1 Proof of Proposition 2

Proof. Recall problem ($\mathcal{P}_1$):

$$\begin{aligned} &\min_{\mathbf{P}} \langle \mathbf{D}, \mathbf{P} \rangle + \epsilon \sum_{i,j} P_{i,j} \log P_{i,j} \\ &s.t. P_{i,j} \geq 0,\ \sum_j P_{i,j} = 1,\ \frac{1}{m} \sum_i P_{i,j} = \frac{1}{n} \end{aligned} \tag{$\mathcal{P}_1$}$$

We manipulate ($\mathcal{P}_1$) using method of Lagrangian multipliers. Introducing multipliers $\alpha_i, i = 1, \ldots, m$ and $\beta_j, j = 1, \ldots, n$. Define

$$\mathcal{L} \triangleq \langle \mathbf{D}, \mathbf{P} \rangle + \epsilon \sum_{i,j} P_{i,j} \log P_{i,j} - \sum_i \alpha_i \left(\sum_j P_{i,j} - 1 \right) - \sum_j \beta_j \left(\sum_i P_{i,j} - \frac{m}{n} \right)$$

The optimizer of ($\mathcal{P}_1$) is the optimizer of the following min-max problem,

$$\min_{\mathbf{P}: P_{i,j} \geq 0} \max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \mathcal{L}(\mathbf{P}, \boldsymbol{\alpha}, \boldsymbol{\beta}),$$

The minimizer $\mathbf{P}^*$ can be obtained by setting

$$\frac{\partial \mathcal{L}}{\partial P_{i,j}} = D_{i,j} + \epsilon \left(\log P_{i,j} + 1 \right) - \alpha_i - \beta_j = 0,$$

which gives the minimizer $\mathbf{P}^*$

$$P_{i,j}^* = \exp \left(-\frac{D_{i,j} - \alpha_i - \beta_j}{\epsilon} - 1 \right).$$

Substitute the $P_{i,j}^*$'s into $\mathcal{L}$, simplify, gives

$$\mathcal{L}(\mathbf{P}^*, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_i \alpha_i + \frac{m}{n} \sum_j \beta_j - \epsilon \sum_{i,j} \exp \left(-\frac{D_{i,j} - \alpha_i - \beta_j}{\epsilon} - 1 \right)$$

We now study

$$\max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \mathcal{L}(\mathbf{P}^*, \boldsymbol{\alpha}, \boldsymbol{\beta}).$$

$\mathcal{L}(\mathbf{P}^*, \boldsymbol{\alpha}, \boldsymbol{\beta})$ is a concave function of $[\boldsymbol{\alpha}, \boldsymbol{\beta}]$, whose maximizer can be obtained by setting its derivative to zero. We therefore let

$$\frac{\partial \mathcal{L}(\mathbf{P}^*, \boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \alpha_i} = 1 - \sum_j \exp \left(-\frac{D_{i,j} - \alpha_i - \beta_j}{\epsilon} - 1 \right) = 0,$$

which gives the optimal α_i^*,

$$\alpha_i^* = -\epsilon \log \sum_j \exp \left(-\frac{D_{i,j} - \beta_j}{\epsilon} - 1 \right)$$

Resubstitute into $\mathcal{L}(\mathbf{P}^*, \boldsymbol{\alpha}, \boldsymbol{\beta})$, simplify, and we arrive at

$$\mathcal{L}(\mathbf{P}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}) = \sum_i \left[-\epsilon \log \sum_j \exp \left(-\frac{D_{i,j} - \beta_j}{\epsilon} \right) + \frac{1}{n} \sum_j \beta_j \right].$$

We want to solve $\max_{\boldsymbol{\beta}} \mathcal{L}(\mathbf{P}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta})$, which is equivalent to

$$\min_{\boldsymbol{\beta}} \sum_i \left[\epsilon \log \sum_j \exp\left(\frac{\beta_j - D_{i,j}}{\epsilon} \right) - \frac{1}{n} \sum_j \beta_j \right].$$

Also, substituting the $\boldsymbol{\alpha}^*$ into $\mathbf{P}^*$ gives

$$\mathbf{P}^* = \frac{\exp\left(\frac{\beta_j - D_{i,j}}{\epsilon}\right)}{\sum_j \exp\left(\frac{\beta_j - D_{i,j}}{\epsilon}\right)}.$$

□

2 Bilingual Lexicon Induction Results Using vocabularies of 200K

This section supplements Section 6.2 of the paper.

Table 1: P@1 values on the large test dictionary. Source and target vocabularies are both 200K

source \ target		en	es	fr	it	pt	de
en	NN		60.62	61.66	52.89	42.19	58.37
	ISF		75.00	76.20	68.69	58.32	70.74
	CSLS		**75.18**	76.35	69.08	**58.81**	**71.06**
	GNN		75.17	**76.95**	**69.16**	58.64	69.69
es	NN	65.14		67.21	68.17	72.17	54.56
	ISF	76.98		80.61	79.88	82.14	67.53
	CSLS	76.94		80.36	79.87	82.95	**67.67**
	GNN	**77.82**		**81.49**	**80.83**	**83.72**	66.62
fr	NN	66.86	67.70		65.92	52.12	62.62
	ISF	78.34	79.92		77.72	65.71	74.37
	CSLS	78.49	80.30		78.07	66.62	**74.75**
	GNN	**79.16**	**80.59**		**78.44**	**67.02**	73.31
it	NN	57.10	70.50	67.79		58.38	57.06
	ISF	70.34	81.80	80.80		72.35	69.75
	CSLS	70.05	81.93	80.57		73.12	**69.91**
	GNN	**71.37**	**82.87**	**81.80**		**74.10**	68.85
pt	NN	47.81	75.31	54.36	58.99		44.69
	ISF	60.93	85.54	69.58	73.37		58.07
	CSLS	60.60	85.66	69.14	73.25		**58.32**
	GNN	**62.49**	**87.13**	**70.90**	**74.90**		57.08
de	NN	60.61	50.05	59.51	52.93	39.17	
	ISF	**72.97**	**64.30**	**74.98**	**69.91**	**57.39**	
	CSLS	72.21	63.60	73.96	68.73	55.97	
	GNN	72.33	64.03	73.52	69.62	56.58	