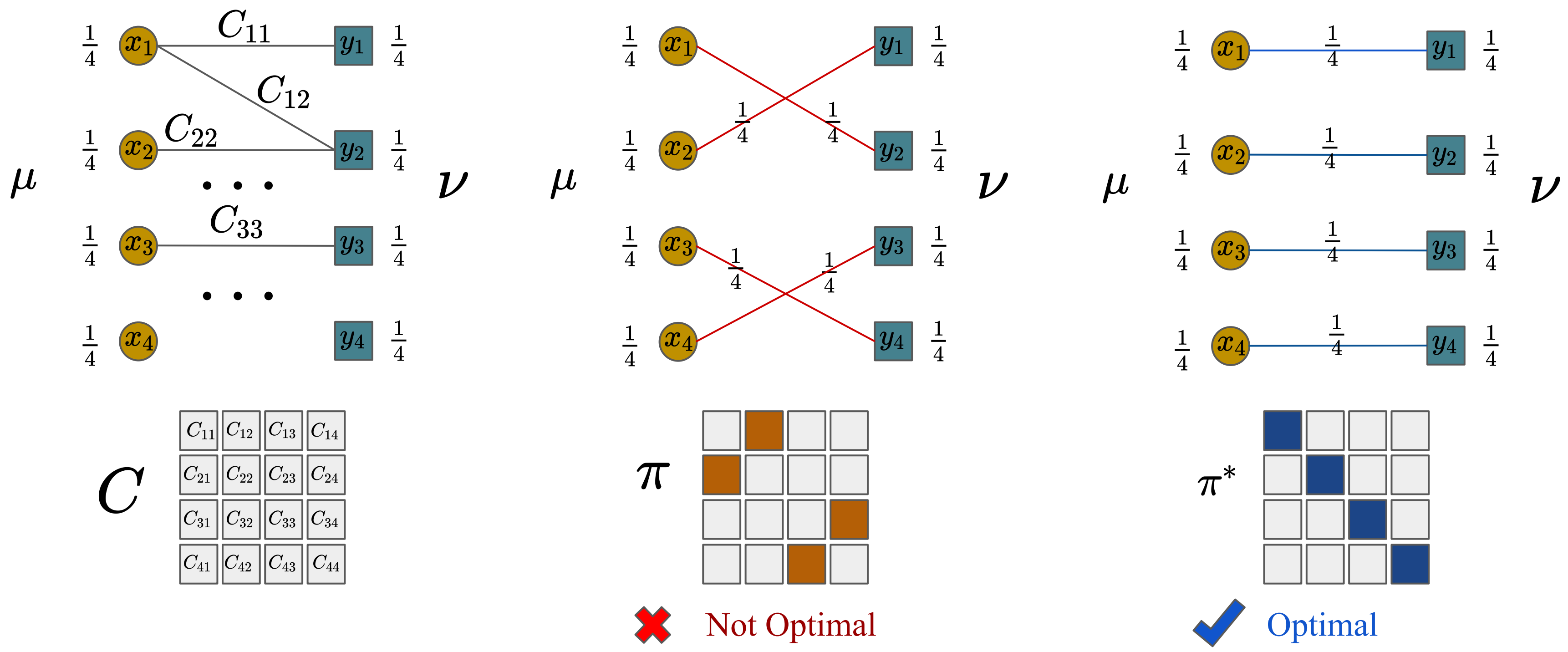
On Transportation of Mini-batches: A Hierarchical Approach

Khai Nguyen[1], Dang Nguyen[2], Quoc Nguyen[2], Tung Pham[2], Hung Bui[2], Dinh Phung[3], Trung Le[3], Nhat Ho[1]

[1]Department of Statistics and Data Sciences, University of Texas at Austin
[2]VinAI Research
[3]Monash University

Optimal Transport



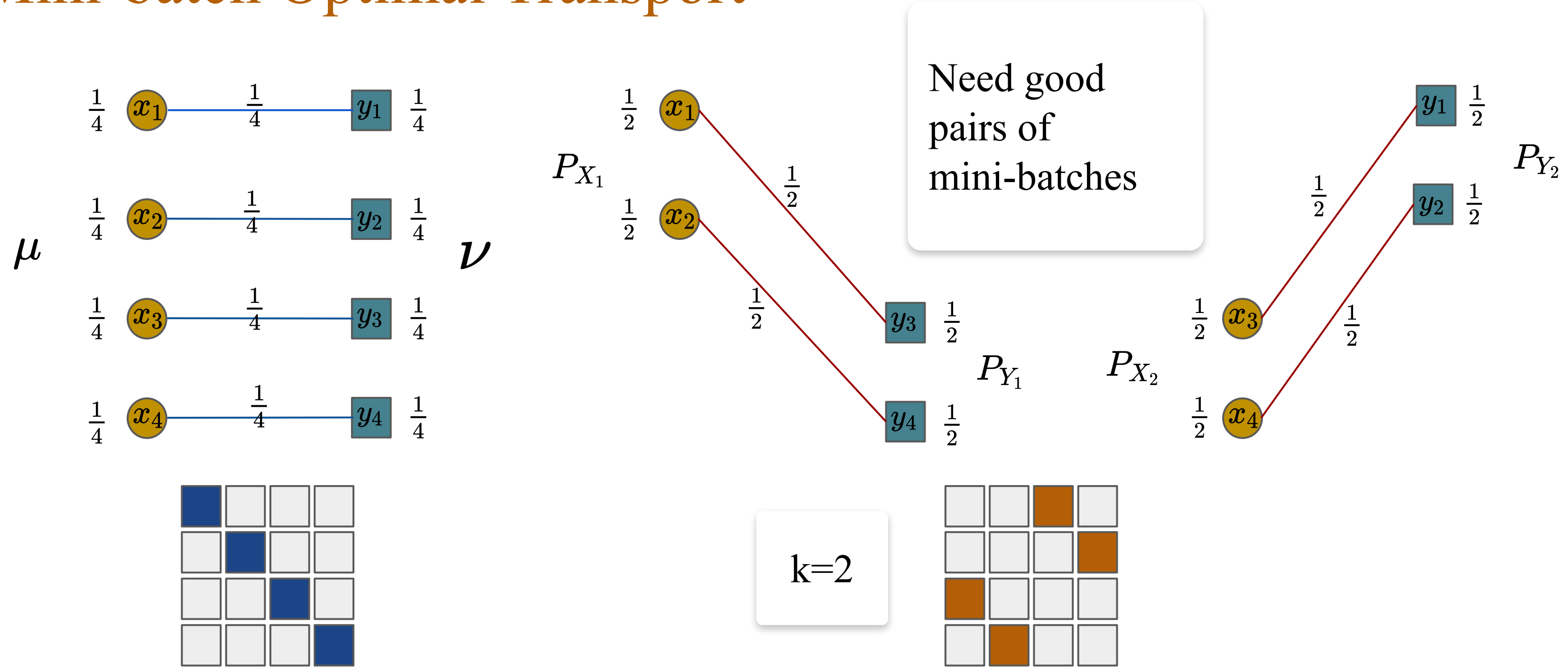
Mini-batch Optimal Transport

The number of supports is large? e.g., millions

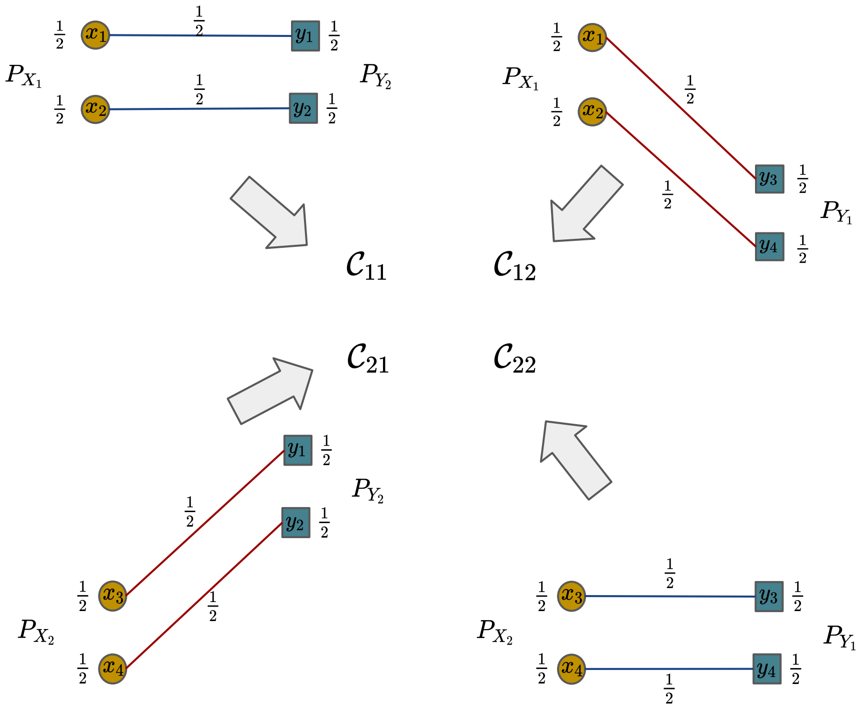
Repeated computation? e.g., deep learning

- ❑ Impossible to store the cost matrix C in the computational graph
- ❑ Slow computation of OT losses which leads to slow training

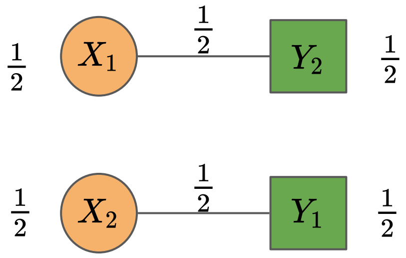
Mini-batch Optimal Transport



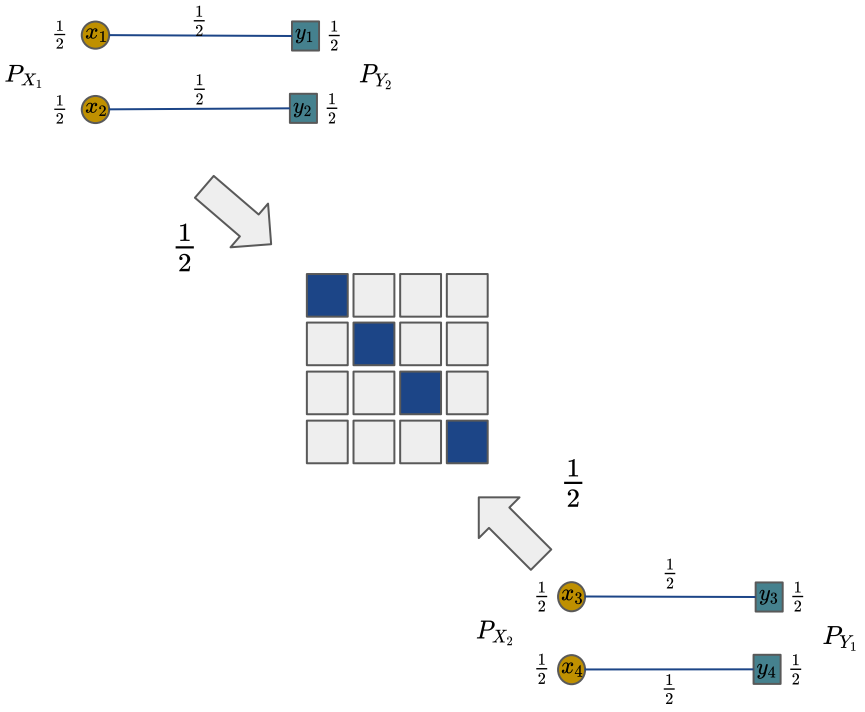
Batch of Mini-batches Optimal Transport



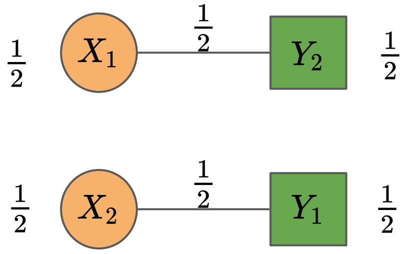
Matching pairs of mini-batches with optimal transport ground cost



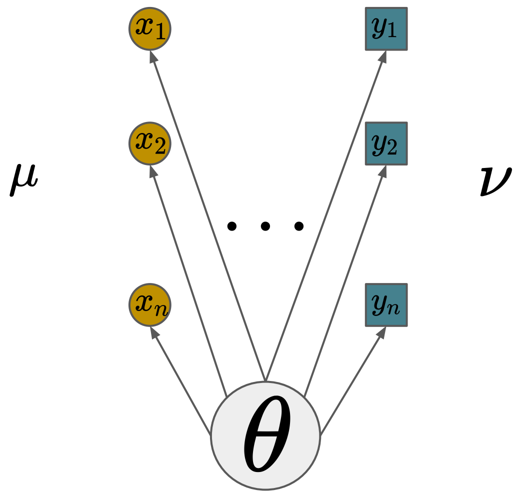
Batch of Mini-batches Optimal Transport



Matching pairs of mini-batches with optimal transport ground cost



Training deep networks with BoMb-OT loss



Supports are functions of parameters of neural networks

Set $\nabla_\theta^k = 0$

For $i = 1$ to k

For $j = 1$ to k

Compute $X_i(\theta), Y_j(\theta); \mathcal{C}_{ij} = OT(P_{X_i(\theta)}, P_{Y_j(\theta)})$

Step 1

Compute plan $\gamma \leftarrow OT(\mu_k^{\otimes m}, \nu_k^{\otimes m}, \mathcal{C})$

Step 2

For $i = 1$ to k

For $j = 1$ to k

Compute $X_i(\theta), Y_j(\theta)$

Compute $\nabla_\theta^k = \nabla_\theta^k + \nabla_\theta \gamma_{ij} OT(P_{X_i(\theta)}, P_{Y_j(\theta)})$

Update θ based on the stochastic gradient ∇_θ^k

Step 3

Experiments on Deep Domain Adaptation

Scenario	k	m-OT	BoMb-OT	m-UOT	BoMb-UOT
S → M	1	90.98 ± 1.00	90.98 ± 1.00	99.10 ± 0.05	99.10 ± 0.05
	2	92.66 ± 0.71	$\mathbf{93.36} \pm \mathbf{0.51}$	98.99 ± 0.08	$\mathbf{99.15} \pm \mathbf{0.07}$
	4	93.70 ± 0.53	$\mathbf{94.79} \pm \mathbf{0.18}$	99.08 ± 0.08	$\mathbf{99.19} \pm \mathbf{0.05}$
	8	94.17 ± 0.39	$\mathbf{95.23} \pm \mathbf{0.44}$	98.94 ± 0.03	$\mathbf{99.09} \pm \mathbf{0.07}$
U → M	1	92.64 ± 0.45	92.64 ± 0.45	98.14 ± 0.21	98.14 ± 0.21
	2	92.85 ± 0.34	$\mathbf{94.76} \pm \mathbf{0.12}$	98.34 ± 0.11	$\mathbf{98.44} \pm \mathbf{0.05}$
	4	93.83 ± 0.56	$\mathbf{95.64} \pm \mathbf{0.26}$	98.55 ± 0.04	$\mathbf{98.60} \pm \mathbf{0.02}$
	8	94.69 ± 0.73	$\mathbf{95.90} \pm \mathbf{0.33}$	98.62 ± 0.08	$\mathbf{98.70} \pm \mathbf{0.06}$

Adapting classification on digits datasets

k	m-OT	BoMb-OT	m-UOT	BoMb-UOT
1	65.29 ± 0.26	65.29 ± 0.26	72.07 ± 0.07	72.07 ± 0.07
2	65.46 ± 0.33	$\mathbf{66.98} \pm \mathbf{0.09}$	72.52 ± 0.14	$\mathbf{73.72} \pm \mathbf{0.13}$
4	65.51 ± 0.17	$\mathbf{67.71} \pm \mathbf{0.05}$	72.95 ± 0.06	$\mathbf{74.65} \pm \mathbf{0.03}$

Adapting classification on VISDA dataset

Experiments on Deep Domain Adaptation

Adapting classification on Office-Home datasets

Methods	A2C	A2P	A2R	C2A	C2P	C2R	P2A	P2C	P2R	R2A	R2C	R2P	Avg
RESNET-50 (*)	34.90	50.00	58.00	37.40	41.90	46.20	38.50	31.20	60.40	53.90	41.20	59.90	46.10
DANN (*)	44.30	59.80	69.80	48.00	58.30	63.00	49.70	42.70	70.60	64.00	51.70	78.30	58.30
CDAN-E (*)	52.50	71.40	76.10	59.70	69.90	71.50	58.70	50.30	77.50	70.50	57.90	83.50	66.60
ALDA (*)	52.20	69.30	76.40	58.70	68.20	71.10	57.40	49.60	76.80	70.60	57.30	82.50	65.80
ROT (*)	47.20	71.80	76.40	58.60	68.10	70.20	56.50	45.00	75.80	69.40	52.10	80.60	64.30
m-OT	49.76	68.37	74.40	59.64	64.69	68.63	56.12	46.69	74.37	67.27	54.45	77.95	63.53
m-UOT	54.99	74.45	**80.78**	65.66	**74.93**	74.91	64.70	**53.42**	80.01	**74.58**	59.88	**83.73**	70.17
BoMb-OT (Ours)	50.16	69.57	74.84	60.24	65.18	69.15	57.48	47.42	74.88	67.39	54.19	78.59	64.09
BoMb-UOT (Ours)	**56.23**	**75.24**	80.53	**65.80**	74.57	**75.38**	**66.15**	53.21	**80.03**	74.25	**60.12**	83.30	**70.40**

Other applications including deep generative models, color transfer, approximate Bayesian computation, gradient flow are in the paper.

Conclusion

- ❑ Solving an additional optimal transport problem to match mini-batches could improve the performance of applications that use mini-batch OT losses.
  - ❑ Three steps algorithm.

- ❑ Using different types of transportation e.g., unbalanced optimal transport (UOT) could improve further the performance.

Thank you for listening!

Khai Nguyen: khainb@utexas.edu @KhaiBaNguyen