Target point y^*

$$P_z^{t+1} = \min_y \left\{ c_{zy} + P_y^t, P_z^t \right\}$$

initialize with

$$\begin{aligned} P_z^0 &= 0 \text{ if } z = y^* \\ &= +\infty \text{ else} \end{aligned}$$

1 Regularized optimal transport

Assume $\sum_x n_x = \sum_y m_y$.

Start from the unregularized case

$$v_y = \max_x \{\Phi_{xy} - u_x\}$$

We had

$$\sum_y \mu_{xy} = n_x \text{ and } \sum_x \mu_{xy} = m_y$$

$$v_y = \max_x \{\Phi_{xy} - u_x\}$$

$$\mu_{xy} > 0 \implies x \in \arg\max_x \{\Phi_{xy} - u_x\}.$$

This problem could be interpreted at a linear programming problem

$$\begin{aligned} \max_{\mu \geq 0} \quad & \sum_{xy} \mu_{xy} \Phi_{xy} \\ s.t. \quad & \sum_y \mu_{xy} = n_x \text{ and } \sum_x \mu_{xy} = m_y \end{aligned}$$

with dual

$$\begin{aligned} \min_{u,v} \quad & \sum_x n_x u_x + \sum_y m_y v_y \\ s.t. \quad & u_x + v_y \geq \Phi_{xy} \end{aligned}$$

Now assume that instead of producing Φ_{xy} with firm y, a worker x produces $\Phi_{xy} + \sigma\eta_x$ where $(\eta_x)_x$ is a random vector and $\sigma > 0$ is a scale parameter. We will assume that η is i.i.d. Gumbel. The indirect utility of a firm y who draws shock η is then

$$\max_x \{\Phi_{xy} - u_x + \sigma\eta_x\}$$

this is random. The expected indirect utility of a firm y is

$$v_y = E\left[\max_x \{\Phi_{xy} - u_x + \sigma\eta_x\}\right] = \sigma \log \sum_x \exp\left(\frac{\Phi_{xy} - u_x}{\sigma}\right)$$

THe probability that y chooses x is

$$\frac{\mu_{xy}}{m_y} = \frac{\exp\left(\frac{\Phi_{xy}-u_x}{\sigma}\right)}{\sum_{x'}\exp\left(\frac{\Phi_{x'y}-u_{x'}}{\sigma}\right)} = \frac{\exp\left(\frac{\Phi_{xy}-u_x}{\sigma}\right)}{\exp\left(\frac{v_y}{\sigma}\right)} = \exp\left(\frac{\Phi_{xy}-u_x-v_y}{\sigma}\right)$$

thus

$$\mu_{xy} = m_y \exp\left(\frac{\Phi_{xy}-u_x-v_y}{\sigma}\right)$$

and the constraints on μ_{xy} are

$$\begin{aligned} \sum_y \mu_{xy} &= n_x \\ \sum_x \mu_{xy} &= m_y \end{aligned}$$

that is

$$\begin{aligned} \sum_y m_y \exp\left(\frac{\Phi_{xy}-u_x-v_y}{\sigma}\right) &= n_x \\ \sum_x m_y \exp\left(\frac{\Phi_{xy}-u_x-v_y}{\sigma}\right) &= m_y \end{aligned}$$

Define $a_x = u_x$ and $b_y = v_y - \sigma \ln m_y$ and rewrite the system of equations as

$$\begin{aligned} \sum_y \exp\left(\frac{\Phi_{xy}-a_x-b_y}{\sigma}\right) &= n_x \\ \sum_x \exp\left(\frac{\Phi_{xy}-a_x-b_y}{\sigma}\right) &= m_y \end{aligned}$$

Question: can we view these equations as the first order conditions of an optimization problem, ie.

$$\min_{(a_x),(b_y)} F(a,b)$$

We need to have

$$\begin{aligned} \frac{\partial F}{\partial a_x} &= n_x - \sum_y \exp\left(\frac{\Phi_{xy}-a_x-b_y}{\sigma}\right) \\ \frac{\partial F}{\partial b_y} &= m_y - \sum_x \exp\left(\frac{\Phi_{xy}-a_x-b_y}{\sigma}\right) \end{aligned}$$

$$F(a,b) = \sum_x n_x a_x + \sum_y m_y b_y + \sigma \sum_{xy} \exp\left(\frac{\Phi_{xy}-a_x-b_y}{\sigma}\right)$$

Claim: this problem is equivalent with

$$\begin{aligned}
\max_{\mu\geq 0} \quad & \sum_{xy}\mu_{xy}\Phi_{xy}-\sigma\sum_{xy}\mu_{xy}\ln\mu_{xy} \\
s.t. \quad & \sum_{y}\mu_{xy}=n_x\left[\tilde{a}_x\right] \\
& \sum_{x}\mu_{xy}=m_y\left[\tilde{b}_y\right]
\end{aligned}$$

Indeed, by the Karusch-Kunh-Tucker conditions in the latter problem

$$\Phi_{xy}-\sigma\left(1+\ln\mu_{xy}\right)-\tilde{a}_x-\tilde{b}_y=0$$

that is

$$\mu_{xy}=\exp\left(\frac{\Phi_{xy}-\tilde{a}_x-\tilde{b}_y-\sigma}{\sigma}\right)$$

so set $a_x=\tilde{a}_x$ and $b_y=\tilde{b}_y+\sigma$ and we have

$$\mu_{xy}=\exp\left(\frac{\Phi_{xy}-a_x-b_y}{\sigma}\right)$$

where a_x and b_y are determined by

$$\begin{aligned}
\sum_{y}\exp\left(\frac{\Phi_{xy}-a_x-b_y}{\sigma}\right) &= n_x \\
\sum_{x}\exp\left(\frac{\Phi_{xy}-a_x-b_y}{\sigma}\right) &= m_y
\end{aligned}$$

If σ is very large, then the solution solves

$$\begin{aligned}
\min \quad & \sum_{xy}\mu_{xy}\ln\mu_{xy} \\
s.t. \quad & \sum_{y}\mu_{xy}=n_x \\
& \sum_{x}\mu_{xy}=m_y
\end{aligned}$$

Whose solution is $\mu_{xy}=n_x m_y$. Berstein-Schrodinger systems.

Computation. We need to minimize

$$F\left(a,b\right)=\sum_{x}n_x a_x+\sum_{y}m_y b_y+\sigma\sum_{xy}\exp\left(\frac{\Phi_{xy}-a_x-b_y}{\sigma}\right)$$

Gradient descent:

$$
\begin{aligned}
a_x^{t+1} &= a_x^t - \epsilon \frac{\partial F(a,b)}{\partial a_x} \\
b_y^{t+1} &= b_y^t - \epsilon \frac{\partial F(a,b)}{\partial b_x}.
\end{aligned}
$$

Coordinate descent:
Take an initial guess of b^1.
Set a^1 in order to minimize $\min_a F\left(a, b^1\right)$.
Set b^2 in order to minimize $\min_b F\left(a^1, b\right)$
Set a^2 in order to minimize $\min_a F\left(a, b^2\right)$.

Consider $\min_a F(a, b^t)$. THat leads to setting a^t such that

$$
\frac{\partial F}{\partial a_x}\left(a^t, b^t\right) = 0
$$

ie

$$
n_x - \sum_y \exp\left(\frac{\Phi_{xy} - a_x^t - b_y^t}{\sigma}\right) = 0
$$

but we have

$$
n_x = e^{-a_x^t/\sigma} \sum_y \exp\left(\frac{\Phi_{xy} - b_y^t}{\sigma}\right)
$$

and thus

$$
\exp\left(a_x^t/\sigma\right) = \frac{\sum_y \exp\left(\frac{\Phi_{xy} - b_y^t}{\sigma}\right)}{n_x}
$$

and similarly,

$$
\exp\left(b_y^{t+1}/\sigma\right) = \frac{\sum_x \exp\left(\frac{\Phi_{xy} - a_x^t}{\sigma}\right)}{m_y}.
$$

This is the IPFP / Sinkhorn / matrix scaling. etc..