Structured OT
Optimization Sub-Routines

Subgradients of f.

$$\partial f(\Gamma) = \arg\max_{\kappa \in \mathcal{B}_F} \langle \Gamma, \kappa \rangle$$

i.e. linear optimization over the base polytope

solve by Edmond's greedy algorithm in $O(N \log N)$

Projections onto transport polytope

Mirror map: $\Phi_\Pi(\Gamma) = \sum_{ij} \Gamma_{ij} \log(\Gamma_{ij})$

Projection Step: $P_{\Pi(\mathbf{a},\mathbf{b})}(w) = \arg \min_{\Gamma \in \Pi(\mathbf{a},\mathbf{b})} \mathrm{KL}(\Gamma \| w)$

Solve by Sinkhorn-Knopp

Projections onto base polytope

Mirror map: $\Phi_{\mathcal{B}_F}(\kappa) = \frac{1}{2}\|k\|_2^2$

Projection Step: $P_{\mathcal{B}_F}(w) = \arg \min_{\Gamma \in \Pi(\mathbf{a},\mathbf{b})} \|\kappa - w\|_2^2$

Solve by Fujishige Min-Norm-Point Algorithm $O(n^6)$

For our decomposable fun's, can solve in $O(n \log n)$

Structured OT
Algorithm Comparison

- Convex formulation: Mirror Descent (MDA)
- Min max formulation:
  - Saddle Point Mirror Descent (SPMD)
  - Saddle Point Mirror Prox (SPMP)
- Theoretically: cost-per-iteration vs convergence rate tradeoff
- Empirically: comparable performance SPMP better in large scale regime