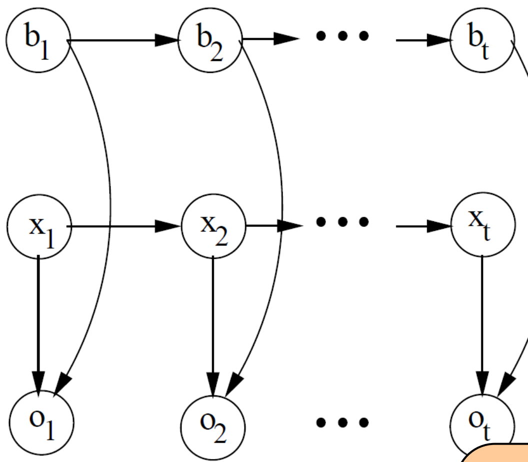
Symbolic Variable Elimination in Discrete and Continuous Graphical Models

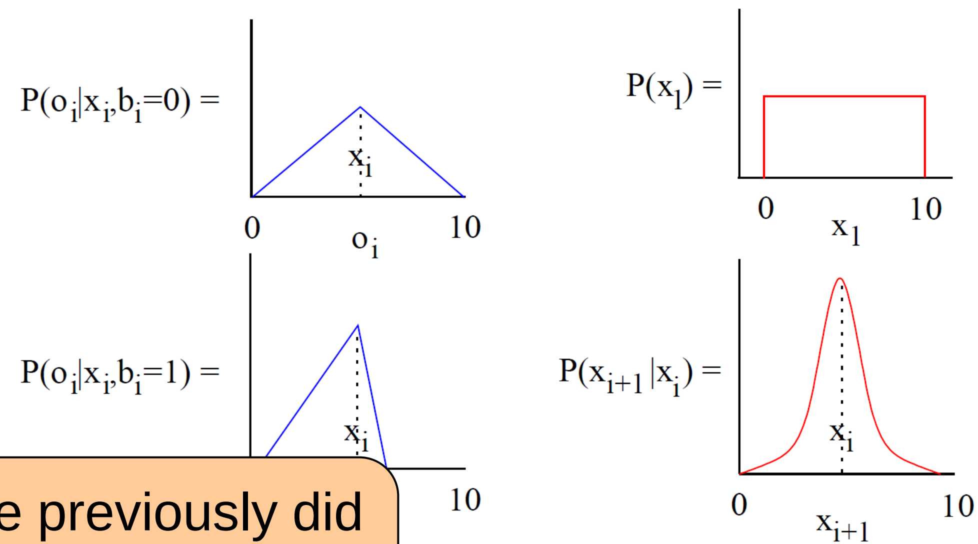
Scott Sanner
Ehsan Abbasnejad

NICTA

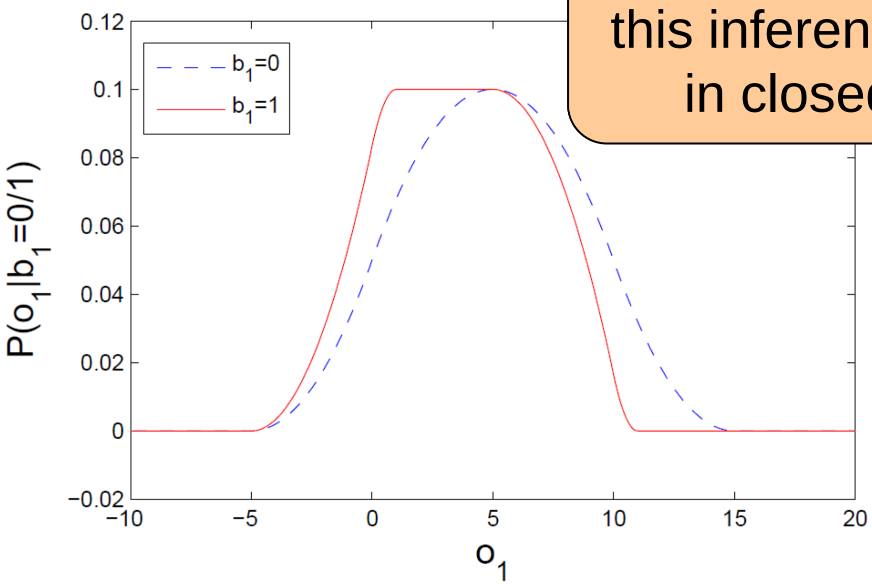
THE AUSTRALIAN NATIONAL UNIVERSITY

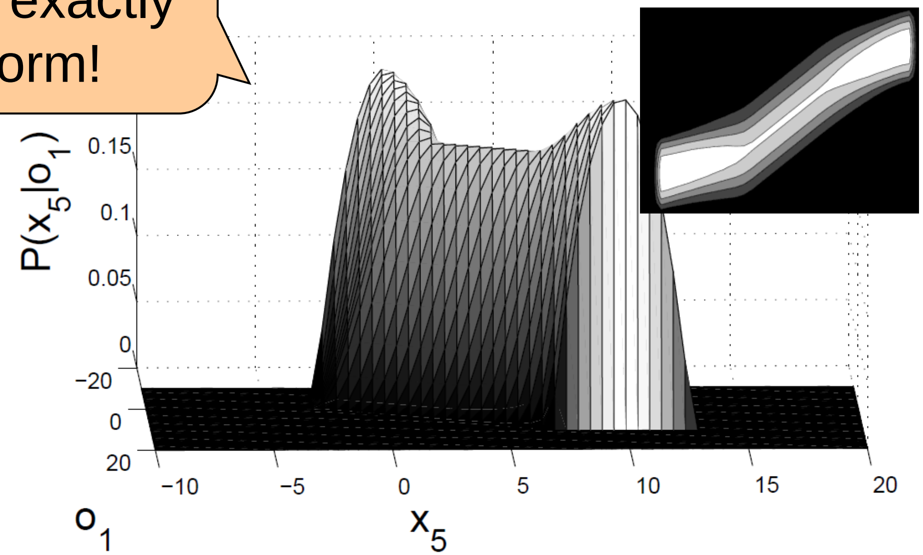
Inference for Dynamic Tracking





No one previously did this inference exactly in closed-form!





Exact Closed-form Continuous Inference

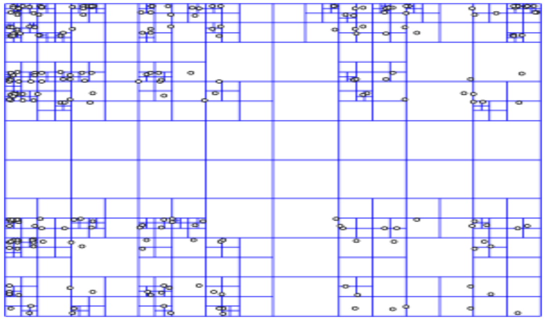
- Fully Gaussian
  - Most inference including conditional
- Fully Uniform
  - 1D, n-D hyperrectangular cases

Yes, but not a solution you can write on 1 sheet of paper

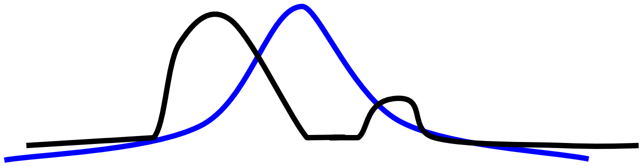
- General Uniform
- Piecewise, Asymmetrical
  - Exact (conditional) inference possible in closed-form?

Already Solved?

- How is inference done in piecewise models?
  - (Adaptively) discretize model:
    - Approximate, $O(N^D)$
    - Adaptivity is an artform



  - Projective approximation: variational, EP
    - Choose approximating class
    - Often Gaussian – good luck!



  - Sampling: Monte Carlo… Gibbs
    - May not converge for (near) deterministic distributions
    - Not for unobserved evidence – E[x | o=?]

No expressive, exact, closed-form solutions! Yet.

What has everyone been missing?

Symbolic representations and operations on piecewise functions

Continuous variant of work started in (Boutilier, Reiter, Price, IJCAI-01)

Piecewise Functions (Cases)

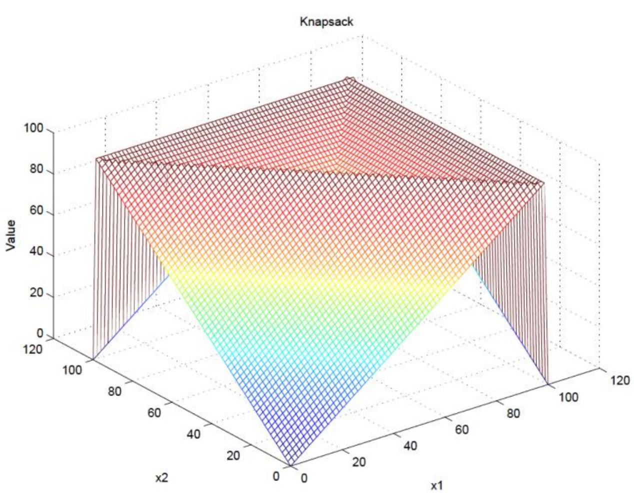
$$z = f(x,y) = \begin{cases} \boxed{(x > 3) \wedge (y \cdot \ x) : \quad x + y} \\ (x \cdot \ 3) \vee (y > x) : \quad x^2 + xy^3 \end{cases}$$

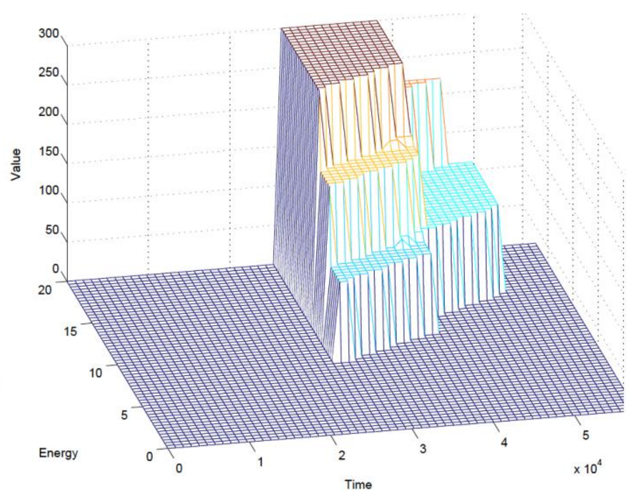
Partition

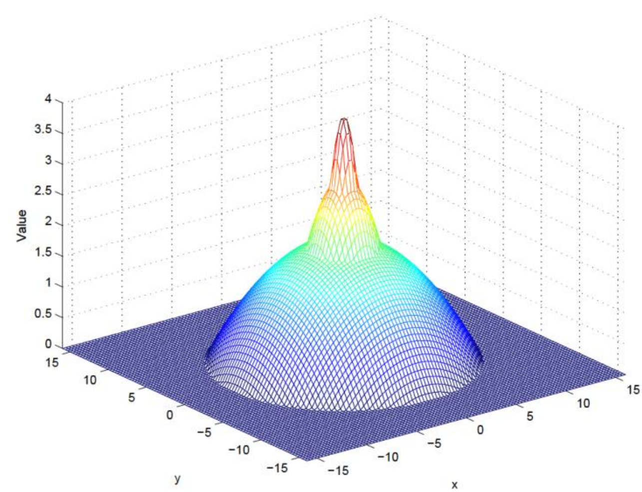
Constraint

Value

Linear constraints and value	Linear constraints, constant value	Quadratic constraints and value







Formal Problem Statement

- General continuous graphical models represented by piecewise functions (cases)

$$f = \begin{cases} \phi_1 : & f_1 \\ \vdots & \vdots \\ \phi_k : & f_k \end{cases}$$

- For $\sum \prod$ variable elimination, need exact closed-form solution inferred via the following piecewise calculus:
  - $f_1 \oplus f_2$, $f_1 \otimes f_2$
  - $\max(f_1, f_2)$, $\min(f_1, f_2)$
  - $\int_x f(x)$

Question: how do we perform these operations in closed-form?

Polynomial Case Operations: $\oplus$, $\otimes$

$$\begin{cases} \phi_1 : & f_1 \\ \phi_2 : & f_2 \end{cases} \oplus \begin{cases} \psi_1 : & g_1 \\ \psi_2 : & g_2 \end{cases} = \qquad ?$$

Polynomial Case Operations: ⊕, ⊗

$$\begin{cases} \phi_1 : & f_1 \\ \phi_2 : & f_2 \end{cases} \oplus \begin{cases} \psi_1 : & g_1 \\ \psi_2 : & g_2 \end{cases} = \begin{cases} \phi_1 \wedge \psi_1 : & f_1 + g_1 \\ \phi_1 \wedge \psi_2 : & f_1 + g_2 \\ \phi_2 \wedge \psi_1 : & f_2 + g_1 \\ \phi_2 \wedge \psi_2 : & f_2 + g_2 \end{cases}$$

- Similarly for ⊗
  - Polynomials closed under +, *
- What about max?
  - Max of polynomials is not a polynomial ☹

Polynomial Case Operations: max

$$\max\left(\begin{cases} \phi_1 : & f_1 \\ \phi_2 : & f_2 \end{cases}, \begin{cases} \psi_1 : & g_1 \\ \psi_2 : & g_2 \end{cases} \right) = \qquad ?$$

Polynomial Case Operations: max

$$
\max\left(\begin{cases} \phi_1 : & f_1 \\ \phi_2 : & f_2 \end{cases}, \begin{cases} \psi_1 : & g_1 \\ \psi_2 : & g_2 \end{cases} \right) = \begin{cases} \phi_1 \wedge \psi_1 \wedge f_1 > g_1 : & f_1 \\ \phi_1 \wedge \psi_1 \wedge f_1 \cdot g_1 : & g_1 \\ \phi_1 \wedge \psi_2 \wedge f_1 > g_2 : & f_1 \\ \phi_1 \wedge \psi_2 \wedge f_1 \cdot g_2 : & g_2 \\ \phi_2 \wedge \psi_1 \wedge f_2 > g_1 : & f_2 \\ \phi_2 \wedge \psi_1 \wedge f_2 \cdot g_1 : & g_1 \\ \phi_2 \wedge \psi_2 \wedge f_2 > g_2 : & f_2 \\ \phi_2 \wedge \psi_2 \wedge f_2 \cdot g_2 : & g_2 \end{cases}
$$

- Still a piecewise polynomial!

Size blowup?
We'll get to that…

Integration: $\int_x$

- $\int_x$ closed for polynomials
  - But how to compute for case?

$$\int_x \begin{cases} \phi_1 : & f_1 \\ \vdots & \vdots \\ \phi_k : & f_k \end{cases} dx = \int_x \sum_{i=1}^{k} [\phi_i] \cdot f_i \, dx$$

$$= \sum_i \int_x [\phi_i] \cdot f_i \, dx$$

  - Just integrate case partitions, $\oplus$ results!

Partition Integral

1. Determine integration bounds

$$\int_x [\phi_1] \cdot f_1 \, dx$$

$$\phi_1 := [x > -1] \wedge [x > y - 1] \wedge [x \cdot \; z] \wedge [x \cdot \; y + 1] \wedge [y > 0]$$

$$f_1 := x^2 - xy$$

$$LB := \begin{cases} y - 1 > -1 : & y - 1 \\ y - 1 \cdot \; -1 : & -1 \end{cases} \qquad UB := \begin{cases} z < y + 1 : & z \\ z \geq y + 1 : & y + 1 \end{cases}$$

UB and LB are symbolic!

$$[\phi_{cons}] \int_{x = LB}^{UB} f_1 \, dx$$

What constraints here?
- independent of x
- pairwise UB > LB

How to evaluate?

Definite Integral Evaluation

- How to evaluate integral bounds?

$$\int_{x=LB}^{UB} x^2 - xy \;=\; \frac{1}{3}x^3 - \frac{1}{2}x^2 y \Big|_{LB}^{UB}$$

$$LB := \begin{cases} y-1 > -1 : & y-1 \\ y-1 \cdot\; -1 : & -1 \end{cases} \qquad UB := \begin{cases} z < y+1 : & z \\ z \geq y+1 : & y+1 \end{cases}$$

- Can do polynomial operations on cases!

$$f_1\Big|_{LB}^{UB} = \left[\frac{1}{3}UB \quad UB \quad UB \ominus \frac{1}{2}UB \quad UB \quad (y)\right]$$

$$\ominus \left[\frac{1}{3}LB \quad LB \quad LB \ominus \frac{1}{2}LB \quad LB \quad (y)\right]$$

Symbolically, exactly evaluated!

Exact Graphical Model Inference!

(directed and undirected)

- Can do general probabilistic inference

$$p(x_2|x_1) = \frac{\int_{x_3} \cdots \int_{x_n} \bigotimes_{i=1}^{k} case_i \; dx_n \cdots dx_3}{\int_{x_2} \cdots \int_{x_n} \bigotimes_{i=1}^{k} case_i \; dx_n \cdots dx_2}$$

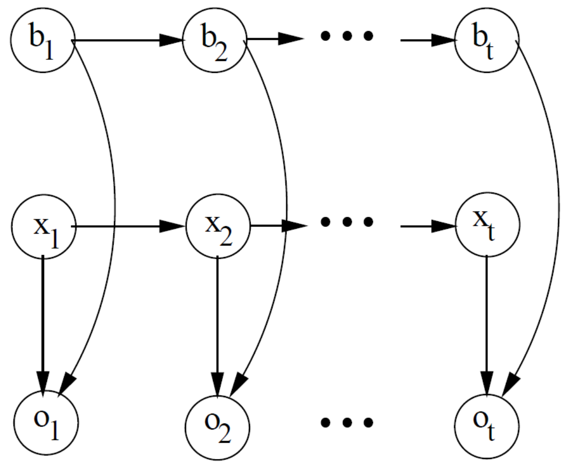
- Or an exact expectation of *any* polynomial

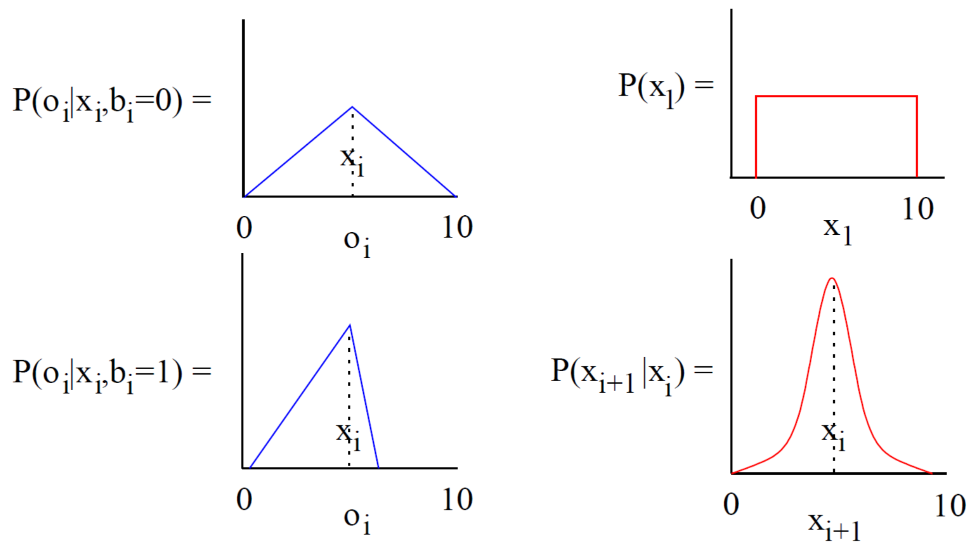
$$\mathbb{E}_{\mathbf{x} \sim p(\mathbf{x}|\mathbf{o})}[poly(\mathbf{x})|\mathbf{o}] = \int_{\mathbf{x}} p(\mathbf{x}|\mathbf{o}) poly(\mathbf{x}) d\mathbf{x}$$

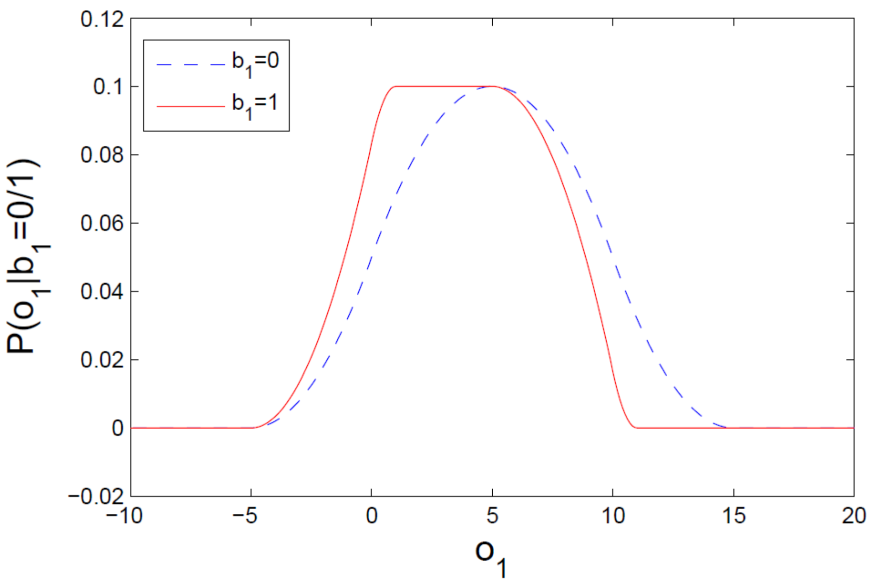
  – *poly* = Mean, variance, skew, curtosis, …, x^2+y^2+xy

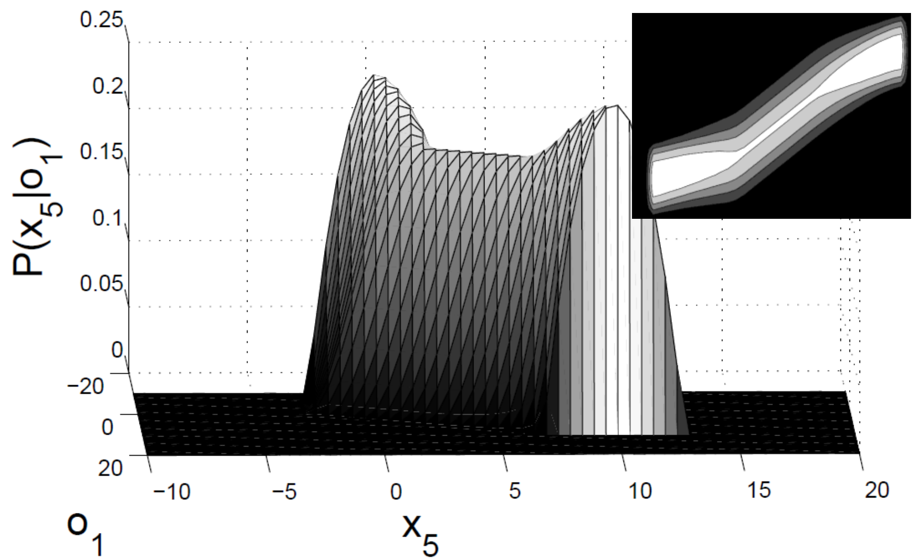
All computed by **Symbolic Variable Elimination (SVE)**

Voila: Closed-form Exact Inference via SVE!

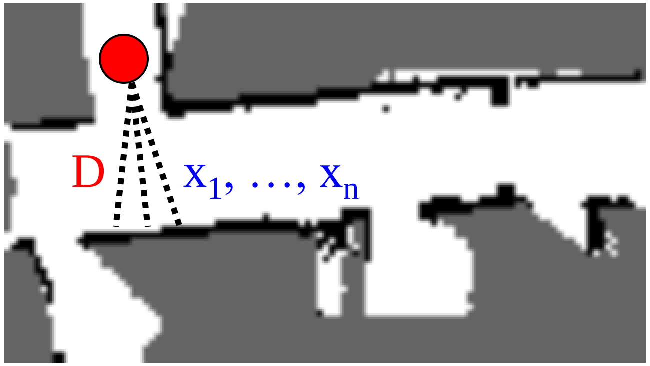




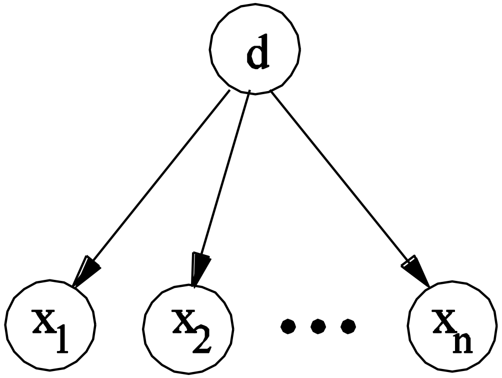


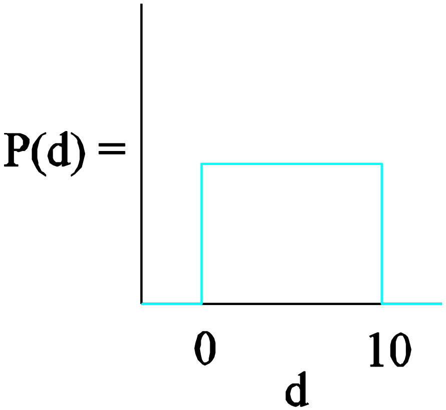


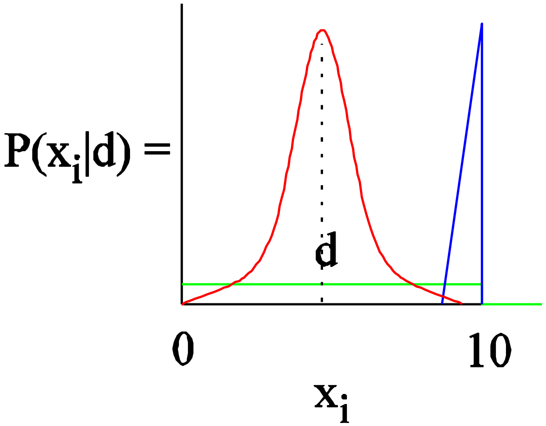
Bayesian Robotics



- D: true distance to wall
- $x_1, \ldots, x_n$: measurements
- want: $E[D \mid x_1,\ldots,x_{n-1}, x_n=?]$

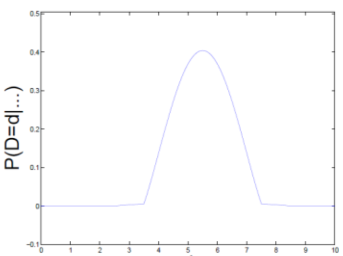




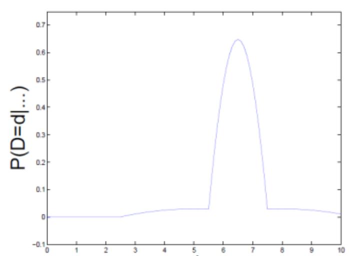


Bayesian Robotics: Inference Examples

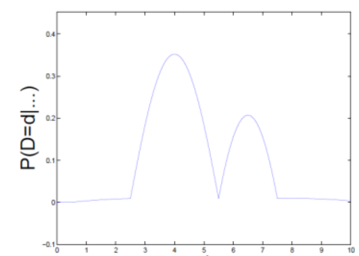
- Posterior distance shown in graph
- Expected distance calculation shown below



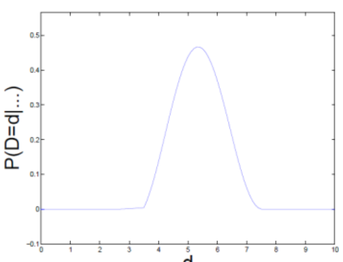
$$\mathbb{E}[D|x_1 = 6, x_2 = 5] = 4.98$$



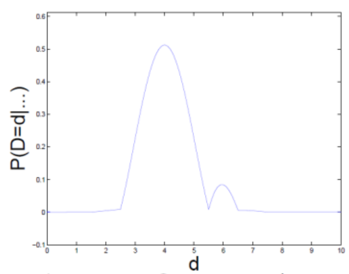
$$\mathbb{E}[D|x_1 = 8, x_2 = 5] = 6.0$$



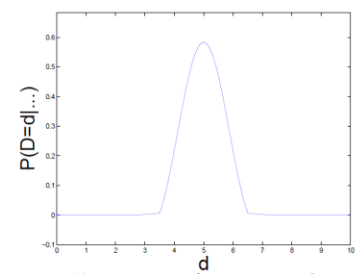
$$\mathbb{E}[D|x_1 = 5, x_2 = 3, x_3 = 8] = 4.39$$



$$\mathbb{E}[D|x_1 = 5, x_2 = 5, x_3 = 6] = 4.98$$



$$\mathbb{E}[D|x_2 = 1, x_2 = 3, x_3 = 4, x_4 = 8] = 5.45$$



$$\mathbb{E}[D|x_1 = 5, x_2 = 4, x_3 = 6, x_4 = 5] = 4.89$$

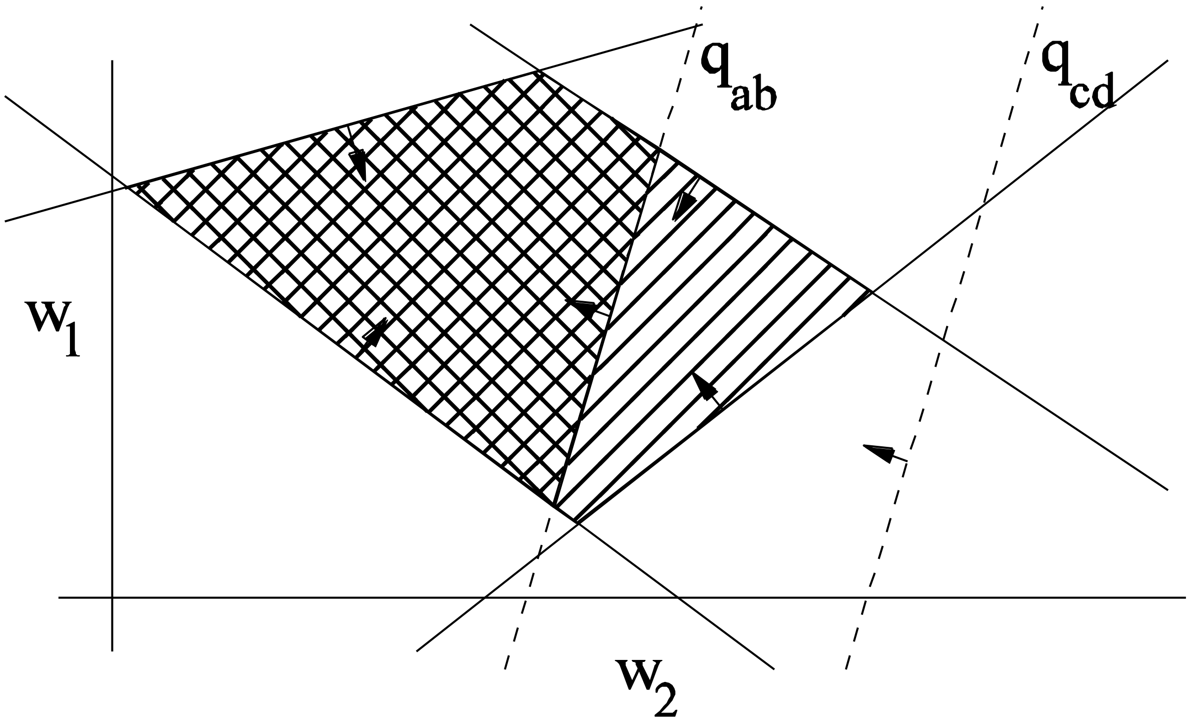
Extra: A New Conjugate Prior for Bayesian Inference

- General Bayesian Inference

$$p(\vec{\theta}|D_{n+1}) \propto p(d_{n+1}|\vec{\theta})p(\vec{\theta}|D_n)$$

- Prior and likelihood are case statements
  - ➔ then posterior is a case statement

Example: Bayesian Preference Learning for Linear Utilities



- If prior is uniform over weights
  and each preference likelihood linearly constrains weights
  ➔ posterior is uniform over convex polytope

- All case statements!

Inference is equivalent to finding volume of a convex polytope!

Computational Complexity?

- In theory for SVE on graphical models
  - Best-case complexity Ω(#operations)
  - Worst-case complexity is O(exp(#operations))
    - Not explicitly tree-width dependent!
    - **But worse:** integral may invoke 100's of operations!

Fortunately data structures mitigate worst-case complexity

BDD / ADDs

Quick Introduction

Function Representation (Tables)

- How to represent functions: $B^n \rightarrow R$?
- How about a fully enumerated table…
- …OK, but can we be more compact?

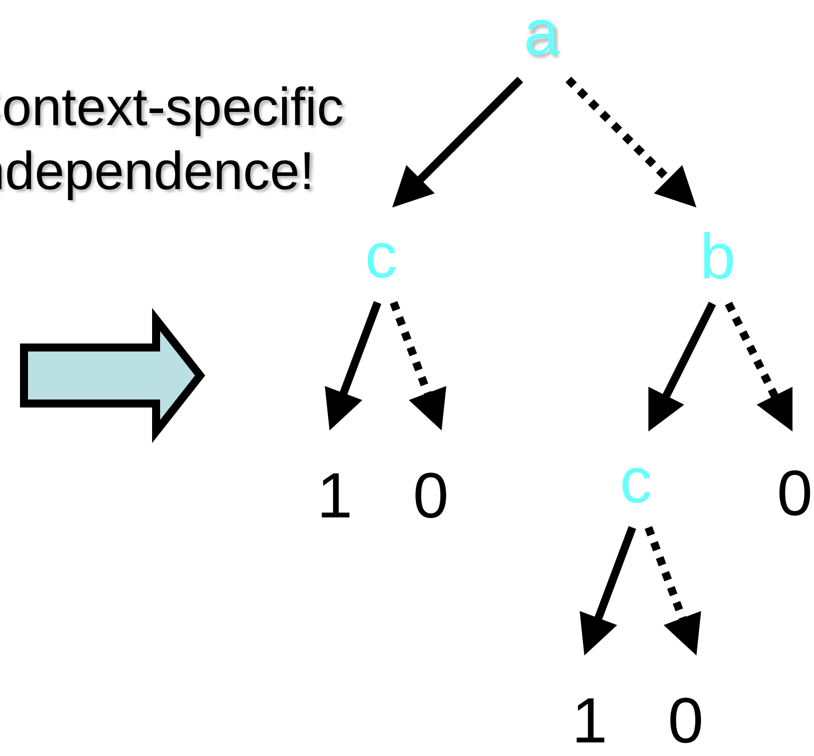
a	b	c	F(a,b,c)
0	0	0	0.00
0	0	1	0.00
0	1	0	0.00
0	1	1	1.00
1	0	0	0.00
1	0	1	1.00
1	1	0	0.00
1	1	1	1.00

Function Representation (Trees)

- How about a tree? Sure, can simplify.

a	b	c	F(a,b,c)
0	0	0	0.00
0	0	1	0.00
0	1	0	0.00
0	1	1	1.00
1	0	0	0.00
1	0	1	1.00
1	1	0	0.00
1	1	1	1.00

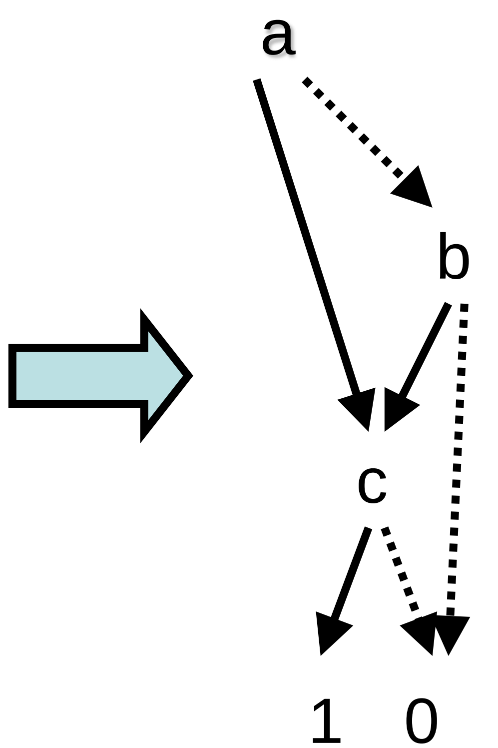
Context-specific independence!



Function Representation (ADDs)

- Why not a directed acyclic graph (DAG)?

a	b	c	F(a,b,c)
0	0	0	0.00
0	0	1	0.00
0	1	0	0.00
0	1	1	1.00
1	0	0	0.00
1	0	1	1.00
1	1	0	0.00
1	1	1	1.00

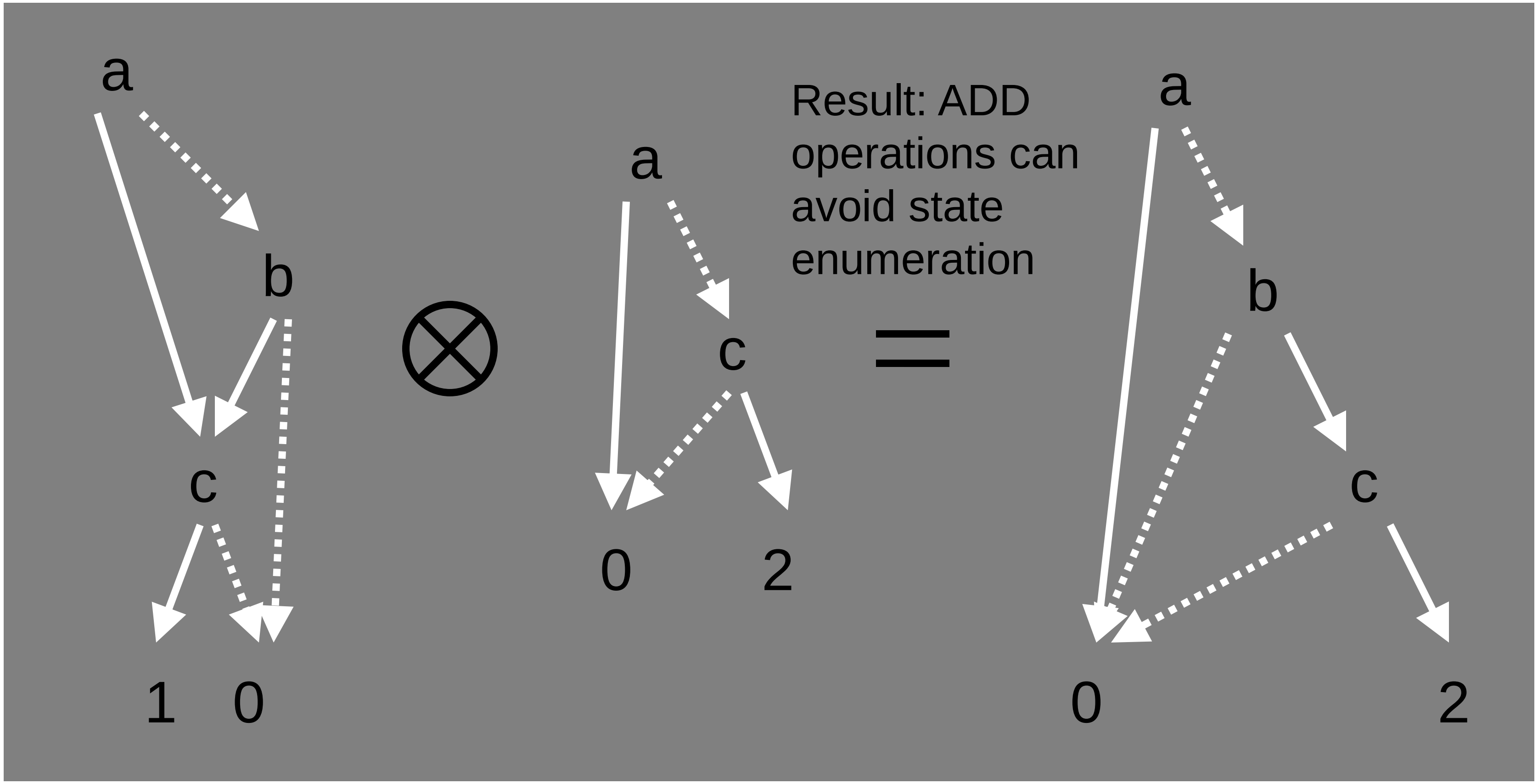


Algebraic Decision Diagram (ADD)

Think of BDDs as {0,1} subset of ADD range

Binary Operations (ADDs)

- Why do we order variable tests?
- Enables us to do efficient binary operations…



Case $\rightarrow$ XADD

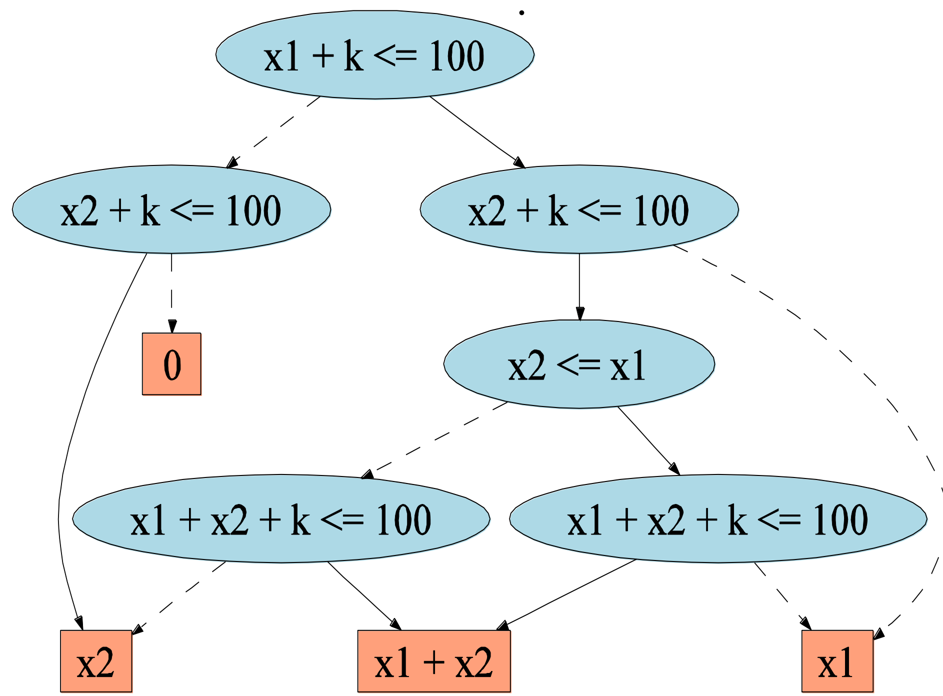
XADD = continuous variable e**x**tension
of **a**lgebraic **d**ecision **d**iagram

Efficient XADD data structure for cases
• strict ordering of atomic inequality tests

➔ compact, minimal case representation
➔ efficient case operations

Case → XADD

$$V = \begin{cases} x_1 + k > 100 \wedge x_2 + k > 100 : & 0 \\ x_1 + k > 100 \wedge x_2 + k \leq 100 : & x_2 \\ x_1 + k \leq 100 \wedge x_2 + k > 100 : & x_1 \\ x_1 + x_2 + k > 100 \wedge x_1 + k \leq 100 \wedge x_2 + k \leq 100 \wedge x_2 > x_1 : & x_2 \\ x_1 + x_2 + k > 100 \wedge x_1 + k \leq 100 \wedge x_2 + k \leq 100 \wedge x_2 \leq x_1 : & x_1 \\ x_1 + x_2 + k \leq 100 : & x_1 + x_2 \\ \vdots & \vdots \end{cases}$$

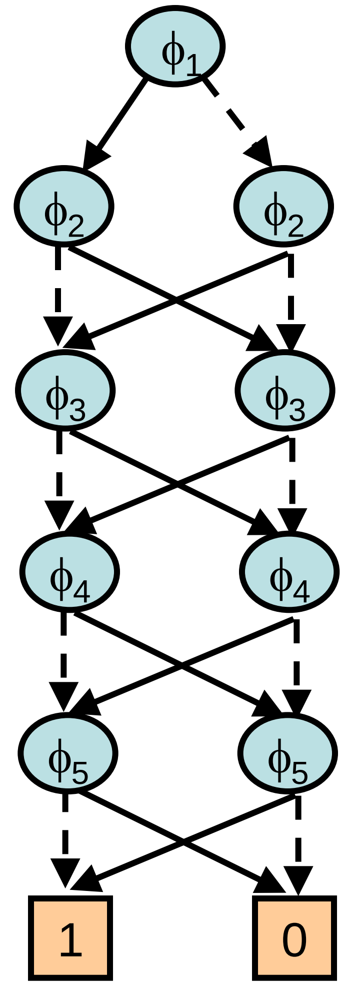


*With non-trivial extensions over ADD, can reduce to a minimal canonical form!

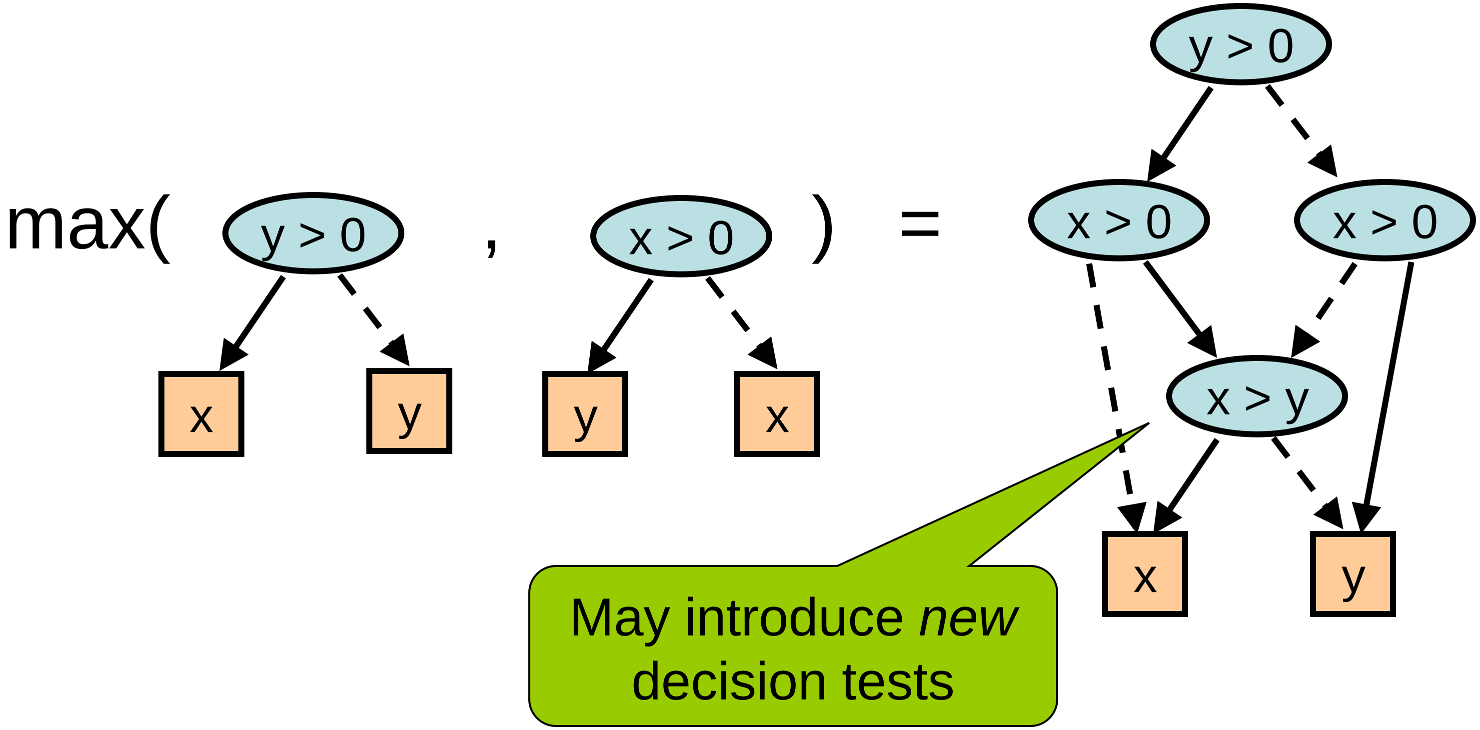
Compactness of (X)ADDs

- Linear in number of decisions ϕ_i

- Case version has exponential number of partitions!



XADD Maximization



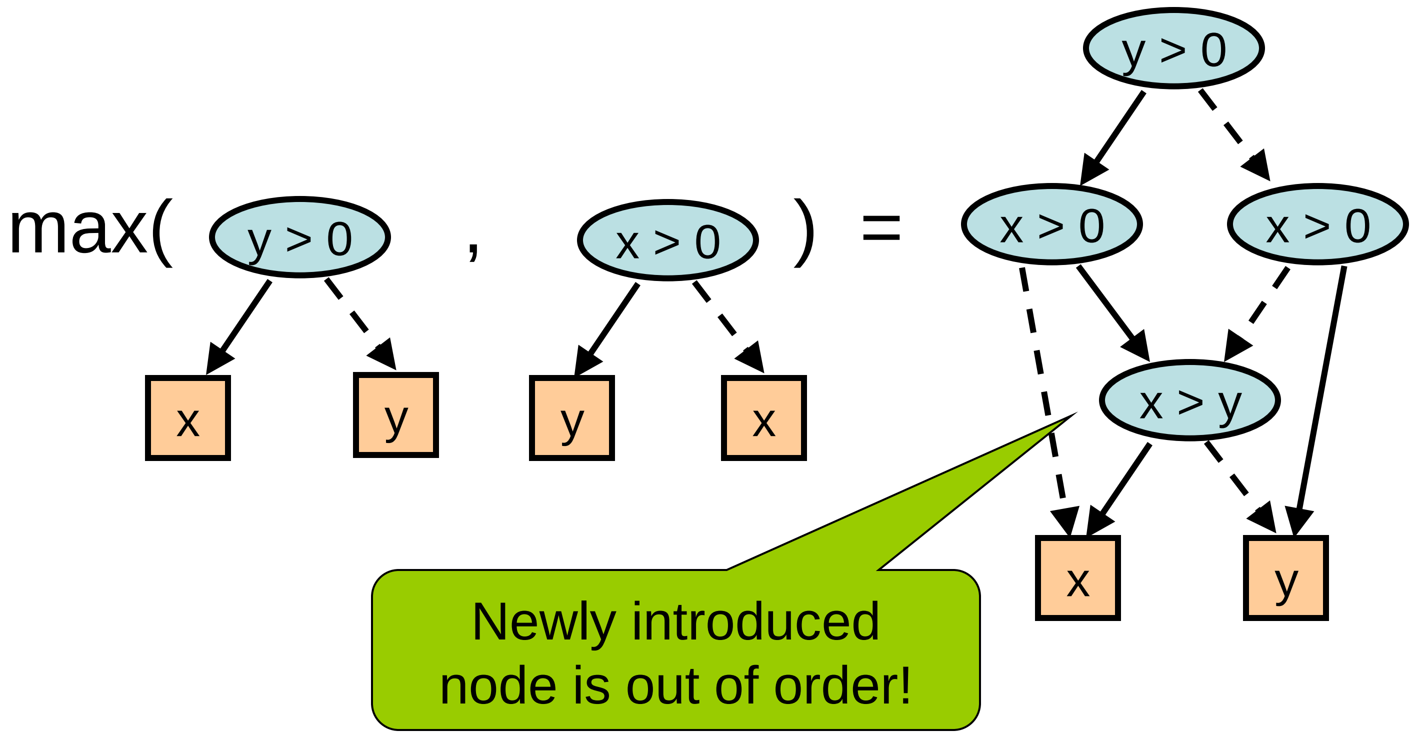
Operations exploit structure: O(|f||g|)

Maintaining XADD Orderings

- Max may get decisions out of order

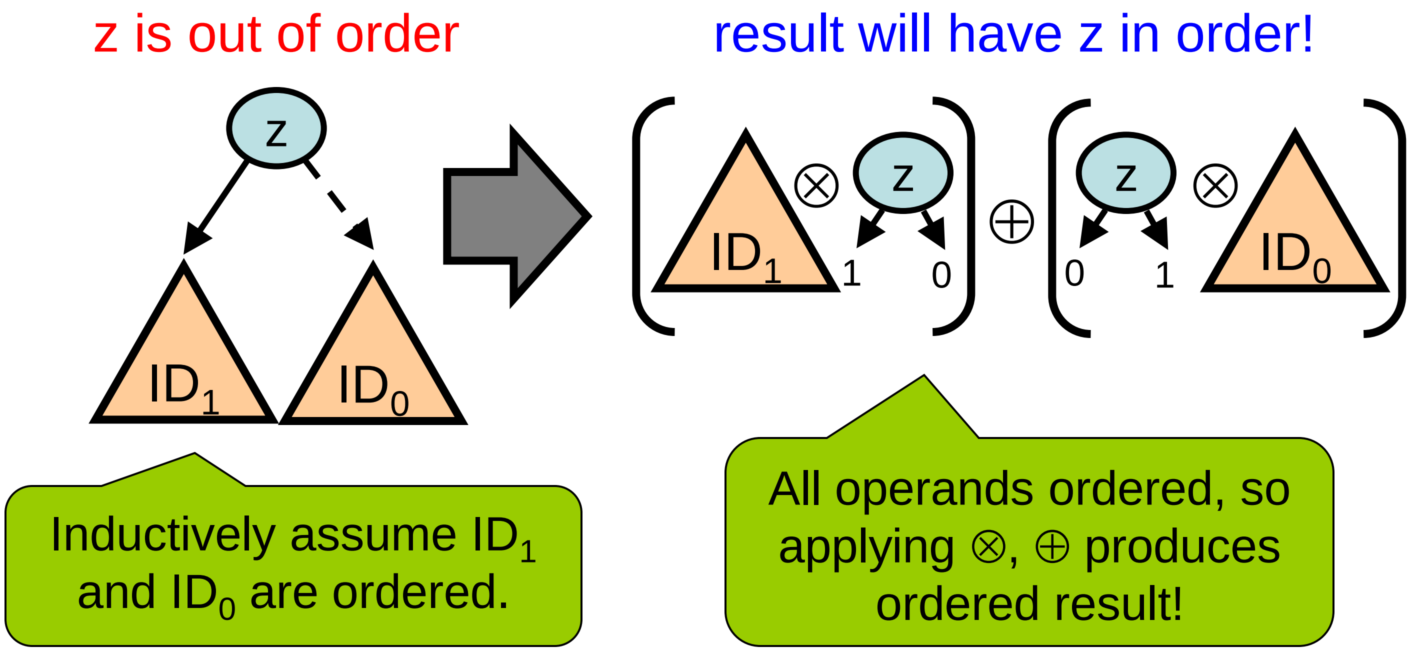
Decision ordering (root→leaf)

- x > y
- y > 0
- x > 0

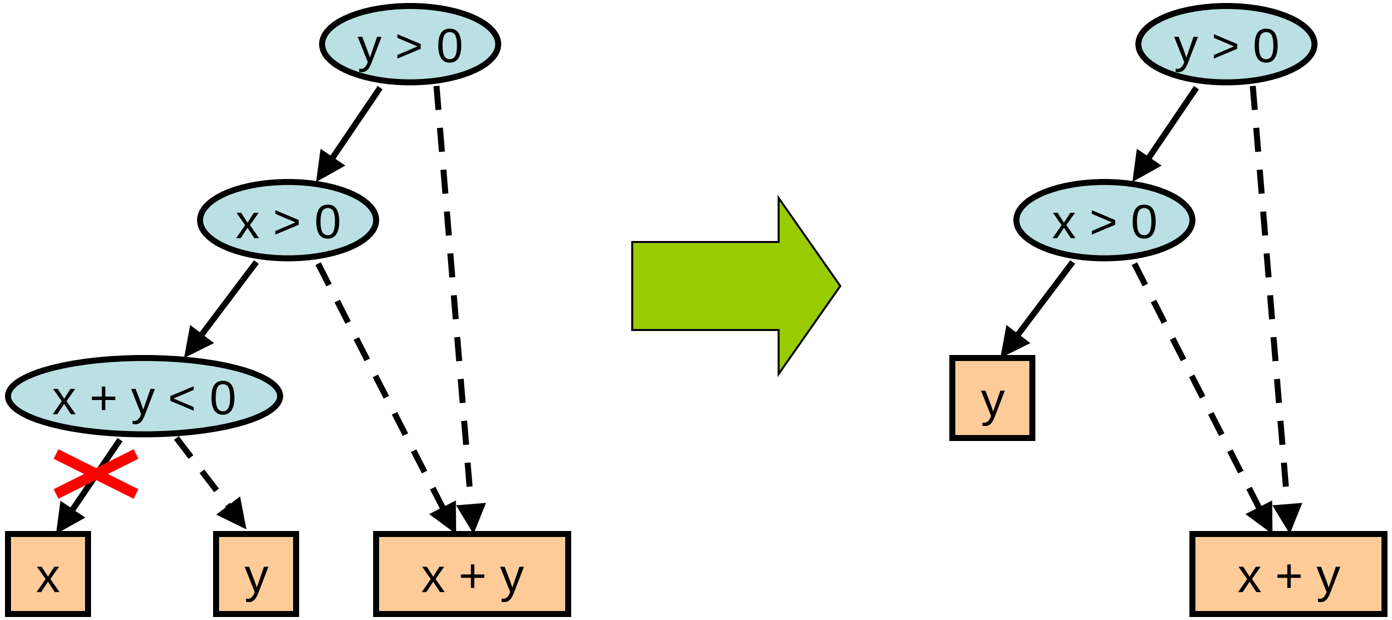


Correcting XADD Ordering

- Obtain *ordered* XADD from *unordered* XADD
  - key idea: binary operations maintain orderings



Maintaining Minimality



Node unreachable –
x + y < 0 always
false if x > 0 & y > 0

If **linear**, can detect with
feasibility checker of LP
solver & prune

More subtle
prunings as
well.

XADD Makes Possible all Previous Inference

Could not even do a single integral or maximization without it!

Recap

- Defined a calculus for piecewise functions
    - $f_1 \oplus f_2$, $f_1 \otimes f_2$
    - $\max(f_1, f_2)$, $\min(f_1, f_2)$
    - $\int_x f(x)$
    - See Zamani & Sanner (AAAI-12) for $\max_x f(x)$, $\min_x f(x)$
- **Defined XADD to efficiently compute with cases**
- **Makes possible**
    - **Closed-form inference in continuous graphical models**
    - **Exact (conditional) expectations of any polynomial**
    - **New paradigms for Bayesian inference**

Piecewise Calculus + XADD

= Expressive, Exact, Closed-form Inference in Discrete & Continuous Variable Graphical Models

Thank you!

Questions?