Sum of Squares Circuits

Lorenzo Loconte
University of Edinburgh, UK

Stefan Mengel
University of Artois, CNRS, CRIL, FR

Antonio Vergari
University of Edinburgh, UK

Paper · Code

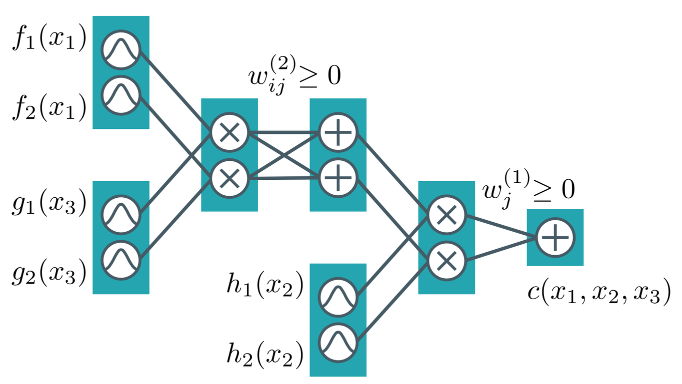
TL;DR: "We show an expressiveness hierarchy including many tractable probabilistic model classes, when represented as deep structured computational graphs called circuits"

Circuits are computational graphs unifying probabilistic models that support efficient and exact inference

Tasks
- lossless (de)compression[1] $p(x_i \mid x_1, \ldots, x_{i-1})$
- neurosymbolic reasoning[2] $\mathbb{E}_{\mathbf{x}\sim p(\mathbf{x})}[\kappa(\mathbf{x})]$

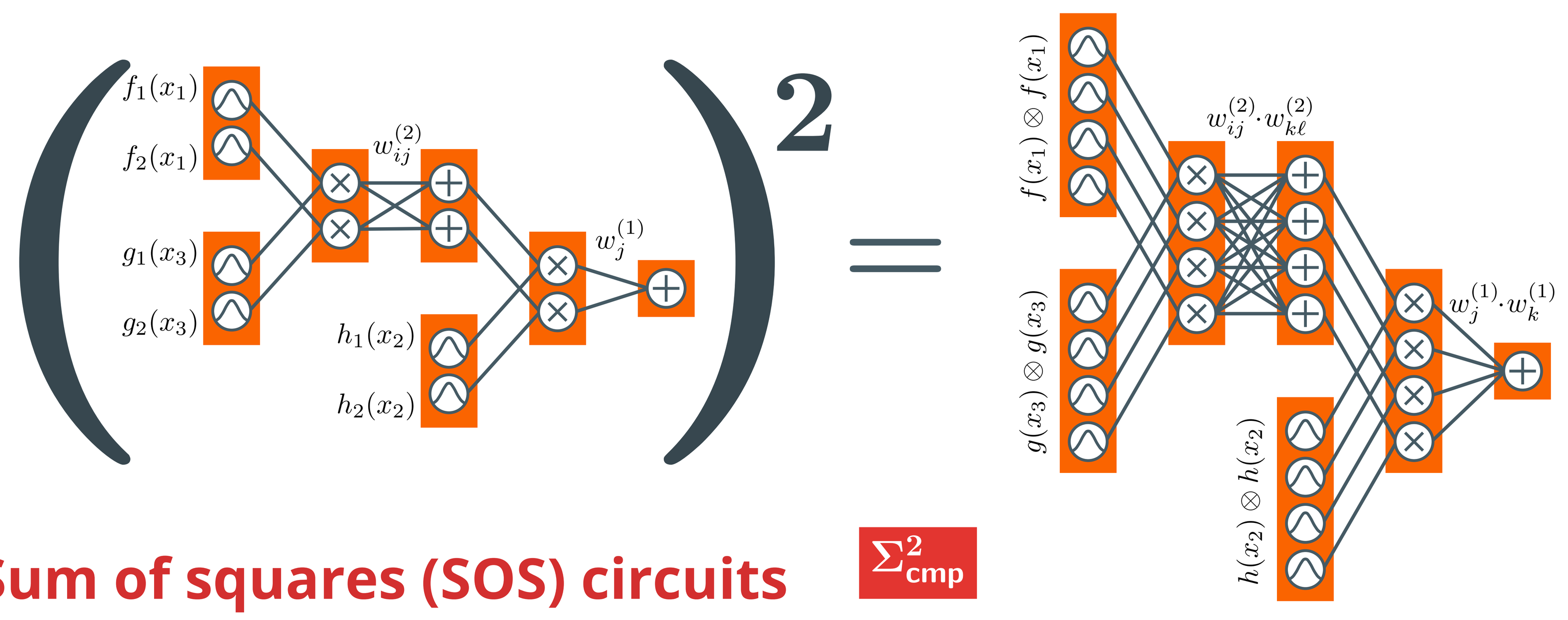
Monotonic circuits $+_{\mathsf{sd}}$

$p(\mathbf{x}) \propto c(\mathbf{x})$, with $c(\mathbf{x}) \geq 0$, positive parameters only



Squared circuits $\pm^2_{\mathbb{R}}$

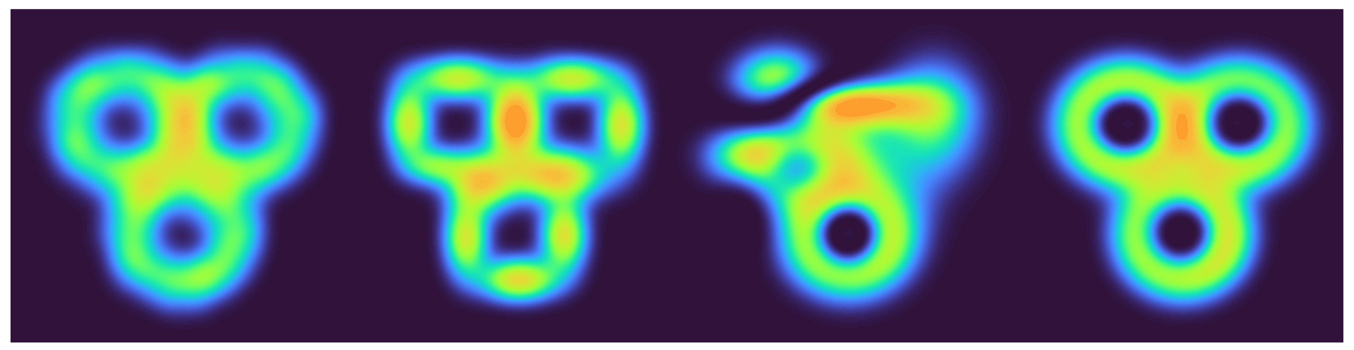
$p(\mathbf{x}) \propto c^2(\mathbf{x})$, with $c(\mathbf{x}) \in \mathbb{R}$, parameters are reals



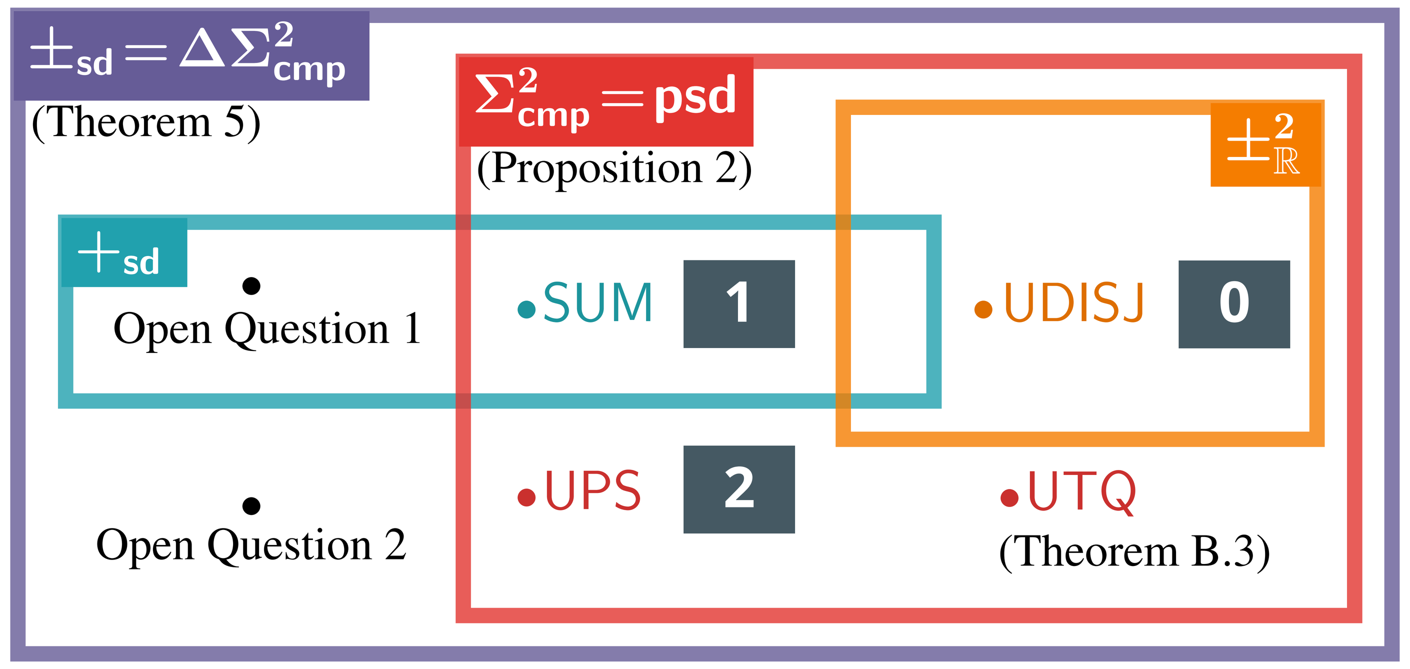
Sum of squares (SOS) circuits Σ^2_{cmp}

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

Why SOS?

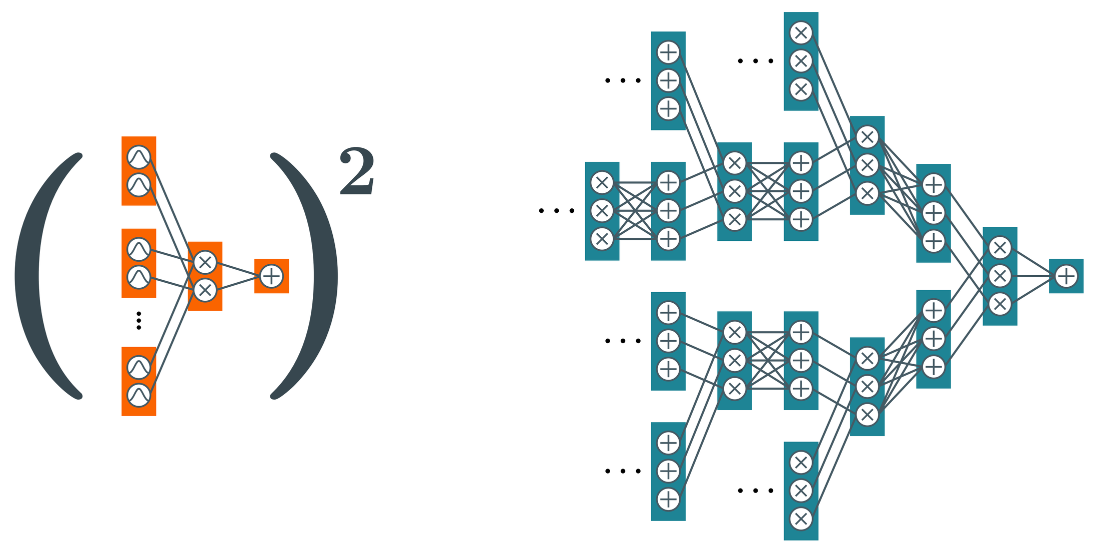


We build an expressiveness hierarchy around sum of squares circuits, with open questions!



0 Squared circuits are more expressive than monotonic circuits

Theorem 0 (Loconte et al. [3])
There exists distributions $\mathcal{F}$ over $\mathbf{X}$ that can be compactly represented as squared circuits ($\pm^2_{\mathbb{R}}$), but the smallest monotonic circuit ($+_{\mathsf{sd}}$) computing any $F \in \mathcal{F}$ has size $2^{\Omega(\sqrt{|\mathbf{X}|})}$.



We understand why other models are expressive: we show models PSD SNF BM are SOS circuits

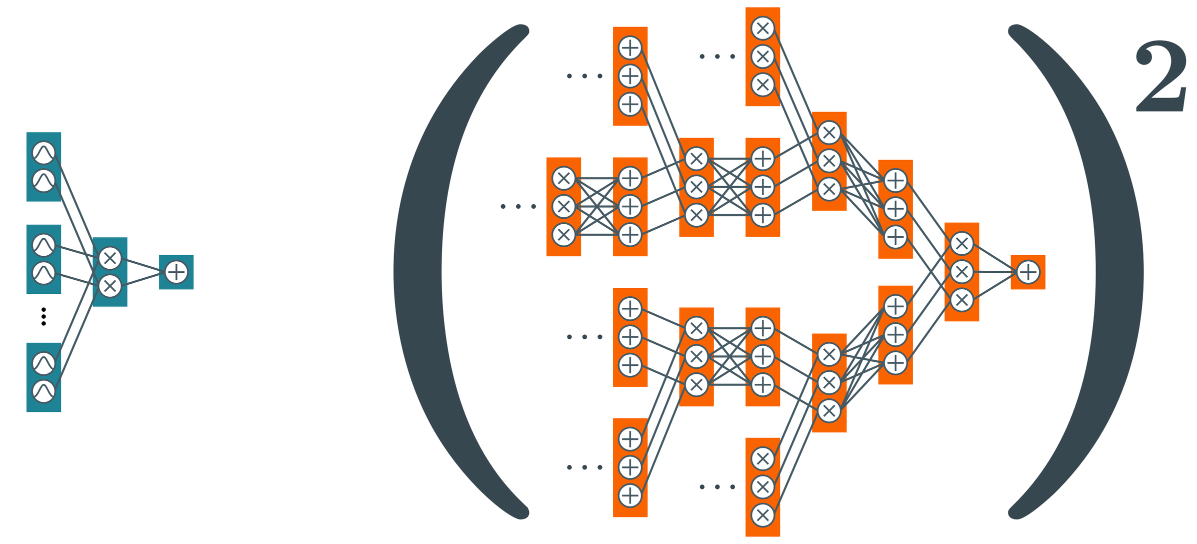
PSD Finite PSD models[4] $p(\mathbf{x}) \propto \boldsymbol{\kappa}(\mathbf{x})^\top \mathbf{A} \boldsymbol{\kappa}(\mathbf{x})$

SNF Squared neural family[5] $p(\mathbf{x}) \propto \mu(\mathbf{x}) ||\mathsf{nn}_\sigma(\mathbf{t}(\mathbf{x}))||_2^2$

BM Complex Born machines[6] $p(\mathbf{x}) \propto \psi(\mathbf{x})\psi(\mathbf{x})^\dagger$

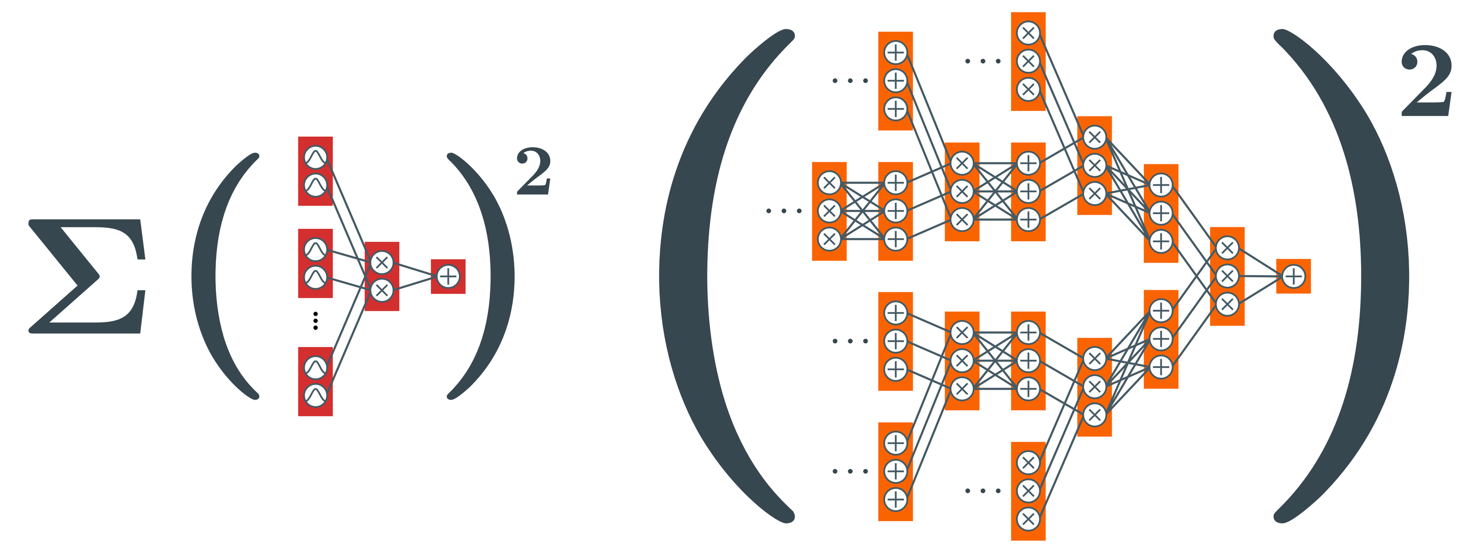
1 Monotonic and squared circuits are incomparable in terms of expressiveness

Theorem 1
There exists distributions $\mathcal{F}$ over $\mathbf{X}$ that can be compactly represented as monotonic circuits ($+_{\mathsf{sd}}$), but the smallest squared circuit ($\pm^2_{\mathbb{R}}$) computing any $F \in \mathcal{F}$ has size $2^{\Omega(\sqrt{|\mathbf{X}|})}$.

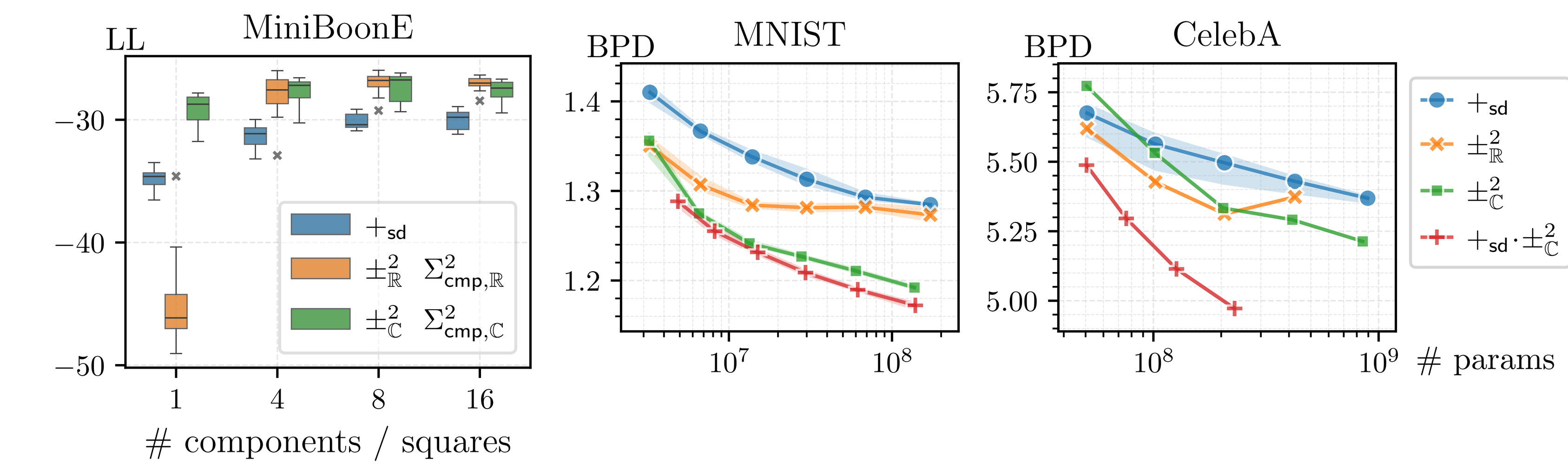


2 SOS circuits can be more expressive than both monotonic and squared circuits

Theorem 2
There exists distributions $\mathcal{F}$ over $\mathbf{X}$ that can be compactly represented as SOS circuits (Σ^2_{cmp}), but the smallest circuits in $+_{\mathsf{sd}}$ or in $\pm^2_{\mathbb{R}}$ computing any $F \in \mathcal{F}$ have size $2^{\Omega(\sqrt{|\mathbf{X}|})}$.



SOS circuits are expressive distribution estimators



References

[1] A. Liu, S. Mandt, and G. V. den Broeck. "Lossless Compression with Probabilistic Circuits". In: *ICLR*. 2022.
[2] K. Ahmed et al. "Semantic probabilistic layers for neuro-symbolic learning". In: *NeurIPS*. 2022.
[3] L. Loconte et al. "Subtractive Mixture Models via Squaring: Representation and Learning". In: *ICLR*. 2024.
[4] A. Rudi and C. Ciliberto. "PSD Representations for Effective Probability Models". In: *NeurIPS*. 2021.
[5] R. Tsuchida, C. S. Ong, and D. Sejdinovic. "Squared Neural Families: A New Class of Tractable Density Models". In: *NeurIPS*. 2023.
[6] I. Glasser et al. "Expressive power of tensor-network factorizations for probabilistic modeling". In: *NeurIPS*. 2019.