Length-Lex Set Variables

Supervisor: Pascal Van Hentenryck

Justin Yip
Brown University

Collaborators: Carmen Gervet, Grégoire Dooms

Introduction

Domain Representation	Subset-Bound	Length-Lex	ROBDD
Propagation	Loose Weak BC	Precise Strong BC	Very Precise AC
Space	$O(n)$	$O(c)$	Potentially exponential
Efficiency	Very Fast	Fast O(poly(c))	Potentially slow

Set variable is a natural modeling vehicle. However its domain may contains an exponential number (to the universe size) of sets.

Explicit representation is undesirable.

Length-Lex representation tackles this problem by approximating the set domain precisely, cheaply and efficiently.

Length-Lex Propagators	Time	Reference
BC on unary	$\tilde{O}(c)$	Gervet and Van Hentenryck, AAAI06
BC on binary	$O(c^3)$	Van Hentenryck, Yip, Gervet and Dooms, AAAI08
BC* on knapsack	$O(n^4/\varepsilon)$	Malitsky et al., CP08
BC on knapsack	$\tilde{O}(c)$	Yip, Van Hentenryck, SAC09
BC on binary	$O(c^2)$	Unpublished

n : universe size, c : cardinality of largest set, usually $n >> c$

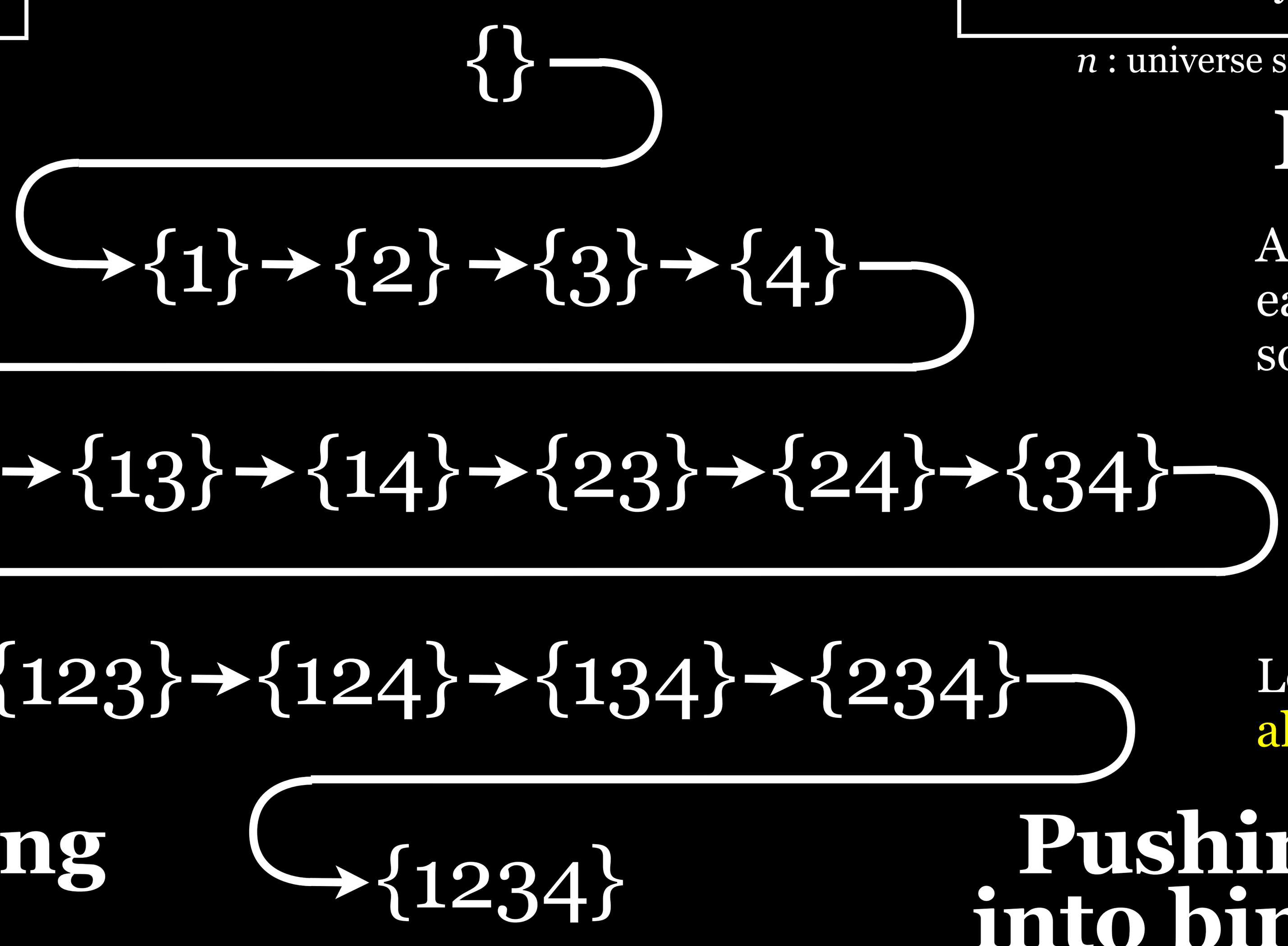
Length-Lex ordering «

Length-Lex orders sets first by cardinality, then breaks tie by lex.

$(s \ll t) \Leftrightarrow (|s|<|t|)$ or $(|s|=|t|$ and $s \leq_{lex} t)$

Length-Lex domain $\langle l,u,n \rangle$ is an interval, l,u are the lower and upper bound, and n is the universe size.

e.g. <{23},{134},4> denotes the set of sets { {23}, {24}, {34}, {123}, {124}, {134} }.



Bound consistency

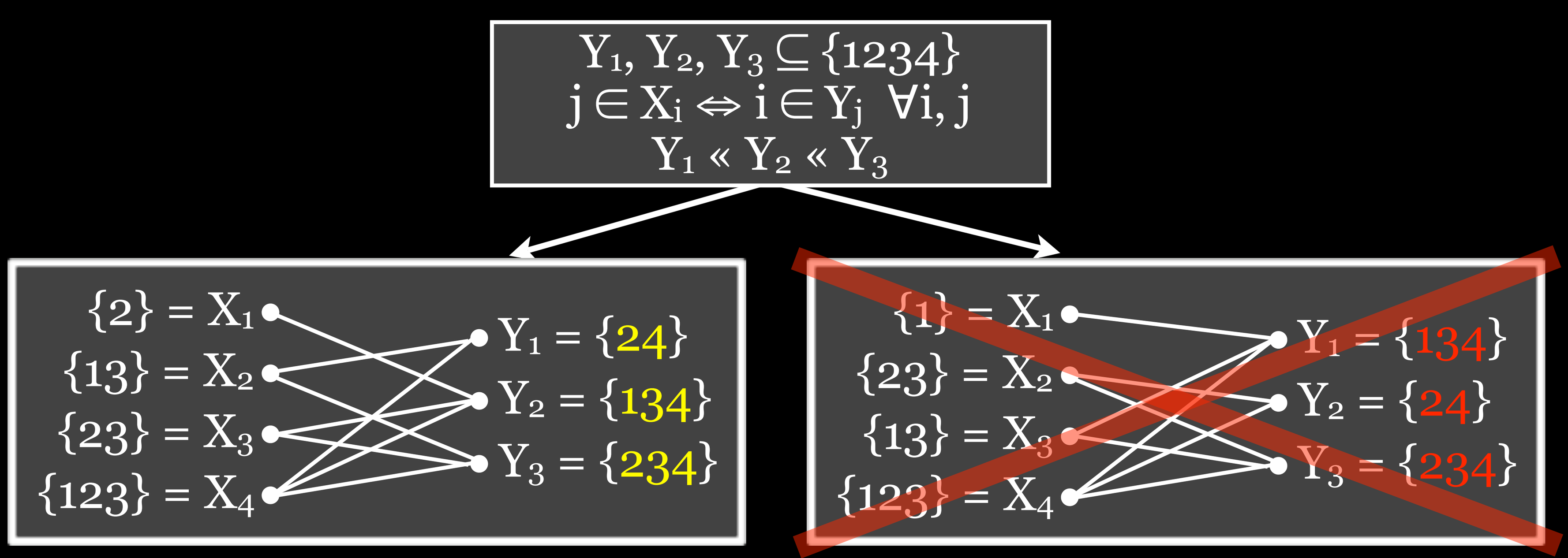
A constraint is bound-consistent (BC) if each lower (upper) bound belongs to some solutions.

e.g. C(X) ≡ {1,3} ⊆ X
X : <{13},{134},4> is BC.
X : <{1},{134},4> is not BC since lower bound {1} violates the constraint.

Length-Lex features a total-order hence always possible to enforce BC.

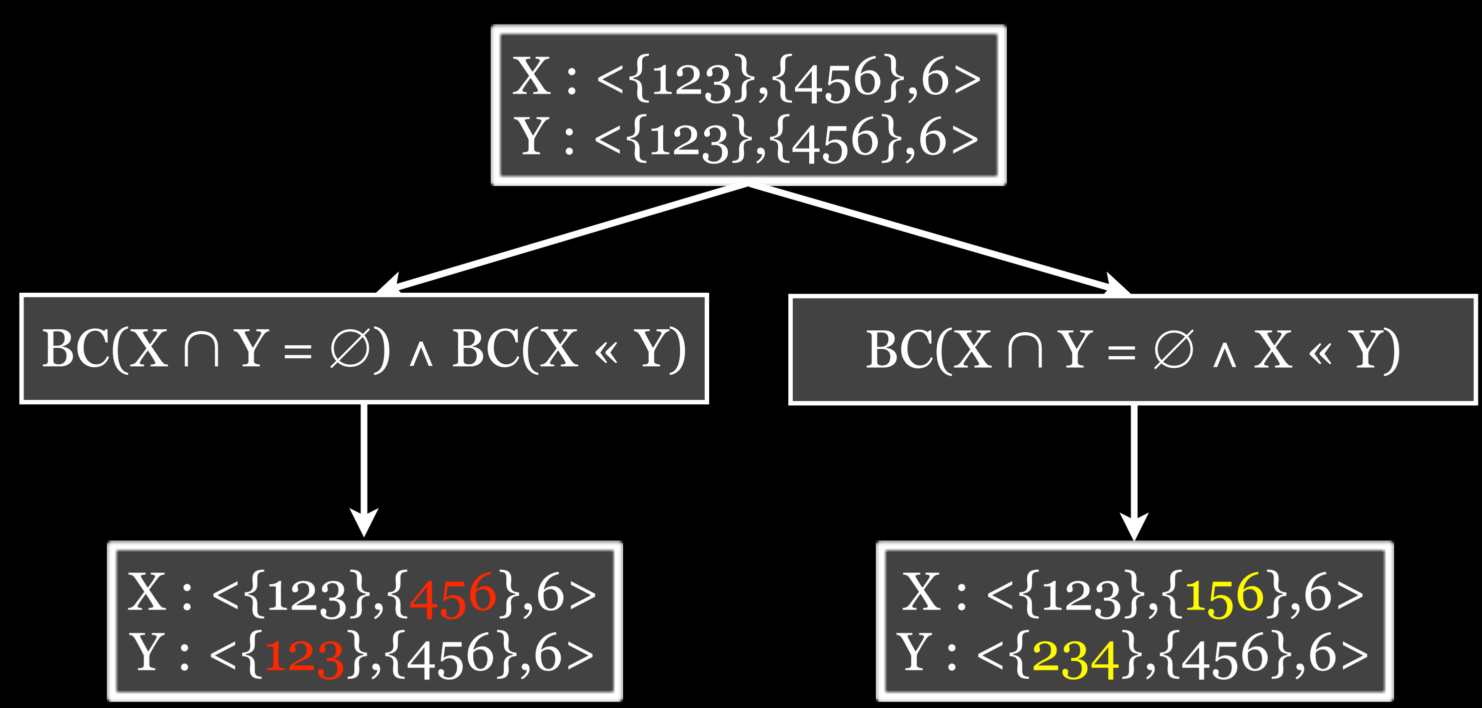
Dual model for breaking value symmetries

Given a CSP with 4 variables $X_i \subseteq \{123\}$, and values 1,2,3 are interchangeable. A dual model can break value symmetries.



Pushing LL-ordering into binary constraints

Combining constraints achieves more propagations.



Experiments

44000+ lines of C++ code
Partially integrated into COMET

All Length-Lex models apply above techniques (breaking value symmetries, pushing LL-ordering constraint).

Length-Lex efficiently solve all instances that can be solved by other approaches. For details, please refer to "Evaluation of Length-Lex Set Variables, CP09"

Social Golfer Problem

	Techniques	# successful instances	Slowdown (arith-mean)	Slowdown (geo-mean)	CPU
Length-Lex	Push-LL, Value-sym	60	1	1	2.53GHz
ROBDD/seq	Push-LL	17	650.74	21.71	2.8GHz
ROBDD/dyn	Push-order	20	82.11	12.82	2.8GHz
Set-int	Value-sym	26	22936.9	257.49	Sparc 900MHz
Cardinal		15	413.64	76.35	2.4GHz
PairAtMost1		11	282.07	9.66	3.8GHz

Steiner Triple Systems

	Techniques	# successful instances	Slowdown (arith-mean)	Slowdown (geo-mean)	CPU
Length-Lex	Push-LL, Value-sym	9	1	1	2.53GHz
ROBDD/ Bounds	Push-LL	4	14.03	10.24	2.8GHz
ROBDD/ Domain	Push-LL	4	558.68	54.56	2.8GHz
01Matrix-Lex	Value-Sym	2	5.81	2.36	1GHz
01Matrix-LexSum	Push-LL, Value-sym	2	5.81	2.36	1GHz
Cardinal		7	13.86	11.09	2.4GHz
Valprec	Value-Sym	4	14.03	10.24	Sparc 900MHz

Balanced Incomplete Block Designs

	Techniques	# successful instances	Slowdown (arith-mean)	Slowdown (geo-mean)	CPU
Length-Lex	Push-LL, Value-sym	29	1	1	2.53GHz
Max-Variety	Value-sym	25	15.82	6.58	Sparc 360MHz
01Matrix-Lex	Value-sym	29	1183.35	12.9	1GHz
01Matrix-LexSum	Push-LL, Value-sym	8	4273.64	3868.45	1GHz

Error Correcting Code Problem

Easy Instances	Techniques	# successful instances	Slowdown (arith-mean)	Slowdown (geo-mean)	CPU
Length-Lex	Push-LL, Value-sym	51	1	-	2.53GHz
ROBDD	Push-LL	51	36.77	-	2.8GHz

Hard Instances	Techniques	# successful instances	Slowdown (arith-mean)	Slowdown (geo-mean)	CPU
Length-Lex	Push-LL, Value-sym	6	1	1	2.53GHz
ROBDD	Push-LL	4	221.39	50.24	2.8GHz

Steiner Systems

	Techniques	# successful instances	Slowdown (arith-mean)	Slowdown (geo-mean)	CPU
Length-Lex	Push-LL, Value-sym	7	1	1	2.53GHz
ROBDD/ Bounds	Push-LL	6	509.57	20.04	2.8GHz
ROBDD/ Domain	Push-LL	6	2728.07	65.74	2.8GHz
Hybrid	Value-Sym	5	27.38	21.95	1GHz
Subset-bound	Value-Sym	5	2.59	2.24	1GHz

(Indeed, empirical results explain everything)