

On the Hardness of Approximating Distributions with Tractable Probabilistic Models

John Leland YooJung Choi
Arizona State University



Problem Setup

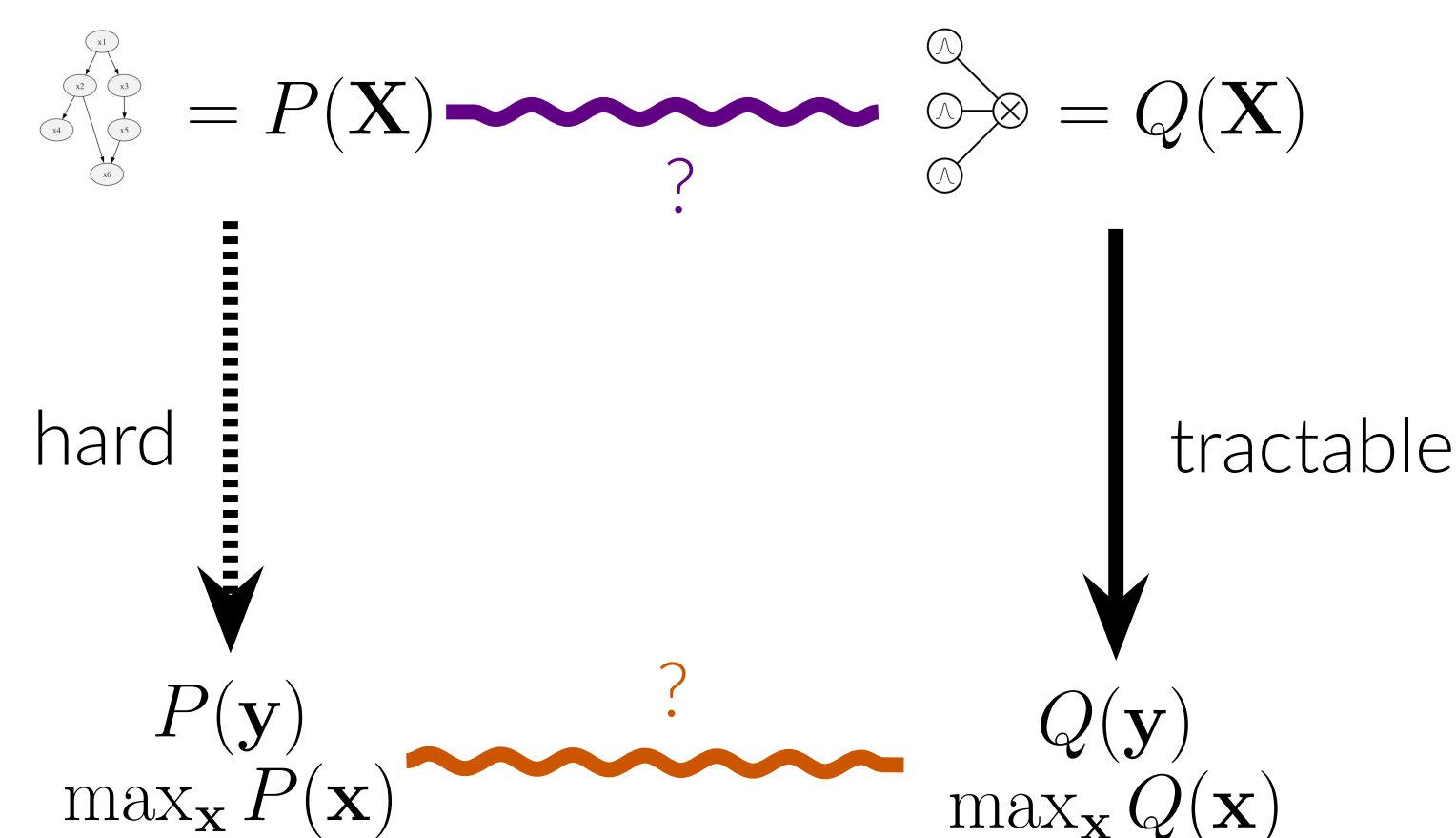


Figure: Approximation scheme from an intractable model to a tractable model

Tractable Probabilistic Models (TPMs) are probabilistic models that meet the following requirements:

- ▶ The model can perform exact inference queries.
- ▶ The queries can be performed in polynomial time in the size of the model.

Problem: It is NP-hard to represent distributions exactly with TPMs.

- ▶ Possible solution: Approximate the distribution instead!
 - Define approximate modeling as: $D_f(P||Q) < k\epsilon^2$ or $D_{TV}(P||Q) < \epsilon$.
 - If $\mathcal{D}(P||Q) < \epsilon$ for a distance $\mathcal{D}$, we call Q an ϵ - $\mathcal{D}$ -approximator of P .

This leads us to two questions: (1) **What is the complexity of approximate modeling using TPMs?** and (2) **Does approximate modeling guarantee approximate inference?**

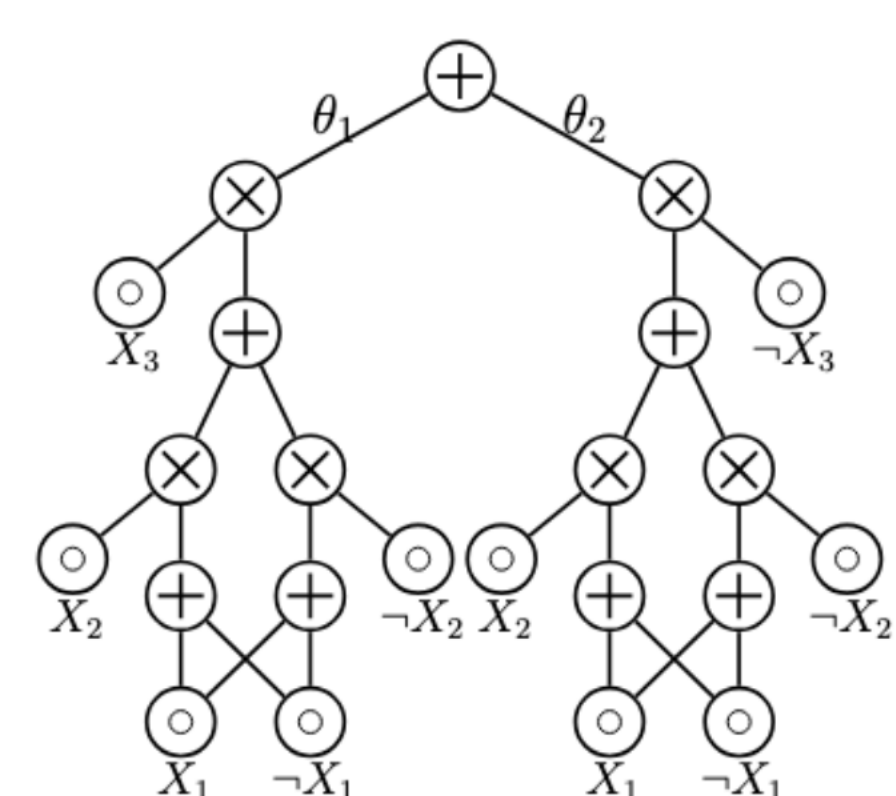


Figure: A smooth, decomposable, and deterministic PC

- ▶ Probabilistic Circuits (PCs) are a unifying framework for TPMs [3].
 - PCs are closely related to Boolean Circuits in Negation Normal Form (NNF).
 - Negation Normal form (NNF) circuits are over $\wedge$, $\vee$ and literals for leaves.
- ▶ Circuits impose structural constraints to access to more queries. Two powerful constraints are *decomposability* and *determinism*.
 - Enforcing decomposability guarantees linear-time *marginal* inference in PCs.
 - Enforcing both decomposability and determinism guarantees linear-time *maximum-a-posteriori* (MAP) inference.
- ▶ But, there is an exponential gap in expressive efficiency between decomposable and deterministic decomposable circuits! [2]

Main Question: Can we avoid this size blowup by allowing some small approximation error?

Hardness of D_f -approximation

Theorem 1

Given a (potentially unnormalized) probability distribution $\hat{P}$ and a k -convex f -divergence D_f , for any $0 < \epsilon < \frac{1}{4}$, it is NP-hard to represent the $k\epsilon^2$ - D_f -approximation of its normalized distribution P as a model that can tractably compute marginals.

- ▶ We prove this by a reduction from $\text{SAT}(f)$
 - Define another Boolean function f' such that $f' = (Y \wedge f) \vee (\neg Y \wedge X_1 \wedge \dots \wedge X_n)$
 - Place a uniform distribution, $\hat{P}$, over the models of f' .
 - If a model that can tractably marginalize, Q , is such that $D_{TV}(P||Q) < 1/4$, then we can solve $\text{SAT}(f)$ by thresholding $Q(Y = 1) \geq 1/4$.

The class of polynomial-sized models which can tractably marginalize, cannot contain $k\epsilon^2$ - $\mathcal{D}_f$ for all distributions unless $P = NP$. **Thus, approximate modeling is NP-hard for any TPM!**

- ▶ This result applies to decomposable PCs [3], Probabilistic Generating Circuits [7], Determinantal Point Processes [1], Sum of Squares Circuits [5], and Positive Unital Circuits [6].

Large Dec. & Det. PCs for Approximate Modeling

Approximate modeling is NP-hard, but can we overcome the exponential gap caused by enforcing determinism?

- ▶ We show the exponential blowup extends to the approximate setting in both Boolean and Probabilistic Circuits.
- ▶ Connecting from approximating Boolean circuits to Probabilistic Circuits is non-trivial:
 - Approximation for PCs measured in terms of divergences.
 - PCs can represent arbitrary distributions.

Consider a decomposable PC, P_n , which is the distribution constructed by taking the DNNF of the *Sauerhoff* function S_n and replacing $\wedge$ with $\otimes$, $\vee$ with $\oplus$ and literals with indicators. Then P_n has size $O(n^2)$, but we prove any deterministic, decomposable PC approximating P_n must be exponentially sized!

Theorem 2

A deterministic, decomposable PC that is a ϵ - D_{TV} -Approximator of P_n , where $\epsilon = \frac{1}{16} - \Omega(1/\text{Poly}(n^2))$, has size $2^{\Omega(n)}$.

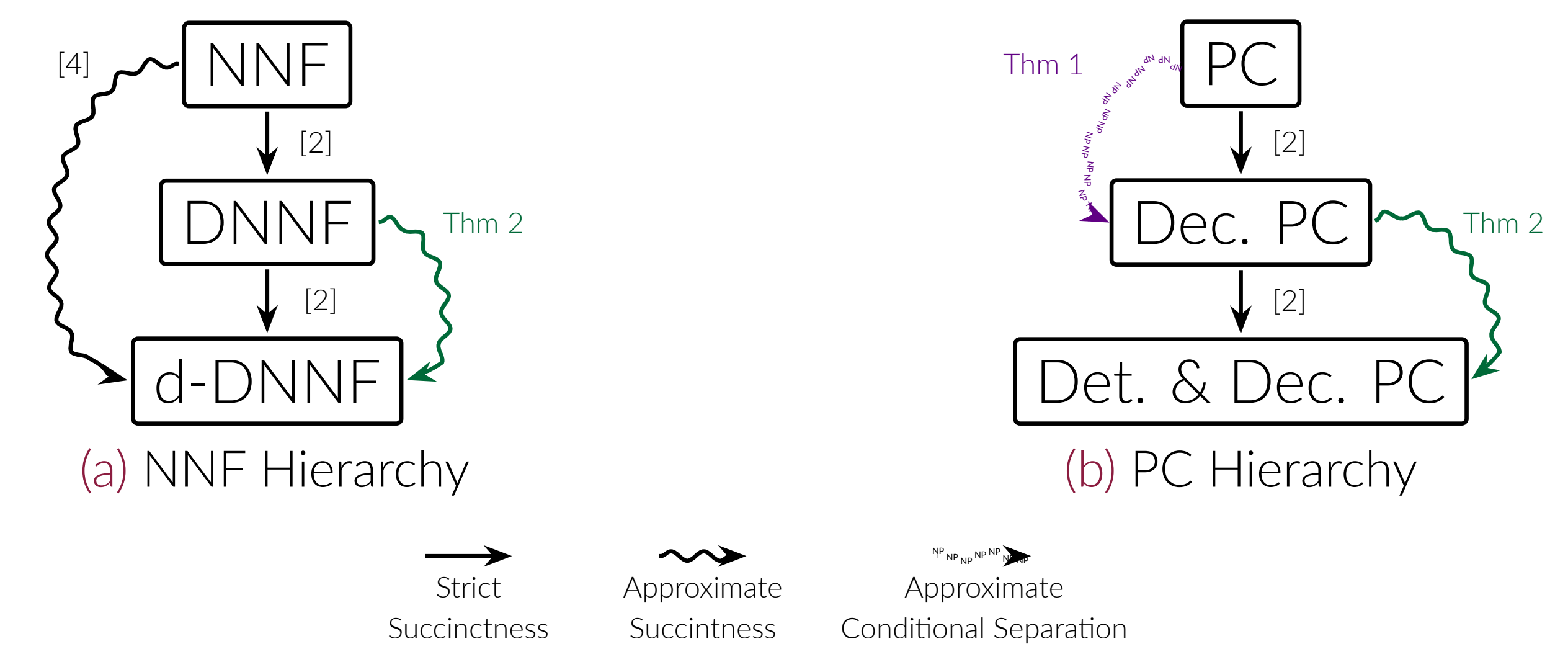


Figure: Updated succinctness relationships for NNF and PCs.

Thus we cannot avoid the exponential gap in expressive efficiency between decomposable and deterministic, decomposable circuits even in the approximate case!

Inferences Guarantees in Approximate Modeling

Table: Inference Guarantees

	$D_f(P Q)$	$D_{TV}(P Q)$
a-Marginal	✓	✓
a-Maximum-a-posteriori	✓	✓
a-Conditional	○	○
r-Marginal	○	○
r-Maximum-a-posteriori	○	○
r-Conditional	○	○

Legend: ✓ indicates the distance guarantees the inference type; ○ indicates the distance cannot guarantee the inference type unless $P = NP$.

* a denotes *absolute* approximation, r denotes *relative* approximation.

Future Work

- ▶ Can a polynomial-time algorithm enable learning an ϵ -approximator for a broad family of distributions with a more relaxed ϵ ?
- ▶ Is there a commonly used condition which can guarantee relative approximation?
- ▶ Can we extend approximate lower bounds to non-deterministic circuits?

References

- [1] Borodin. Determinantal point processes, 2009.
- [2] Bova et al. Knowledge compilation meets communication complexity. In *IJCAI*, 2016.
- [3] Choi et al. Probabilistic circuits: A unifying framework for tractable probabilistic models, 2020.
- [4] de Colnet et al. Lower bounds for approximate knowledge compilation. In *KR*, 2021.
- [5] Loconte et al. Sum of squares circuits, 2025.
- [6] Zuidberg Dos Martires. A quantum information theoretic approach to tractable probabilistic models. In *UAI*, 2025.
- [7] Zhang et al. Probabilistic generating circuits. In *ICML*, 2021.