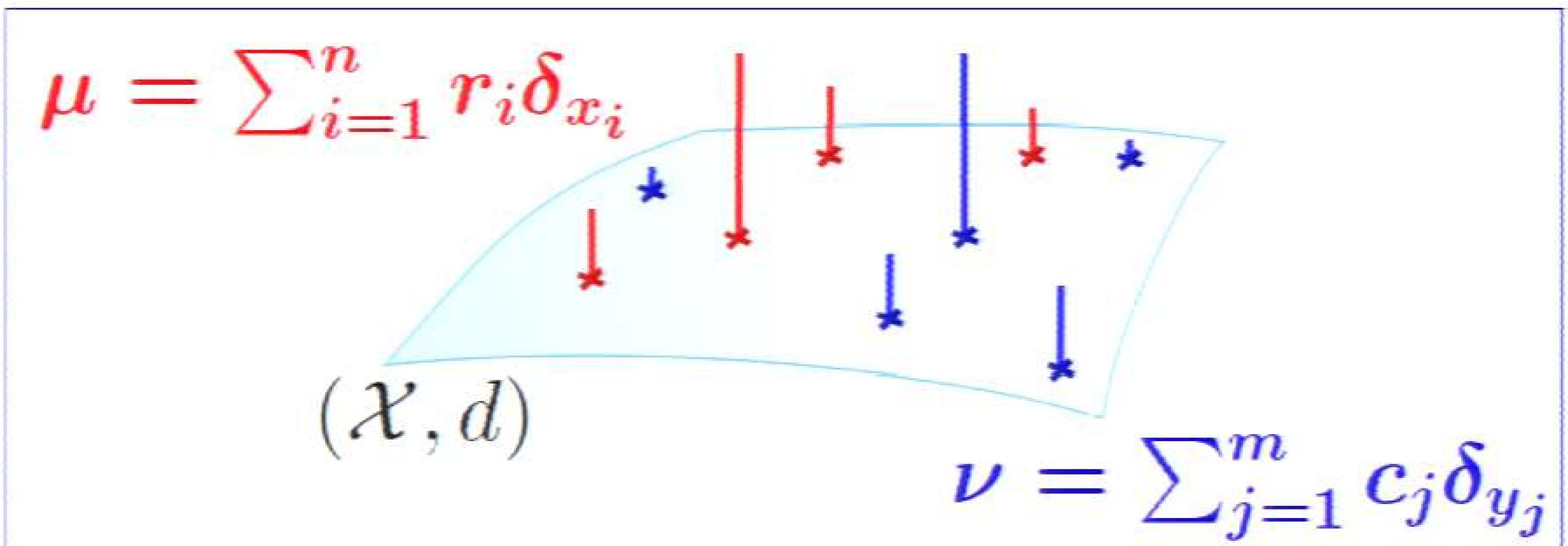
Sinkhorn Distances: Lightspeed Computation of Optimal Transport

very very fast Earth Mover's on GPGPU's

Marco Cuturi (Kyoto University) - Poster 89

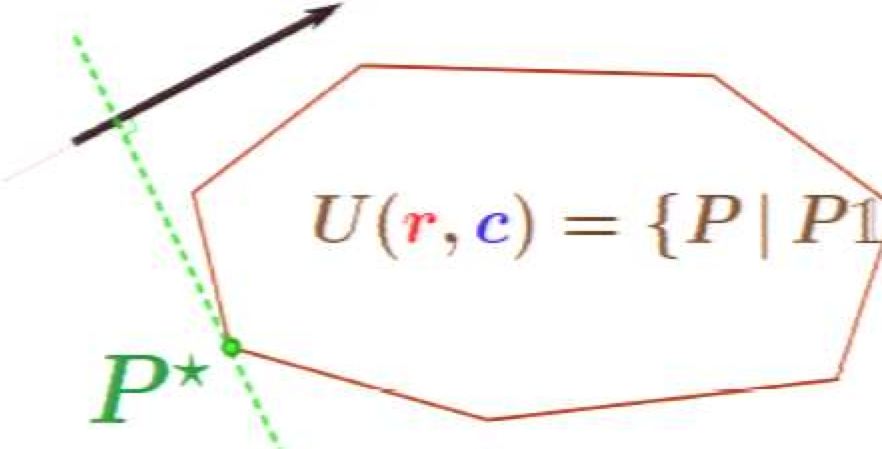
Comparing probability measures is crucial in ML



Optimal Transport, Wasserstein, EMD ...

+ Arguably the most natural distance for probabilities

— Hard to compute. Linear Program, $O(n^3 \log(n))$



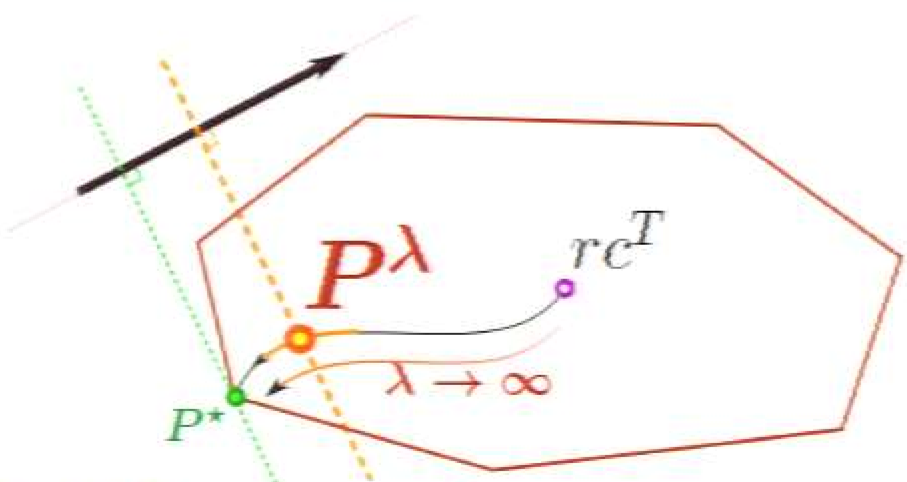
$$\mathrm{EMD}(\boldsymbol{\mu}, \boldsymbol{\nu}) = \min_{P \in U(r,c)} \langle P, M \rangle$$

$\mathbb{R}_+^{n \times m}$

Regularize OT, Use Sinkhorn's Algorithm

λ : add an entropic regularization in $h(P)$

P^{λ} can be factorized: $\mathbf{diag}(u)e^{-\lambda M}\,\mathbf{diag}(v)$



$$\mathbf{EMD}_{\lambda}(\mu, \nu) = \min_{P \in U(r,c)} \langle P, M \rangle - \frac{1}{\lambda} h(P)$$

$\mathbb{R}_{+}^{n \times m}$

Significance of this work

Better performance in preliminary experiments

- EMD_λ works better than EMD using SVM's

Faster, Parallel, seamless GPGPU execution

- $O(n^2)$, **Vectorized**, only uses matrix products
- Also provides accurate **upper/lower** bound on EMD
- compute *one* distance between histograms, $n = 4096$
  - 6,900 s (≈ 2 hours) for EMD using `FastEMD`,
  - 0.2 s for EMD_λ single core, 0.03 s on GPGPU