. In particular, when $\sigma \leq \min_i \gamma_i^\star$, the presented approach provides an alternative to dual ascent for solving (SEA).

To compute $\mathbf{P}_\sigma^{\mathrm{se}}$, one can rely on the well-studied convergence of alternating Bregman projection methods [4]. The core idea is to alternate projection onto $\mathcal{H}_\xi$ with the projection onto $\mathcal{S}$. As $\mathcal{H}_\xi$ is not affine, one needs to resort to the Dykstra procedure [13] described in Algorithm 1. Note that Dykstra's algorithm can be applied to any Bregman divergence including KL [8] with guarantees [1].

Algorithm 1 *Dykstra* for computing $\mathbf{P}^{\mathrm{se}}_{\sigma}$

1: **Input**: cost $\mathbf{C}$, perplexity ξ, scaling σ
2: $(\mathbf{P}_s, \mathbf{\Xi}) \leftarrow \left(\exp(-\mathbf{C}/\sigma), \mathbf{1}\mathbf{1}^\top\right)$
3: **while** not converged **do**
4: $\quad \mathbf{P}_h \leftarrow \mathrm{Proj}^{\mathrm{KL}}_{\mathcal{H}_\xi}(\mathbf{P}_s \odot \mathbf{\Xi})$
5: $\quad \mathbf{\Xi} \leftarrow \mathbf{\Xi} \odot \mathbf{P}_s \oslash \mathbf{P}_h$
6: $\quad \mathbf{P}_s \leftarrow \mathrm{Proj}^{\mathrm{KL}}_{\mathcal{S}}(\mathbf{P}_h)$
7: **end while**
8: **Output:** $\mathbf{P}_s$

The projection-based strategy necessitates rescaling the data beforehand. This factor σ shrinks the data such that row-wise entropies are controlled when projecting onto $\mathcal{H}_\xi$. As such, choosing a σ too high might result in some entropies being unsaturated while a σ too small generally leads to slow convergence.

We now describe how to perform the two KL projection steps.

Projection onto $\mathcal{S}$. The KL projection onto $\mathcal{S}$ of $\mathbf{K} \in \mathbb{R}_+^{n\times n}$ amounts to the following problem.

$$\operatorname*{arg\,min}_{\mathbf{P}\in\mathcal{S}} \ \mathrm{KL}(\mathbf{P}|\mathbf{K}) \,. \tag{31}$$

For this problem the Lagrangian reads, where $\mathbf{W} \in \mathbb{R}^{n\times n}$ is a dual variable:

$$\mathcal{L}(\mathbf{P}, \mathbf{W}) = \mathrm{KL}(\mathbf{P}|\mathbf{K}) + \langle \mathbf{W}, \mathbf{P} - \mathbf{P}^\top \rangle \,. \tag{32}$$

Similarly as before, if we cancel the gradient of $\mathcal{L}$ with respect to $\mathbf{P}$ we obtain $\log(\mathbf{P}^\star \oslash \mathbf{K}) + \mathbf{W} - \mathbf{W}^\top = \mathbf{0}$. Thus $\mathbf{P}^\star = \exp(\mathbf{W} - \mathbf{W}^\top) \odot \mathbf{K}$. We must also have the primal feasibility that is $\mathbf{P}^\star = \mathbf{P}^{\star\top}$. Plugging the expression in this condition leads to $\mathbf{W} - \mathbf{W}^\top = \frac{1}{2}\log(\mathbf{K}^\top \oslash \mathbf{K})$. Hence plugging it back we get $\mathbf{P}^\star = \exp(\frac{1}{2}\log(\mathbf{K}^\top \oslash \mathbf{K})) \odot \mathbf{K} = \left(\mathbf{K}^\top \oslash \mathbf{K}\right)^{\odot\frac{1}{2}} \odot \mathbf{K} = \left(\mathbf{K} \odot \mathbf{K}^\top\right)^{\odot\frac{1}{2}}$. Overall the projection reads:

$$\operatorname*{arg\,min}_{\mathbf{P}\in\mathcal{S}} \ \mathrm{KL}(\mathbf{P}|\mathbf{K}) = \left(\mathbf{K} \odot \mathbf{K}^\top\right)^{\odot\frac{1}{2}} \,. \tag{33}$$

Projection onto $\mathcal{H}_\xi$. Concerning the entropic projection, on can compute $\mathrm{Proj}^{\mathrm{KL}}_{\mathcal{H}_\xi} : \mathcal{S} \to \mathcal{H}_\xi$ using a slight adaptation of Lemma 8. For any $\mathbf{P} \in \mathcal{S}$, it holds

$$\forall (i,j), \quad \mathrm{Proj}^{\mathrm{KL}}_{\mathcal{H}_\xi}(\mathbf{P})_{ij} = \frac{\exp\left(-\log P_{ij}/\rho_i\right)}{\sum_\ell \exp\left(-\log P_{i\ell}/\rho_i\right)} \tag{34}$$

where for any i, $\rho_i = \max(\varepsilon_i^\star, 1)$ where $\boldsymbol{\varepsilon}^\star$ solves (EA) with cost $\mathbf{C} = -\log \mathbf{K}$. Note that this projection is more efficient to compute than $\mathbf{P}^{\mathrm{e}}$ as one can stop the search when the upper bound on the root becomes smaller than one.