Sum of Squares Circuits

Lorenzo Loconte
University of Edinburgh, UK

Stefan Mengel
University of Artois, CNRS, CRIL, FR

Antonio Vergari
University of Edinburgh, UK

AAAI-25 17 January 2025

There are many probabilistic models...

Estimate $p(\mathbf{x})$ with a model

GMMs | Trees | PSD | SNFs

HMMs | NTFs | BMs

MAFs | VAEs | GANs

LLMs | Dffusion | EBMs

There are many probabilistic models...

Estimate $p(\mathbf{x})$ with a model

GMMs | Trees | PSD | SNFs

HMMs | NTFs | BMs

MAFs | VAEs | GANs

LLMs | Dffusion | EBMs

Probe $p(\mathbf{x})$ to solve tasks

Lossless data (de-)compression:

$p(x_i \mid x_1, \ldots, x_{i-1})$

Neurosymbolic reasoning:

$\mathbb{E}_{\mathbf{x} \in p(\mathbf{x})}[\kappa(\mathbf{x})] = \int_{\mathcal{X}} p(\mathbf{x})\kappa(\mathbf{x})\mathrm{d}\mathbf{x}$

Liu, Mandt, and Broeck, “Lossless Compression with Probabilistic Circuits”, 2022

Ahmed et al., “Semantic probabilistic layers for neuro-symbolic learning”, 2022

There are many probabilistic models...

Estimate $p(\mathbf{x})$ with a model

Expressive generative models
(often intractable inference)

MAFs | VAEs | GANs

LLMs | Dffusion | EBMs

Probe $p(\mathbf{x})$ to solve tasks

Lossless data (de-)compression:
$p(x_i \mid x_1, \ldots, x_{i-1})$

Neurosymbolic reasoning:
$\mathbb{E}_{\mathbf{x} \in p(\mathbf{x})}[\kappa(\mathbf{x})] = \int_{\mathcal{X}} p(\mathbf{x}) \kappa(\mathbf{x}) \mathrm{d}\mathbf{x}$

There are many probabilistic models...

Estimate $p(\mathbf{x})$ with a model

GMMs | Trees | PSD | SNFs

HMMs | NTFs | BMs

Models supporting tractability
(often not expressive enough)

Probe $p(\mathbf{x})$ to solve tasks

Lossless data (de-)compression:
$p(x_i \mid x_1, \dots, x_{i-1})$

Neurosymbolic reasoning:
$\mathbb{E}_{\mathbf{x} \in p(\mathbf{x})}[\kappa(\mathbf{x})] = \int_{\mathcal{X}} p(\mathbf{x}) \kappa(\mathbf{x}) \mathrm{d}\mathbf{x}$

"How to build probabilistic models that are expressive yet support tractable inference?"

Circuits: why?

Neural networks whose **tractability** and **expressiveness** can be analyzed theoretically

Circuits: why?

Neural networks whose **tractability** and **expressiveness** can be analyzed theoretically

Complexity of **exact inference** is polynomial w.r.t. **circuit size**, under structural assumptions

conditionals: $p(x_i \mid x_1, \ldots, x_{i-1})$

expectations: $\mathbb{E}_{\mathbf{x} \in p(\mathbf{x})}[\kappa(\mathbf{x})]$

...and more!

Choi, Vergari, and Broeck, Probabilistic Circuits: A Unifying Framework for Tractable Probabilistic Modeling, 2020

Vergari et al., "A Compositional Atlas of Tractable Circuit Operations for Probabilistic Inference", 2021

Circuits: why?

Neural networks whose **tractability** and **expressiveness** can be analyzed theoretically

Framework to study expressiveness,
based on **circuit complexity theory**

Valiant, "Negation can be exponentially powerful", 1979

Colnet and Mengel, "A Compilation of Succinctness Results for Arithmetic Circuits", 2021

Circuits: why?

Neural networks whose **tractability** and **expressiveness** can be analyzed theoretically

Framework to study expressiveness, based on **circuit complexity theory**

GMMs Trees PSD SNFs

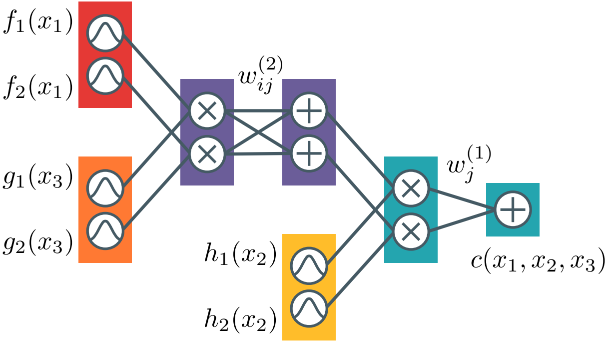
HMMs NTFs BMs

...are all circuits! (more later)

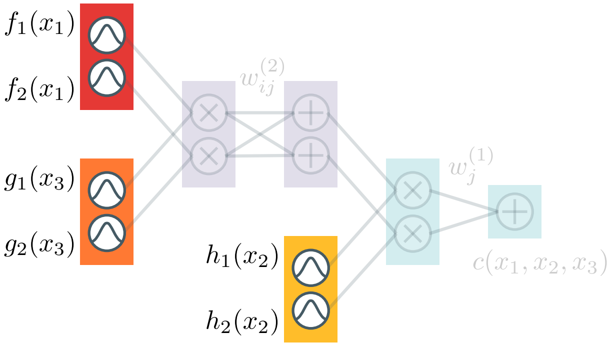
Valiant, “Negation can be exponentially powerful”, 1979

Colnet and Mengel, “A Compilation of Succinctness Results for Arithmetic Circuits”, 2021

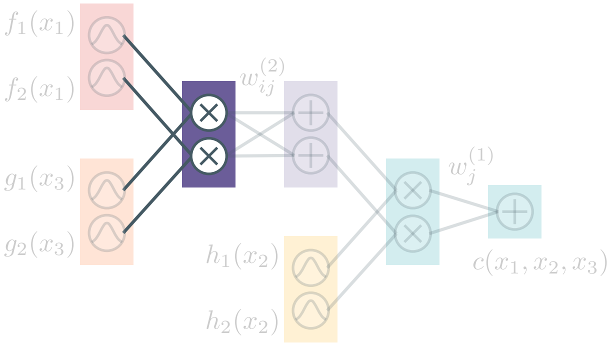
Circuits: what?



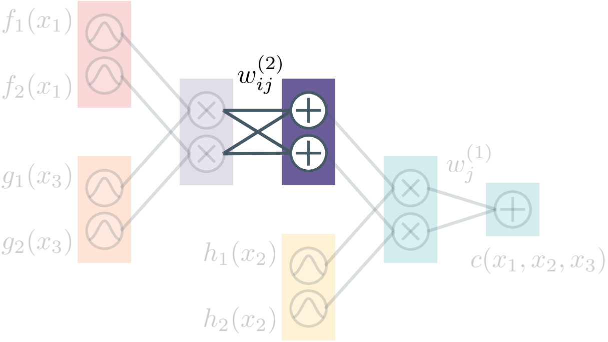
Circuits: what?



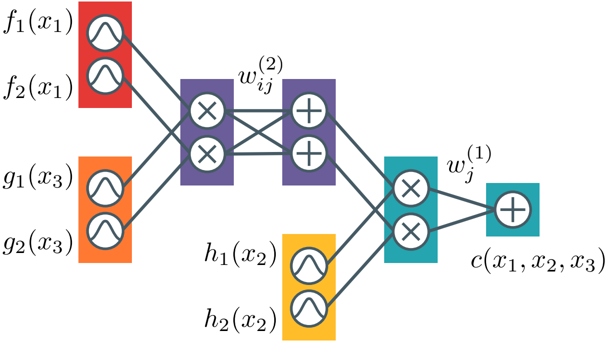
Circuits: what?



Circuits: what?



Circuits: what?



Circuits: how to model distributions?

Monotonic circuits

$$p(\mathbf{x}) = \frac{1}{Z}\, c(\mathbf{x}), \quad c(\mathbf{x}) \geq 0$$

where parameters and input functions are **positive**

Choi, Vergari, and Broeck, Probabilistic Circuits: A Unifying Framework for Tractable Probabilistic Modeling, 2020

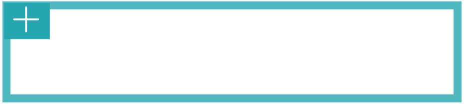
Circuits: how to model distributions?

Monotonic circuits

$$p(\mathbf{x}) = \frac{1}{Z}\, c(\mathbf{x}), \quad c(\mathbf{x}) \geq 0$$

where parameters and input functions are **positive**

 $=$ set of distributions modeled by **polysize** circuits

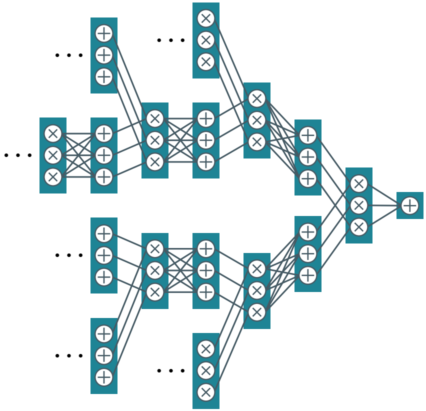


A limitation of monotonic circuits

 = set of distributions modeled by **polysize** circuits

- UDISJ

Loconte et al., "Subtractive Mixture Models via Squaring: Representation and Learning", 2024



$\exists\, p$ requiring **exponentially large** **monotonic circuits**...

Loconte et al., "Subtractive Mixture Models via Squaring: Representation and Learning", 2024

Squared circuits

$$p(\mathbf{x}) = \frac{1}{Z}\, c^2(\mathbf{x}), \quad c(\mathbf{x}) \in \mathbb{R}$$

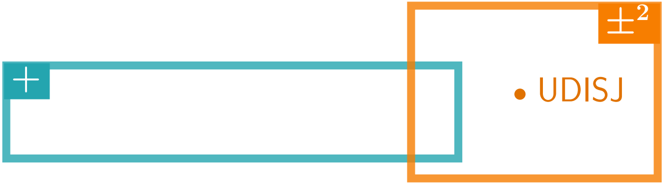
where parameters and input functions can be **negative**

Loconte et al., “Subtractive Mixture Models via Squaring: Representation and Learning”, 2024

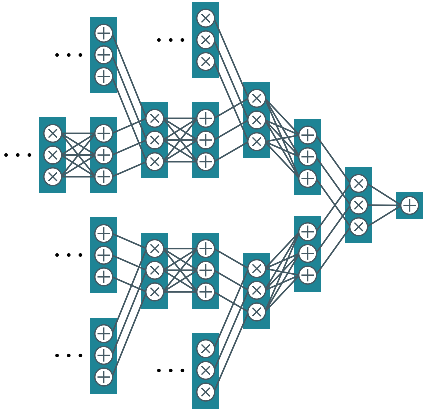
Squared circuits

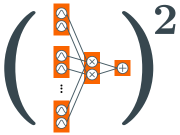
$$p(\mathbf{x}) = \frac{1}{Z}\, c^2(\mathbf{x}), \quad c(\mathbf{x}) \in \mathbb{R}$$

where parameters and input functions can be **negative**



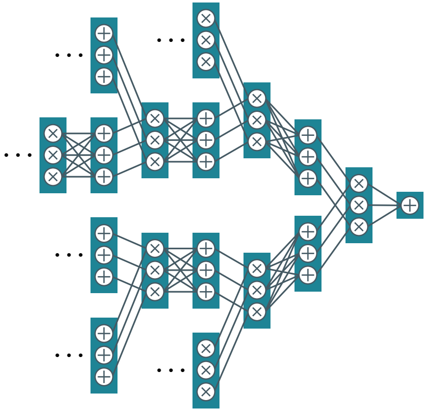
Loconte et al., "Subtractive Mixture Models via Squaring: Representation and Learning", 2024

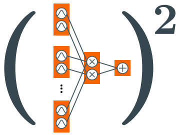




...instead **squared circuits** require **polynomial size**

Loconte et al., "Subtractive Mixture Models via Squaring: Representation and Learning", 2024



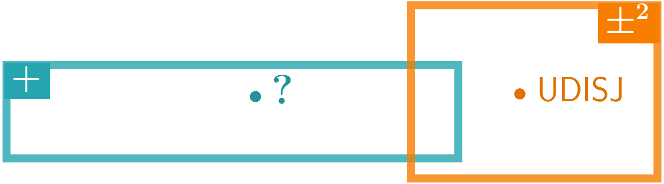


Squared circuits more expressive than **monotonic** ones

Loconte et al., “Subtractive Mixture Models via Squaring: Representation and Learning”, 2024

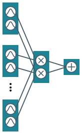
Outline

1. *Can monotonic circuits be more expressive than squared?*

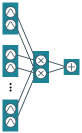


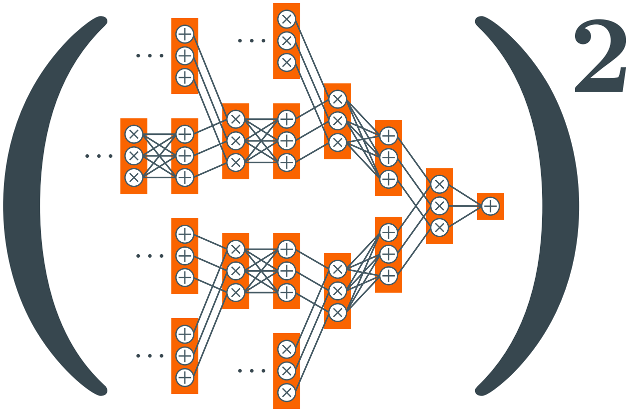
Outline

1. *Can monotonic circuits be more expressive than squared?*



$\exists\, p$ requiring **polysize** **monotonic circuits**...





...but require **exponentially large** **squared circuits**

Theorem 1.

There is a class of non-negative functions $\mathcal{F}$ over $d = k(k+1)$ variables that can be encoded by a PC in $+_{\mathsf{sd}}$ having size $\mathcal{O}(d)$. However, the smallest $c^2 \in \pm^2_{\mathbb{R}}$ computing any $F \in \mathcal{F}$ requires $|c|$ to be at least $2^{\Omega(\sqrt{d})}$.

Squaring alone can reduce expressiveness!

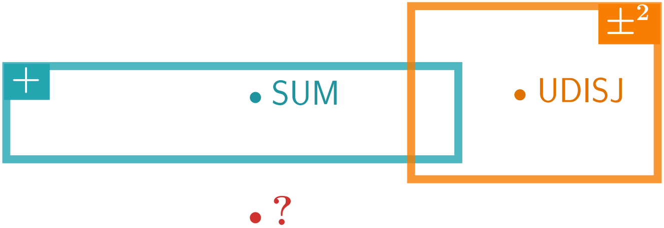
Wang and Van den Broeck, On the Relationship Between Monotone and Squared Probabilistic Circuits, 2024

Outline

1. *Can monotonic circuits be more expressive than squared?*

   $\Longrightarrow$ Yes!

2. *How to build models more expressive than <u>both</u>?*



Sum of squares (SOS) circuits

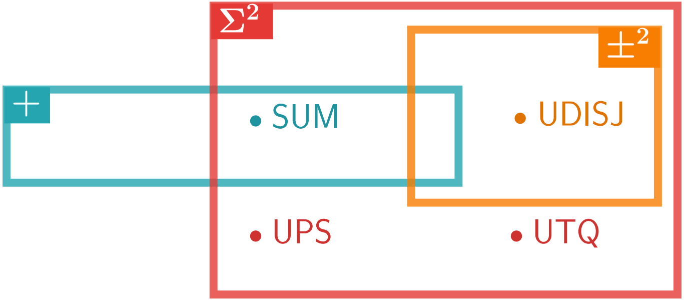
$p(\mathbf{x}) = \frac{1}{Z}\sum_{i=1}^{r} c_i^2(\mathbf{x}), \quad c_i(\mathbf{x}) \in \mathbb{R}$

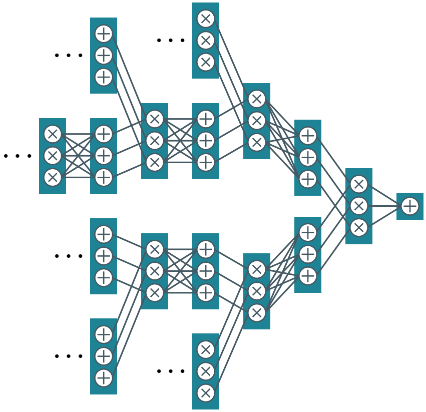
where parameters and input functions can be **negative**

Sum of squares (SOS) circuits

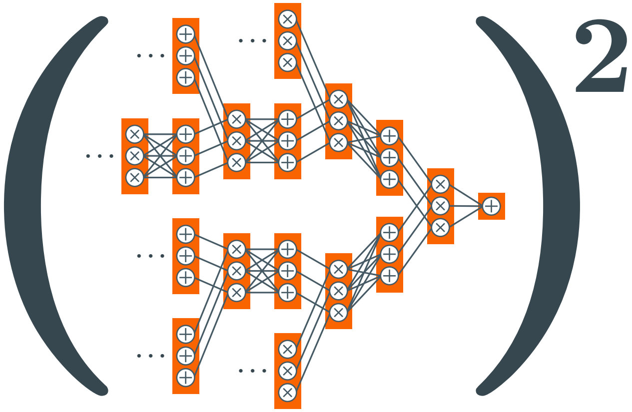
$$p(\mathbf{x}) = \frac{1}{Z}\sum_{i=1}^{r} c_i^2(\mathbf{x}), \quad c_i(\mathbf{x}) \in \mathbb{R}$$

where parameters and input functions can be **negative**

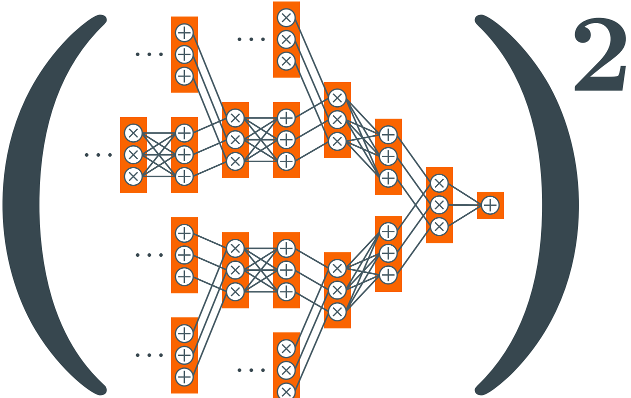


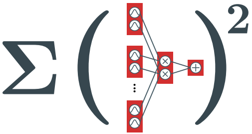


$\exists\, p$ requiring **exponentially large monotonic circuits**...



...and also **exponentially large squared circuits** ...





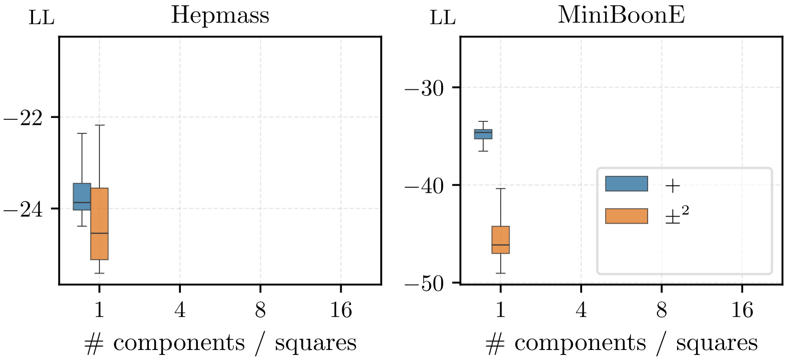
...but a **sum of squares (SOS)** **polysize** circuits

Theorem 2.

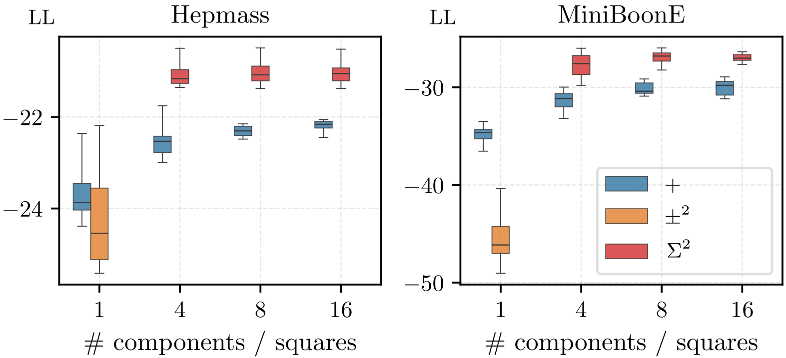
There is a class of non-negative functions $\mathcal{F}$ over d variables that can be represented by a PC in $\Sigma^2_{\textsf{cmp}}$ of size $\mathcal{O}(d^3)$. However, (i) the smallest PC in $+_{\textsf{sd}}$ computing any $F \in \mathcal{F}$ has at least size $2^{\Omega(\sqrt{d})}$, and (ii) the smallest $c^2 \in \pm^2_{\mathbb{R}}$ computing F obtained by squaring a structured-decomposable circuit c, requires $|c|$ to be at least $2^{\Omega(\sqrt{d})}$.

SOS can surpass both expressiveness limitations!

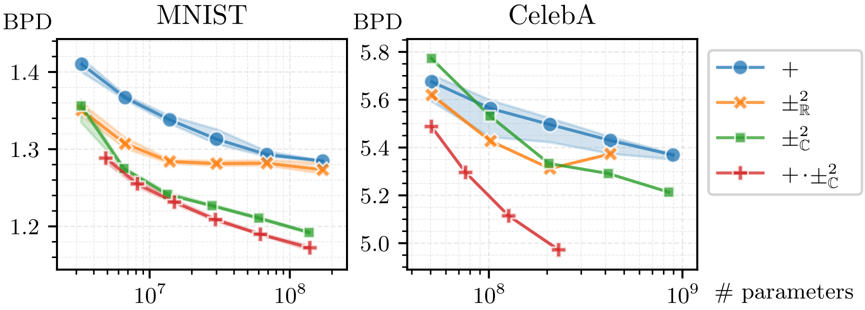
Experiments



Experiments



Experiments



Outline

1. *Can monotonic circuits be more expressive than squared?*

   $\Longrightarrow$ Yes!

2. *How to build models more expressive than <u>both</u>?*

   $\Longrightarrow$ SOS circuits!

3. *How are SOS circuits related to other probabilistic models?*

$$p(\mathbf{x}) \propto \boldsymbol{\kappa}(\mathbf{x})^\top \mathbf{A} \boldsymbol{\kappa}(\mathbf{x})$$

with $\mathbf{A} \in \mathbb{R}^{d \times d}$ PSD

Positive Semi-Definite models

Rudi and Ciliberto, “PSD Representations for Effective Probability Models”, 2021

Tsuchida, Ong, and Sejdinovic, “Squared Neural Families: A New Class of Tractable Density Models”, 2023

Glasser et al., “Expressive power of tensor-network factorizations for probabilistic modeling”, 2019

$$p(\mathbf{x}) \propto \boldsymbol{\kappa}(\mathbf{x})^\top \mathbf{A} \boldsymbol{\kappa}(\mathbf{x})$$

with $\mathbf{A} \in \mathbb{R}^{d \times d}$ PSD

Positive Semi-Definite models

$$p(\mathbf{x}) \propto \mu(\mathbf{x}) ||\mathrm{NN}_{\sigma,\Theta}(\boldsymbol{t}(\mathbf{x}))||_2^2$$

Squared Neural Families

Rudi and Ciliberto, "PSD Representations for Effective Probability Models", 2021

Tsuchida, Ong, and Sejdinovic, "Squared Neural Families: A New Class of Tractable Density Models", 2023

Glasser et al., "Expressive power of tensor-network factorizations for probabilistic modeling", 2019

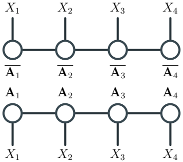
$$p(\mathbf{x}) \propto \boldsymbol{\kappa}(\mathbf{x})^\top \mathbf{A} \boldsymbol{\kappa}(\mathbf{x})$$

with $\mathbf{A} \in \mathbb{R}^{d \times d}$ PSD

Positive Semi-Definite models

$$p(\mathbf{x}) \propto \mu(\mathbf{x}) ||\mathrm{NN}_{\sigma,\Theta}(\boldsymbol{t}(\mathbf{x}))||_2^2$$

Squared Neural Families



$$p(\mathbf{x}) \propto \psi(\mathbf{x}) \psi(\mathbf{x})^\dagger, \psi(\mathbf{x}) \in \mathbb{C}$$

Complex Born machines

Rudi and Ciliberto, “PSD Representations for Effective Probability Models”, 2021

Tsuchida, Ong, and Sejdinovic, “Squared Neural Families: A New Class of Tractable Density Models”, 2023

Glasser et al., “Expressive power of tensor-network factorizations for probabilistic modeling”, 2019

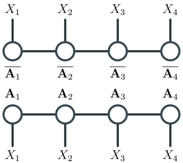
$$p(\mathbf{x}) \propto \boldsymbol{\kappa}(\mathbf{x})^\top \mathbf{A} \boldsymbol{\kappa}(\mathbf{x})$$

with $\mathbf{A} \in \mathbb{R}^{d \times d}$ PSD

Positive Semi-Definite models

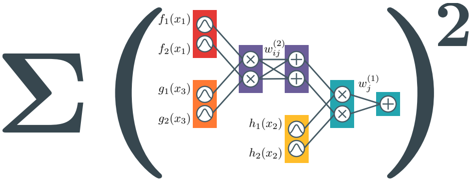
$$p(\mathbf{x}) \propto \mu(\mathbf{x}) ||\mathrm{NN}_{\sigma,\Theta}(\boldsymbol{t}(\mathbf{x}))||_2^2$$

Squared Neural Families



$$p(\mathbf{x}) \propto \psi(\mathbf{x})\psi(\mathbf{x})^\dagger, \psi(\mathbf{x}) \in \mathbb{C}$$

Complex Born machines



...can be reduced to SOS circuits!

$\implies$ *summing squares boosts expressiveness*

$\implies$ *complex parameters help*

Rudi and Ciliberto, “PSD Representations for Effective Probability Models”, 2021

Tsuchida, Ong, and Sejdinovic, “Squared Neural Families: A New Class of Tractable Density Models”, 2023

Glasser et al., “Expressive power of tensor-network factorizations for probabilistic modeling”, 2019

Takeaways

1. A more precise expressiveness characterization of squared circuits

Paper *Code*

loreloc.github.io

@loreloc_

@loreloc

Takeaways

1. A more precise expressiveness characterization of squared circuits

2. SOS circuits can be more expressive than both monotonic & squared circuits

Paper ***Code***

`loreloc.github.io`

`@loreloc_`

`@loreloc`

Takeaways

1. A more precise expressiveness characterization of squared circuits

2. SOS circuits can be more expressive than both monotonic & squared circuits

3. Connect SOS with other models, thus we understand why they are expressive

Paper ***Code***

loreloc.github.io

@loreloc_

@loreloc