Let

- $\mu, \nu \in \Delta^n$ be the source and the target measures,
- $C \in \mathbb{R}_+^{n\times n}$ be the cost matrix,
- $\gamma > 0$ be the entropy regularization parameter,
- $H(P)$ denote the negative entropy for a plan $P \in \mathbb{R}^{n\times n}$, i.e., $H(P) := \sum_{i,j} P_{ij} \ln P_{ij}$.

Consider the OT problem:

$$\begin{aligned} \min_{P\in\mathbb{R}^{n\times n}} & \langle C, P\rangle + \gamma H(P) \\ \text{s.t. } & P \geq 0, \\ & P\mathbf{1} = \mu, \\ & P^\top \mathbf{1} = \nu. \end{aligned} \tag{1}$$

Proposition 0.1. *Let $\beta^* \in \mathbb{R}^n$ be the solution to the dual problem*

$$\max_{\beta} \; -\nu^\top \beta - \gamma \sum_i \mu_i \ln\left(\textstyle\sum_j \exp\left(\frac{-C_{ij}-\beta_j}{\gamma}\right)\right),$$

then the optimal plan P^* for (1) *can be recovered as follows*

$$p^*_{ij} = \mu_i \frac{\exp\left(\frac{-C_{ij}-\beta^*_j}{\gamma}\right)}{\sum_k \exp\left(\frac{-C_{ik}-\beta^*_k}{\gamma}\right)}.$$

Lagrangian:

$$\begin{aligned} L(P,\Lambda,\alpha,\beta) &= \langle C, P\rangle + \gamma H(P) - \langle \Lambda, P\rangle + \alpha^\top(P\mathbf{1}-\mu) + \beta^\top(P^\top \mathbf{1} - \nu) \\ &= -\mu^\top\alpha - \nu^\top\beta - \langle \Lambda - C - \alpha\mathbf{1}^\top - \mathbf{1}^\top\beta, P\rangle + \gamma H(P) \end{aligned}$$

Dual function:

$$\hat{g}(\Lambda,\alpha,\beta) = -\mu^\top\alpha - \nu^\top\beta - \underbrace{\sup_P \left[\langle \Lambda - C - \alpha\mathbf{1}^\top - \mathbf{1}^\top\beta, P\rangle - \gamma H(P)\right]}_{=:h(\Lambda,\alpha,\beta)}.$$

Note that

$$h(\Lambda,\alpha,\beta) = \gamma \sum_{i,j} \sup_{p\in\mathbb{R}} \left[\frac{\Lambda_{ij} - C_{ij} - \alpha_i - \beta_j}{\gamma} p - p\ln p\right].$$

The supremum is attained at a point

$$p^*_{ij} = \exp\left(\frac{\Lambda_{ij} - C_{ij} - \alpha_i - \beta_j}{\gamma} - 1\right), \tag{2}$$

and it holds

$$h(\Lambda,\alpha,\beta) = \gamma \sum_{i,j} \exp\left(\frac{\Lambda_{ij} - C_{ij} - \alpha_i - \beta_j}{\gamma} - 1\right).$$

Thus, we have

$$\hat{g}(\Lambda,\alpha,\beta) = -\mu^\top\alpha - \nu^\top\beta - \gamma\sum_{i,j} \exp\left(\frac{\Lambda_{ij} - C_{ij} - \alpha_i - \beta_j}{\gamma} - 1\right).$$

The dual problem reads as

$$\max_{\Lambda\geq 0,\alpha,\beta} \hat{g}(\Lambda,\alpha,\beta).$$

Observe that for any pair α,β, the function $\hat{g}(\cdot,\alpha,\beta)$ achieves its maximum at $\Lambda = 0$. We therefore consider the function $\tilde{g}(\alpha,\beta) := \hat{g}(0,\alpha,\beta)$. It holds

$$\frac{\partial \tilde{g}}{\partial \alpha_i} = -\mu_i + e^{-\alpha_i/\gamma} \sum_j \exp\left(\frac{-C_{ij} - \beta_j}{\gamma} - 1\right),$$

$$\begin{aligned} \frac{\partial \tilde{g}}{\partial \alpha_i} = 0 \iff e^{-\alpha_i^*(\beta)/\gamma} &= \frac{\mu_i}{\sum_j \exp\left(\frac{-C_{ij}-\beta_j}{\gamma} - 1\right)} \\ \iff \alpha_i^*(\beta) &= \gamma\ln\left(\textstyle\sum_j \exp\left(\frac{-C_{ij}-\beta_j}{\gamma} - 1\right)\right) - \gamma\ln\mu_i \\ &= \gamma\ln\left(\textstyle\sum_j \exp\left(\frac{-C_{ij}-\beta_j}{\gamma}\right)\right) - \gamma - \gamma\ln\mu_i. \end{aligned} \tag{3}$$

Thus, we reduce the problem to maximization of the function

$$\begin{aligned} g(\beta) := \tilde{g}(\alpha^*(\beta),\beta) = &-\gamma\sum_i \mu_i \ln\left(\textstyle\sum_j \exp\left(\frac{-C_{ij}-\beta_j}{\gamma}\right)\right) + \gamma + \gamma H(\mu) - \nu^\top\beta \\ &- \gamma\sum_i e^{-\alpha_i^*(\beta)/\gamma} \sum_j \exp\left(\frac{-C_{ij}-\beta_j}{\gamma} - 1\right) \\ = &-\gamma\sum_i \mu_i \ln\left(\textstyle\sum_j \exp\left(\frac{-C_{ij}-\beta_j}{\gamma}\right)\right) + \gamma H(\mu) - \nu^\top\beta. \end{aligned}$$

Thus, we arrive at the dual problem

$$\max_{\beta} \; -\nu^\top \beta - \gamma \sum_i \mu_i \ln \left(\textstyle\sum_j \exp\left(\frac{-C_{ij} - \beta_j}{\gamma} \right) \right).$$

If β^* is the optimal dual variable, we can recover the optimal plan using (2) and (3):

$$p^*_{ij} = e^{-\alpha_i^*(\beta)/\gamma} \exp\left(\frac{-C_{ij} - \beta_j}{\gamma} - 1 \right) = \mu_i \frac{\exp\left(\frac{-C_{ij} - \beta_j}{\gamma} \right)}{\sum_k \exp\left(\frac{-C_{ik} - \beta_k}{\gamma} \right)}.$$