Low-Rank Optimal Transport

Yessin Moakher

Jan 2025

Method (Low-Rank Optimal Transport through Factor Relaxation with Latent Coupling, Halmos and al., Neurips 2024)

Notations :

$$\Pi_{a,\cdot} := \{\mathbf{P} \mid \mathbf{P}1_m = a\}, \quad \Pi_{\cdot,b} := \{\mathbf{P} \mid \mathbf{P}^T 1_n = b\}, \quad \Pi_{a,b} := \Pi_{a,\cdot} \cap \Pi_{\cdot,b}.$$

The problem :

$$\begin{aligned}
&\mathbf{P} = \arg\min_{\mathbf{P}} \langle \mathbf{P}, \mathbf{M} \rangle_F \\
&\text{s.t.} \quad \mathbf{P} = \mathbf{Q}\,\mathrm{diag}\left(\frac{1}{g_Q}\right)\mathbf{T}\,\mathrm{diag}\left(\frac{1}{g_R}\right)\mathbf{R}^T, \\
&g_Q := \mathbf{Q}^T 1_n, \quad g_R := \mathbf{R}^T 1_m, \\
&\mathbf{Q} \in \Pi_{a,\cdot}, \quad \mathbf{R} \in \Pi_{b,\cdot}, \quad \mathbf{T} \in \Pi_{g_Q,g_R}, \\
&\mathbf{Q} \in \mathbb{R}^+_{n,r}, \quad \mathbf{R} \in \mathbb{R}^+_{m,r}, \quad \mathbf{T} \in \mathbb{R}^+_{r,r},
\end{aligned}$$

General optimization framework :

$$\min_{\mathbf{w_1},\mathbf{w_2}} f(w_1, w_2)$$

- For multivariable optimization : Coordinate Descent algorithm : f convex, differentiable :

$$w_1^{(k+1)} \leftarrow \arg\min_{w} f(w, w_2^{(k)})$$

$$w_2^{(k+1)} \leftarrow \arg\min_{w} f(w_1^{(k+1)}, w)$$

For each optimization problem : Gradient descent algorithm :

$$W_{k+1} \leftarrow \arg\min_{W} \langle W, g \rangle + \frac{1}{\gamma}\|W - W_k\|_2.$$

Generalization of gradient descent : Mirror Descent :

$$W_{k+1} \leftarrow \arg\min_{W} \langle W, g \rangle + \frac{1}{\gamma} D(W, W_k).$$

where D is a Bregman divergence.

In particular, here to take into account the "geometry" of the problem (we optimize over the space of probability distributions), we will use the Kullback-Leiber(KL) divergence between probability distributions.
We do the coordinate Descent on the variables (Q,R) and T :

$$(Q_{k+1}, R_{k+1}) \leftarrow \arg \min_{\mathbf{Q} \in \Pi_{a,\cdot}, \mathbf{R} \in \Pi_{b,\cdot}} \langle (Q, R), \nabla_{Q,R} \mathcal{L}_{\mathrm{LC}} \rangle + \frac{1}{\gamma_k} \mathrm{KL}((Q, R) \parallel (Q_k, R_k)).$$

$$T_{k+1} \leftarrow \arg \min_{T \in \Pi_{g_{Q_{k+1}}, g_{R_{k+1}}}} \langle T, \nabla_T \mathcal{L}_{\mathrm{LC}} \rangle + \frac{1}{\gamma_k} \mathrm{KL}(T \parallel T_k).$$

Now we,