# Computing the Optimal Transport Map Between Probabilistic Circuits

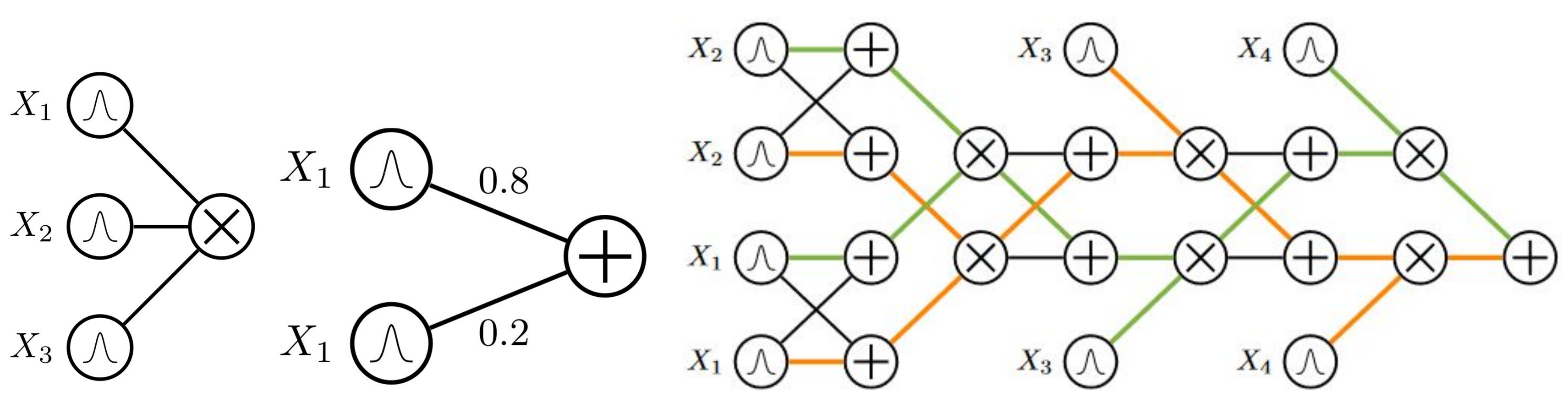

Adrian Ciotinga, Computer Science

Mentor: Dr. YooJung Choi, Assistant Professor

School of Computing and Augmented Intelligence


## Background

### Probabilistic Circuits (PCs)




Product nodes encode a factorized (independent) distribution

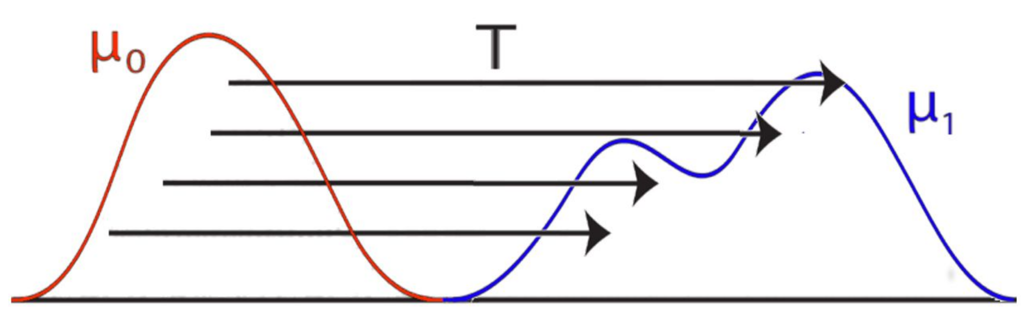
Sum nodes encode a mixture of distributions

- The sum of PCs is a PC, the product of PCs is a PC
- PCs are tractable for computing measures like the Kullback-Leibler Divergence or Cauchy-Schwartz Divergence [1], **but not the Wasserstein distance**

### Optimal Transport and the Wasserstein Distance

$$W_p(\mu, \upsilon) = \Big(\inf_{\gamma\in\Gamma(\mu,\upsilon)} \mathbb{E}_{(x,y)\sim\gamma}[||x-y||_p^p]\Big)^{\frac{1}{p}}$$

- The optimal transport map γ between distributions μ and υ maps each point in one distribution to a distribution of points in the other distribution
- Since γ minimizes the expectation above, **computing the Wasserstein distance requires solving an optimization problem**, which is hard

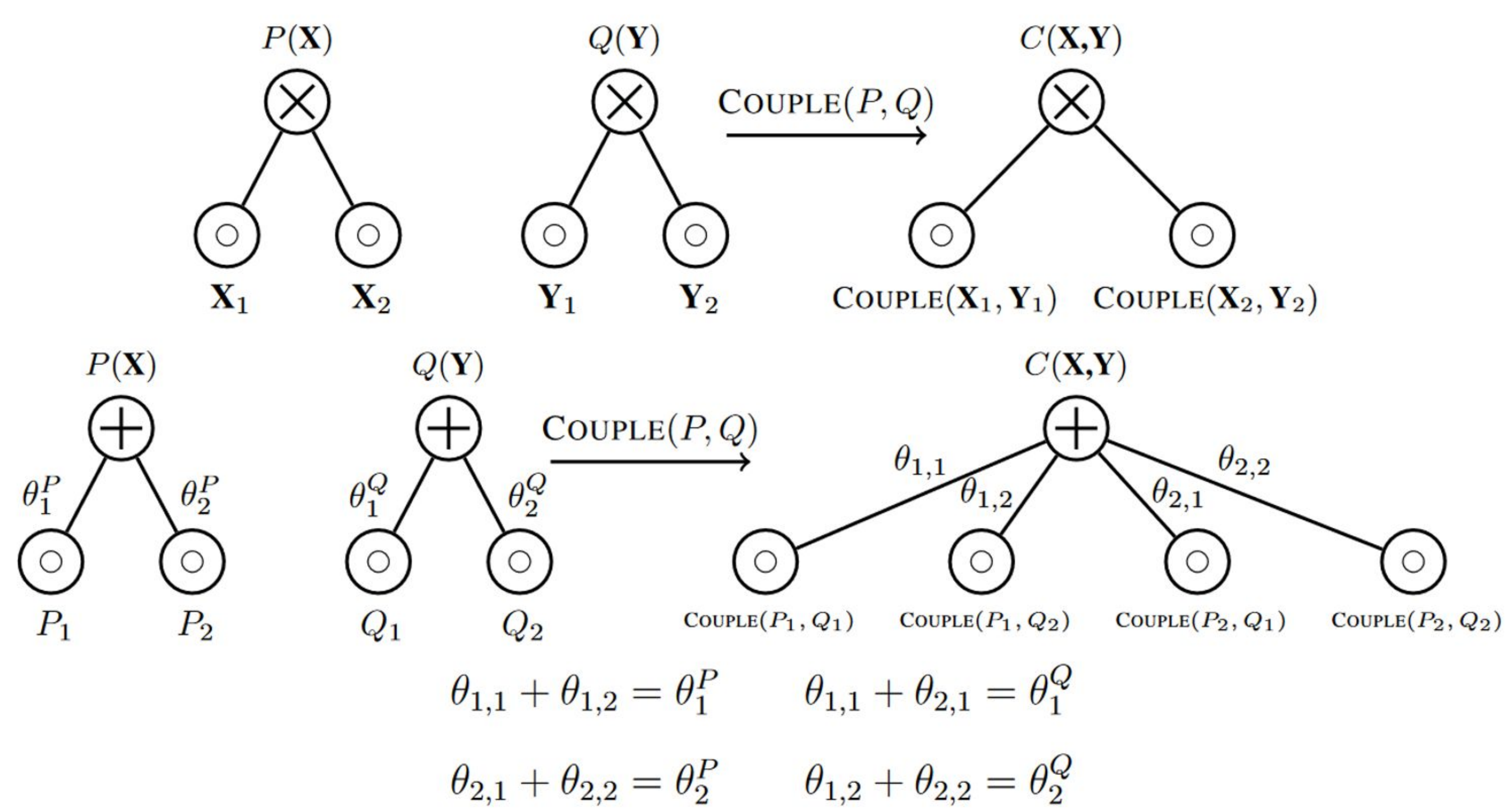



## Motivation

- The Wasserstein distance is commonly used in fields such as computer science and statistics
- In many cases, computing the Wasserstein distance or optimal transport map between distributions is computationally expensive or impractical
- Tractable computation of the Wasserstein distance and optimal transport map between PCs allows for the comparison of and interpolation between two probability distributions on a metric space

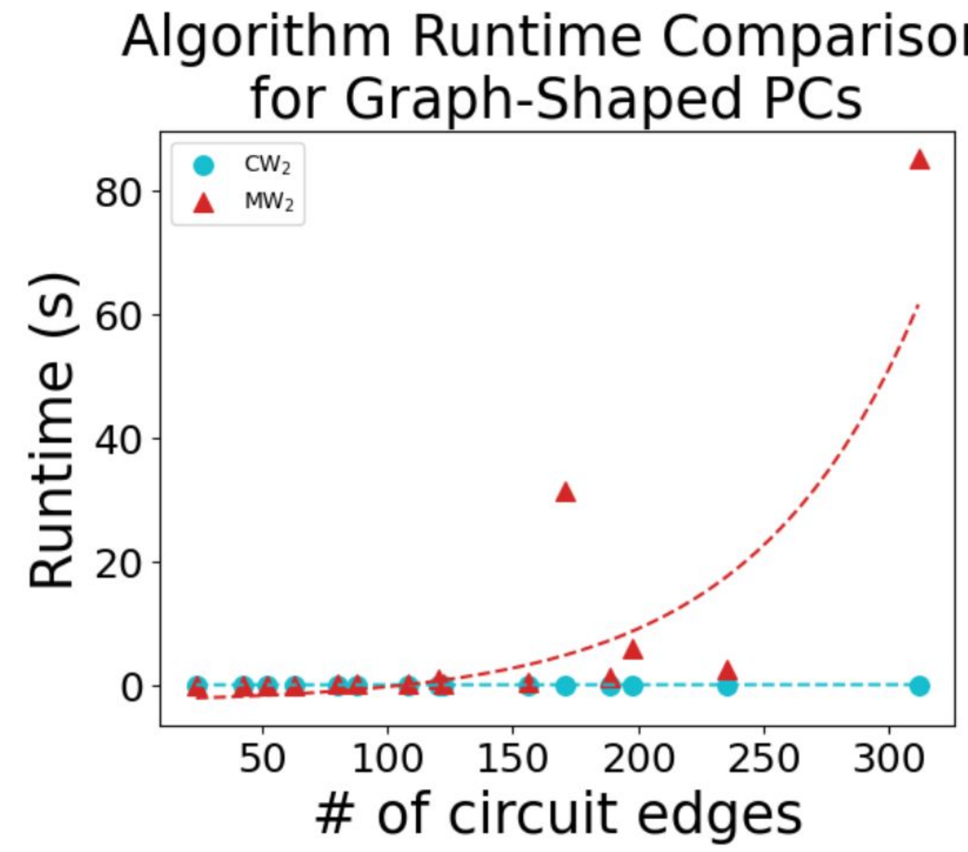
## Algorithm Overview

- First, construct a coupling circuit structure that is the *product* of the original two circuits
- Then, solve a **linear programming problem** at each sum node to get the optimal parameters




- **The resulting optimal transport map is itself a PC**

## Implementation Results




- Our approach is ***exponentially faster*** than an existing approach that restricts the transport map to be a Gaussian Mixture Model (GMM) [2]
- Larger circuit branching factors resulted in an **out-of-memory error for the GMM Wasserstein method**. All experiments were run on a machine with an Intel i7-8750H CPU and 80GB of memory

## Future Work

- Parameter learning of a probabilistic circuit by minimizing the Wasserstein distance to a dataset
- Improving the expressivity of the coupling circuit

## References


[1] Y. Choi, A. Vergari, and G. Van den Broeck. “Probabilistic circuits: A unifying framework for tractable probabilistic modeling.” 2020.

[2] J. Delon and A. Desolneux. “A Wasserstein-Type Distance in the Space of Gaussian Mixture Models.” *SIAM Journal on Imaging Sciences*. Vol. 13, No. 2, pages 936-970, 2020.


FURI

Ira A. Fulton Schools of Engineering

Arizona State University