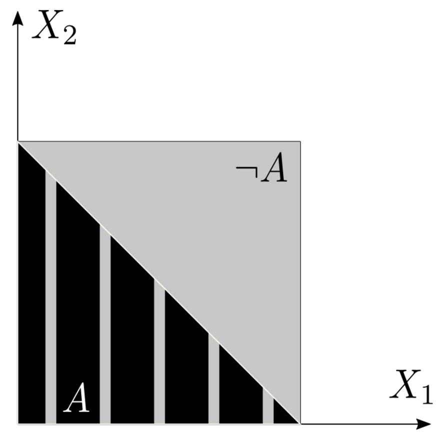
# Is Parameter Learning via Weighted Model Integration Tractable?

Zhe Zeng*[1] Paolo Morettin*[2] Fanqi Yan[3]
Antonio Vergari[1] Andrea Passerini[4] Guy Van den Broeck[1]
[1]University of California, Los Angeles [2]KU Leuven [3]University of Texas, Austin [4]University of Trento

## WMI Problem

Formula

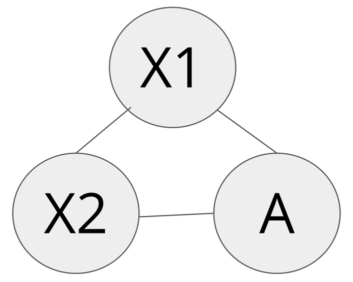
$\Delta = (0 \leq X_1) \wedge (X_1 \leq 1)$
$\wedge (0 < X_2) \wedge (X_2 \leq 1)$
$\wedge (A \to (X_1 + X_2 \leq 1))$




Weight

$w_\ell(X_1, X_2) = X_1 + X_2$
*with* $\ell = (X_1 + X_2 \leq 1)$

Primal Graph




## Weighted Model Integration (WMI)

A framework for hybrid probabilistic inference with algebraic constraints [1].

$$\mathsf{WMI}(\Delta, w; \mathbf{X}, \mathbf{B}) \triangleq \sum_{\mathbf{b} \in \mathbb{B}^{|\mathbf{B}|}} \int_{\Delta(\boldsymbol{x}, \boldsymbol{b})} w(\boldsymbol{x}, \boldsymbol{b})\, d\boldsymbol{x}$$

i.e. integration over the weighted satisfying assignments of a formula over hybrid variables

**Algebraic constraints** are Satisfiability Modulo Theories (SMT) formulas, as combinations of *Boolean* literals & *linear real arithmetic* (LRA) literals in Conjunctive Normal Form (CNF).

**Primal graphs** for SMT(LRA) formula:
nodes ~ variables, edges ~ clauses.

**Per-literal weights** assign weight if the literal is SAT; otherwise assign one.
Together they define a **joint weight**.

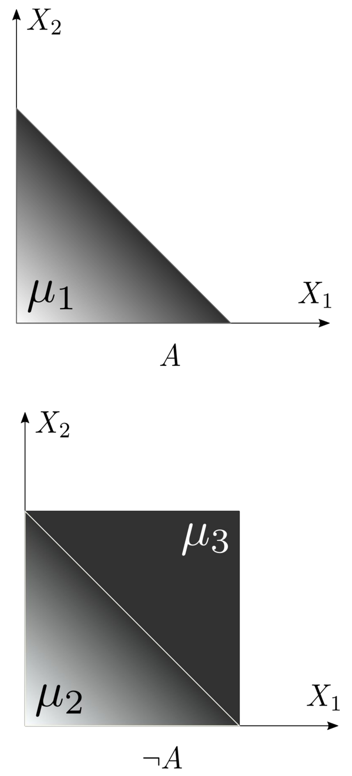
**Answer queries**

$$\Pr(q) = \frac{\mathsf{WMI}(\Delta \wedge q, w; \mathbf{X}, \mathbf{B})}{\mathsf{WMI}(\Delta, w; \mathbf{X}, \mathbf{B})}$$

## Compute WMI

$\mathsf{WMI}(\Delta, \mathcal{W})$

$$= \int_{\mu_1} X_1 + X_2\, dX_1\, dX_2$$
$$+ \int_{\mu_2} X_1 + X_2\, dX_1\, dX_2$$
$$+ \int_{\mu_3} 1\, dX_1\, dX_2 = 2 \cdot \frac{1}{3} + \frac{1}{2}$$




Probabilistic Inference

$$Pr(A) = \frac{\mathsf{WMI}(\Delta \wedge A, \mathcal{W})}{\mathsf{WMI}(\Delta, \mathcal{W})}$$

## Contributions

- Provide theoretical insights on tractability of MLE-based parameter learning of WMI.
- Bridging two fields: Hinge-Loss Markov Random Fields (HL-MRFs) and WMI models, by reducing marginal inference of HL-MRFs to WMI inference.

## Maximum Likelihood Estimation (MLE)

### *for WMI parameters*

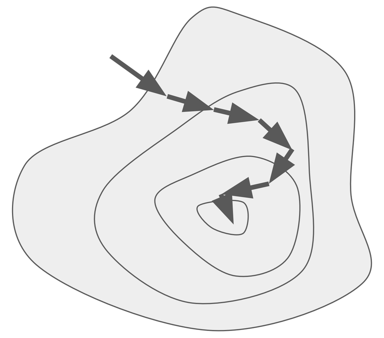
**Optimization objective** log-likelihood of dataset

$$\boldsymbol{\Theta}^* = \arg\max_{\boldsymbol{\Theta}} L(\boldsymbol{\Theta}) = \arg\max_{\boldsymbol{\Theta}} \log \prod_{x \in \mathcal{D}} p(x; \boldsymbol{\Theta})$$

**Convexity?**

**Prop.** The log-likelihood of dataset is ***concave*** if the weight functions are ***log-linear*** in their parameters.

=> This gives guarantees to convergence to the global optimum when using iterative methods, such as gradient ascent.



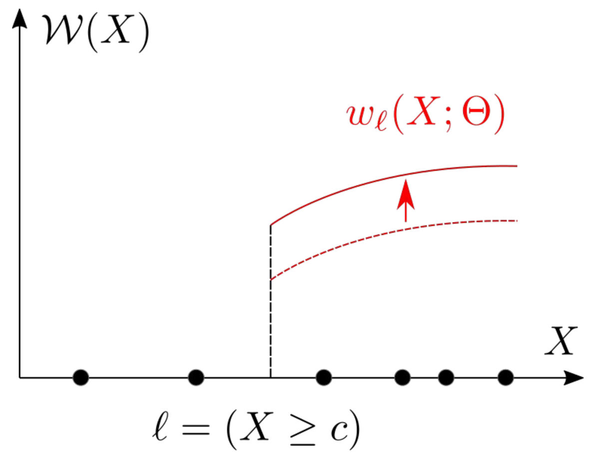
## MLE

$$\boldsymbol{\Theta}^* = \arg\max_{\boldsymbol{\Theta}} L(\boldsymbol{\Theta}; \mathcal{D})$$

***Convexity?***

$$\frac{\partial}{\partial \Theta_\ell^k} L(\boldsymbol{\Theta})$$

***Tractability?***




***Tractability of computing gradients?***

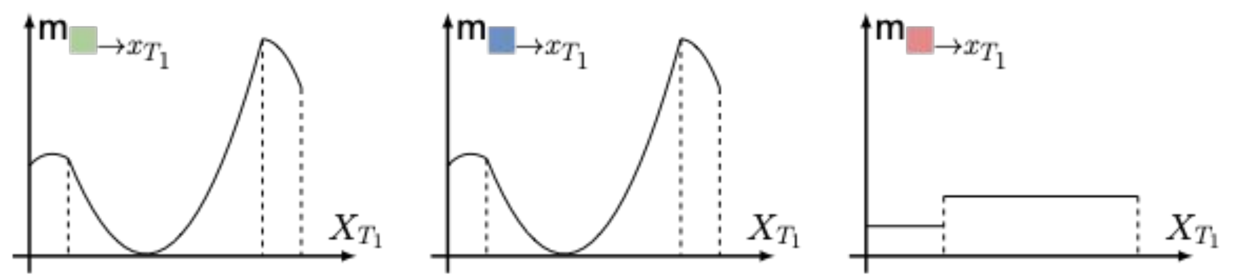
**Partial derivatives** of the log-likelihood with respect to a parameter in log-linear weight function:

$$\frac{\partial}{\partial \Theta_\ell^k} L(\boldsymbol{\Theta}) = \mathbb{E}_{\boldsymbol{x} \sim \mathcal{D}} [\![ \boldsymbol{x} \models \ell ]\!] \cdot f_\ell^k(\boldsymbol{x})]$$
$$- \mathbb{E}_{\boldsymbol{x} \sim p(\boldsymbol{x}; \boldsymbol{\Theta})} [[\![ \boldsymbol{x} \models \ell ]\!] \cdot f_\ell^k(\boldsymbol{x})]$$

**Thm.** The computation of the partial derivative is ***tractable*** if
*i)* WMI model is in the tractable WMI problem class [1];
*ii)* the feature functions are in function families that satisfies *tractable weight conditions (TWCs)*,

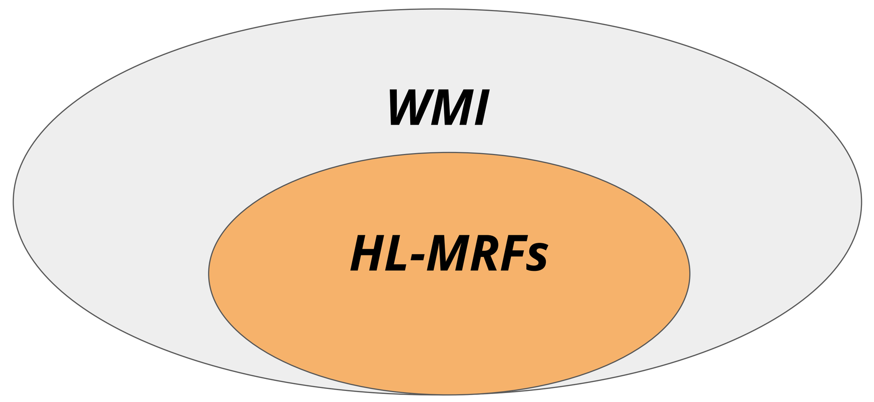
where A family of weight functions satisfies ***TWCs*** iff:
- It is closed under *product*;
- It admits efficient computation of *antiderivatives*;
- It is closed under *definite integration.*

For example, (piecewise) polynomial, log-linear functions ...




=> This characterize which weight family allows tractable computation of the exact partial derivatives.

## WMI & PSL




## Marginal Inference For HL-MRFs via WMI

**Probabilistic Soft Logic (PSL)** is a statistical relational learning (SRL) framework for modeling probabilistic and relational domains, e.g.,

$$0.3 : \mathrm{friend}(A, B) \wedge \mathrm{votesFor}(B, P)$$

A PSL program induces a **Hinge-Loss Markov Random Field (HL-MRF)**, which defines the distribution over interpretations

$$Pr(x_1, x_2) = \exp\{-0.3 \max\{x_1 + x_2 - 1, 0\}\}$$

**Thm.** For any HL-MRF, there exists a WMI model with per-literal weights whose WMI density equals to the HL-MRF density.

For example, the WMI model that is equal to the HL-MRF shown above is

Formula $(x_1 + x_2 - 1 \geq 0) \vee \texttt{True}$

Weight $w_\ell(x_1, x_2) = \exp\{-0.3(x_1 + x_2 - 1)\}$

=> This allows us to characterize the tractability of marginal inference for HL-MRFs via the analysis for WMI.

### References

[1] Zhe Zeng, Paolo Morettin, Fanqi Yan, Antonio Vergari, and Guy Van den Broeck. *Probabilistic inference with algebraic constraints: Theoretical limits and practical approximations*. NeurIPS, 33, 2020.

TPM Workshop at *UAI 2021*