Abstract Domains of Affine Relations

Matt Elder, Junghee Lim, Tushar Sharma, Tycho Andersen, Thomas Reps

University of Wisconsin-Madison

Static Analysis Symposium, 2011

Efficiently find sound linear equalities in programs using machine integers

Efficiently find sound linear equalities in programs using machine integers Machine integers = w-bit ints

Analyzing ints

Difficulties:

- Arithmetic overflow: $12 + 7 \equiv_{16} 3$
- All even numbers are zero divisors: $2 \cdot 8 \equiv_{16} 0$
- All odd numbers have inverses: $3 \cdot 11 \equiv_{16} 1$

Difficulties:

- Arithmetic overflow: $12 + 7 \equiv_{16} 3$
- All even numbers are zero divisors: $2 \cdot 8 \equiv_{16} 0$
- All odd numbers have inverses: $3 \cdot 11 \equiv_{16} 1$

Advantages:

- Can represent some bit-level properties in linear equations: $8x \equiv_{16} 8$ means x is odd
- Fast operations on native ints
- Int domains are finite lattices
- Soundness: capture real semantics

KS: Conjunction of affine constraints **MOS:** Affine set of affine transformers

KS: Conjunction of affine constraints **MOS:** Affine set of affine transformers

KS and MOS are two-vocabulary domains

Definition

If the set of concrete program states is S, a **two-vocabulary domain** abstracts relations from S to S

"Standard" domains abstract program states, Two-vocabulary domains abstract program transitions

Also called:

- Sharir-Pneuli-style domain
- Transformer domain
- Transition domain

How can we adapt KS to directly model *w*-bit ints?

Definition

Symbolic abstraction converts logical formulas to overapproximating domain elements

Definition

Symbolic abstraction converts logical formulas to overapproximating domain elements

Definition

Symbolic concretization converts domain elements to overapproximating logical formulas

How can we perform symbolic abstraction to MOS?

KS and MOS: Which is more precise? Which is more efficient?

- How can we adapt KS to directly model w-bit ints?
- How can we perform symbolic abstraction to MOS?
- Which is more precise, KS or MOS?
- Which is more efficient, KS or MOS?

- How can we adapt KS to directly model w-bit ints?
- How can we perform symbolic abstraction to MOS?
- Which is more precise, KS or MOS?
- Which is more efficient, KS or MOS?

How can we adapt KS to directly model *w*-bit ints?

KS Definition

KS element: Matrix of w-bit ints; each row encodes a constraint

Example

$$\begin{bmatrix} x & y & x' & y' & 1 \\ 5 & 7 & 9 & 12 & 6 \\ 5 & 1 & 9 & 2 & 8 \end{bmatrix} : \quad \begin{array}{c} 5x + 7y + 9x' + 12y' + 6 = 0 \\ and 5x + y + 9x' + 2y' + 8 = 0 \\ \end{array}$$

Generalization of: King and Søndergaard, CAV 2008

KS Definition

KS element: Matrix of w-bit ints; each row encodes a constraint

Example

$$\begin{bmatrix} x & y & x' & y' & 1 \\ 5 & 7 & 9 & 12 & 6 \\ 5 & 1 & 9 & 2 & 8 \end{bmatrix} : \begin{bmatrix} 5 & 7 & 9 & 12 & 6 \\ 5 & 1 & 9 & 2 & 8 \end{bmatrix} \begin{bmatrix} x \\ y \\ x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

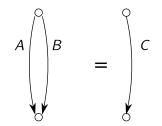
Generalization of: King and Søndergaard, CAV 2008

$$\begin{bmatrix} A_{\text{pre}}, A_{\text{post}}, A_{\text{c}} \end{bmatrix} = \bigcirc \\ \begin{bmatrix} B_{\text{pre}}, B_{\text{post}}, B_{\text{c}} \end{bmatrix} = \bigcirc \\ C$$

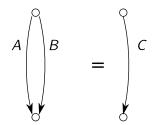
$$\begin{bmatrix} A_{\text{pre}}, A_{\text{post}}, A_{\text{c}} \end{bmatrix} = \bigcirc \\ \begin{bmatrix} B_{\text{pre}}, B_{\text{post}}, B_{\text{c}} \end{bmatrix} = \bigcirc \\ C$$

$$C = {Project} \left(egin{bmatrix} {m{A}_{
m pre}} & {m{A}_{
m post}} & 0 & {m{A}_{
m c}} \ 0 & {m{B}_{
m pre}} & {m{B}_{
m post}} & {m{B}_{
m c}} \end{bmatrix}
ight)$$

KS Join



KS Join



$$C = \operatorname{Project}\left(\begin{bmatrix} -A & A \\ B & 0 \end{bmatrix}\right)$$

Elder, Lim, Sharma, Andersen, Reps

Project is the critical operation!

- 1 Move lost variables to the left
- 2 Do Gaussian elimination
- 3 Drop every row constraining a lost variable
- 4 Drop the lost-variable columns

$$\begin{bmatrix} x & y & x' & y' & 1 \\ 9 & 5 & 12 & 7 & 6 \\ 9 & 5 & 2 & 3 & 8 \end{bmatrix}$$

1 Move lost variables to the left

- 2 Do Gaussian elimination
- 3 Drop every row constraining a lost variable
- 4 Drop the lost-variable columns

$$\begin{bmatrix} x & y & x' & y' & 1 \\ 9 & 5 & 12 & 7 & 6 \\ 9 & 5 & 2 & 3 & 8 \end{bmatrix}$$

1 Move lost variables to the left

- 2 Do Gaussian elimination
- 3 Drop every row constraining a lost variable
- 4 Drop the lost-variable columns

$$\begin{bmatrix} y & y' & x & x' & 1 \\ 5 & 7 & 9 & 12 & 6 \\ 5 & 3 & 9 & 2 & 8 \end{bmatrix}$$

- 1 Move lost variables to the left
- 2 Do Gaussian elimination
- 3 Drop every row constraining a lost variable
- 4 Drop the lost-variable columns

$$\begin{bmatrix} y & y' & x & x' & 1 \\ 5 & 7 & 9 & 12 & 6 \\ 5 & 3 & 9 & 2 & 8 \end{bmatrix}$$

- 1 Move lost variables to the left
- 2 Do Gaussian elimination
- 3 Drop every row constraining a lost variable
- 4 Drop the lost-variable columns

$$\begin{bmatrix} y & y' & x & x' & 1 \\ 5 & 7 & 9 & 12 & 6 \\ 0 & 12 & 0 & 6 & 2 \end{bmatrix}$$

- 1 Move lost variables to the left
- 2 Do Gaussian elimination
- 3 Drop every row constraining a lost variable
- 4 Drop the lost-variable columns

$$\begin{bmatrix} y & y' & x & x' & 1 \\ 5 & 7 & 9 & 12 & 6 \\ 0 & 12 & 0 & 6 & 2 \end{bmatrix}$$

- Move lost variables to the left
- 2 Do Gaussian elimination
- 3 Drop every row constraining a lost variable
- 4 Drop the lost-variable columns

- Move lost variables to the left
- 2 Do Gaussian elimination
- 3 Drop every row constraining a lost variable
- 4 Drop the lost-variable columns

- 1 Move lost variables to the left
- 2 Do Gaussian elimination
- 3 Drop every row constraining a lost variable
- 4 Drop the lost-variable columns

Example (Project onto x and x')

$$\begin{bmatrix} y & y' & x & x' & 1 \\ 5 & 7 & 9 & 12 & 6 \\ 0 & 12 & 0 & 6 & 2 \end{bmatrix} \Rightarrow 12y' + 6x' + 2 = 0$$

- 1 Move lost variables to the left
- 2 Do Gaussian elimination
- 3 Drop every row constraining a lost variable
- 4 Drop the lost-variable columns

Example (Project onto x and x')

$$\begin{bmatrix} y & y' & x & x' & 1 \\ 5 & 7 & 9 & 12 & 6 \\ 0 & 12 & 0 & 6 & 2 \end{bmatrix} \Rightarrow 12y' + 6x' + 2 \equiv_{16} 0$$

- 1 Move lost variables to the left
- 2 Do Gaussian elimination
- 3 Drop every row constraining a lost variable
- 4 Drop the lost-variable columns

Example (Project onto x and x')

$$\begin{bmatrix} y & y' & x & x' & 1 \\ 5 & 7 & 9 & 12 & 6 \\ 0 & 12 & 0 & 6 & 2 \end{bmatrix} \Rightarrow 48y' + 24x' + 8 \equiv_{16} 0$$

- 1 Move lost variables to the left
- 2 Do Gaussian elimination
- 3 Drop every row constraining a lost variable
- 4 Drop the lost-variable columns

Example (Project onto x and x')

$$\begin{bmatrix} y & y' & x & x' & 1 \\ 5 & 7 & 9 & 12 & 6 \\ 0 & 12 & 0 & 6 & 2 \end{bmatrix} \Rightarrow 8x' + 8 \equiv_{16} 0$$

- 1 Move lost variables to the left
- 2 Do Gaussian elimination
- 3 Drop every row constraining a lost variable
- 4 Drop the lost-variable columns

Example (Project onto x and x')

$$\begin{bmatrix} y & y' & x & x' & 1 \\ 5 & 7 & 9 & 12 & 6 \\ 0 & 12 & 0 & 6 & 2 \end{bmatrix} \Rightarrow 8x' \equiv_{16} 8$$

Definition

The **row space** of a matrix is the set of linear combinations of its rows

Definition

The **null space** of a matrix is the set of values whose product with the matrix is zero

The null space of *A* is $\{x \mid Ax = 0\}$

Definition

A **row operation** is a matrix transformation that adds or changes individual rows without changing the matrix's row space

Some Row Operations:

- Scale a row by an odd number
- Add a multiple of one row to another
- Insert some multiple of a row into the matrix

Row Operations

What if we scale a row by an even number?

Example

Scale first row by 2:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

What if we scale a row by an even number?

Example

Scale first row by 2:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

First row space: $\begin{bmatrix} x & y & 0 \end{bmatrix}$ for any *x*, any *y* Second row space: $\begin{bmatrix} x & y & 0 \end{bmatrix}$ for even *x*, any *y*

Row-Echelon form

Definition

A matrix is in **Row-Echelon Form** if each row has fewer leading zeroes than the next row



$$\begin{bmatrix} * & * & * & * & * & * \\ 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & * & * & * \\ 0 & 0 & 0 & 0 & 0 & * \end{bmatrix}$$

Howell form is a normal form for int matrices

Howell form is a normal form for int matrices

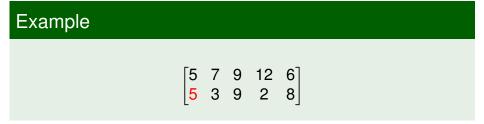
Like Gaussian Elimination, Howellization preserves the row space

- Normal form for row spaces
- Normal form for null spaces
- Normal form for KS

- Normal form for row spaces
- Normal form for null spaces
- Normal form for KS

Simplifies checking KS equality

The matrix is in Row-Echelon form



The matrix is in Row-Echelon form

Example $\begin{bmatrix} 5 & 7 & 9 & 12 & 6 \\ 0 & 12 & 0 & 6 & 2 \end{bmatrix}$

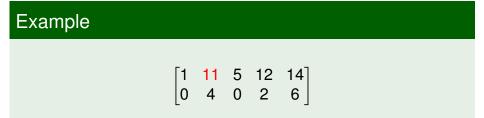
Leading values are powers of 2

$$\begin{bmatrix} 5 & 7 & 9 & 12 & 6 \\ 0 & 12 & 0 & 6 & 2 \end{bmatrix}$$

Leading values are powers of 2

$$\begin{bmatrix} 1 & 11 & 5 & 12 & 14 \\ 0 & 4 & 0 & 2 & 6 \end{bmatrix}$$

Leading values are largest in their columns



Leading values are largest in their columns

Example $\begin{bmatrix} 1 & 3 & 5 & 8 & 2 \\ 0 & 4 & 0 & 2 & 6 \end{bmatrix}$

Every consequence of every row is a linear combination of the matrix rows that have at least as many leading zeros as the consequence

Every consequence of every row is a linear combination of the matrix rows that have at least as many leading zeros as the consequence

Definition

The vectors $2^k v$ are the **consequences** of v

Example

 $\begin{array}{l} \text{The consequences of } \begin{bmatrix} 0 & 4 & 0 & 2 & 6 \end{bmatrix} \text{ are} \\ \left\{ \begin{bmatrix} 0 & 8 & 0 & 4 & 4 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & 8 & 8 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \end{bmatrix} \right\} \end{array}$

Every consequence of every row is a linear combination of the matrix rows that have at least as many leading zeros as the consequence

$$\begin{bmatrix} 1 & 3 & 5 & 8 & 2 \\ 0 & 4 & 0 & 2 & 6 \end{bmatrix}$$

Every consequence of every row is a linear combination of the matrix rows that have at least as many leading zeros as the consequence

$$\begin{bmatrix} 1 & 3 & 5 & 8 & 2 \\ 0 & 4 & 0 & 2 & 6 \end{bmatrix}$$
$$\begin{bmatrix} 0 & 8 & 0 & 4 & 4 \end{bmatrix}$$

Every consequence of every row is a linear combination of the matrix rows that have at least as many leading zeros as the consequence

$$\begin{bmatrix} 1 & 3 & 5 & 8 & 2 \\ 0 & 4 & 0 & 2 & 6 \end{bmatrix}$$
$$\begin{bmatrix} 0 & 0 & 0 & 8 & 8 \end{bmatrix}$$

Every consequence of every row is a linear combination of the matrix rows that have at least as many leading zeros as the consequence

$$\begin{bmatrix} 1 & 3 & 5 & 8 & 2 \\ 0 & 4 & 0 & 2 & 6 \\ 0 & 0 & 0 & 8 & 8 \end{bmatrix}$$

Every consequence of every row is a linear combination of the matrix rows that have at least as many leading zeros as the consequence

$$\begin{bmatrix} 1 & 3 & 5 & 8 & 2 \\ 0 & 4 & 0 & 2 & 6 \\ 0 & 0 & 0 & 8 & 8 \end{bmatrix}$$

Every consequence of every row is a linear combination of the matrix rows that have at least as many leading zeros as the consequence

$$\begin{bmatrix} 1 & 3 & 5 & 0 & 10 \\ 0 & 4 & 0 & 2 & 6 \\ 0 & 0 & 0 & 8 & 8 \end{bmatrix}$$

- 1 Move lost variables to the left
- 2 Howellize the matrix
- 3 Drop every row constraining a lost variable
- 4 Drop the lost-variable columns

$$\begin{bmatrix} x & y & x' & y' & 1 \\ 9 & 5 & 12 & 7 & 6 \\ 9 & 5 & 2 & 3 & 8 \end{bmatrix}$$

Move lost variables to the left

- 2 Howellize the matrix
- 3 Drop every row constraining a lost variable
- 4 Drop the lost-variable columns

$$\begin{bmatrix} x & y & x' & y' & 1 \\ 9 & 5 & 12 & 7 & 6 \\ 9 & 5 & 2 & 3 & 8 \end{bmatrix}$$

Move lost variables to the left

- 2 Howellize the matrix
- 3 Drop every row constraining a lost variable
- 4 Drop the lost-variable columns

$$\begin{bmatrix} y & y' & x & x' & 1 \\ 5 & 7 & 9 & 12 & 6 \\ 5 & 3 & 9 & 2 & 8 \end{bmatrix}$$

- 1 Move lost variables to the left
- 2 Howellize the matrix
- 3 Drop every row constraining a lost variable
- 4 Drop the lost-variable columns

$$\begin{bmatrix} y & y' & x & x' & 1 \\ 5 & 7 & 9 & 12 & 6 \\ 5 & 3 & 9 & 2 & 8 \end{bmatrix}$$

- Move lost variables to the left
- 2 Howellize the matrix
- 3 Drop every row constraining a lost variable
- 4 Drop the lost-variable columns

$$\begin{bmatrix} y & y' & x & x' & 1 \\ 1 & 3 & 5 & 0 & 10 \\ 0 & 4 & 0 & 2 & 6 \\ 0 & 0 & 0 & 8 & 8 \end{bmatrix}$$

- Move lost variables to the left
- 2 Howellize the matrix
- 3 Drop every row constraining a lost variable
- 4 Drop the lost-variable columns

$$\begin{bmatrix} y & y' & x & x' & 1 \\ 1 & 3 & 5 & 0 & 10 \\ 0 & 4 & 0 & 2 & 6 \\ 0 & 0 & 0 & 8 & 8 \end{bmatrix}$$

- 1 Move lost variables to the left
- 2 Howellize the matrix
- 3 Drop every row constraining a lost variable
- 4 Drop the lost-variable columns

$$\begin{bmatrix} y & y' & x & x' & 1 \\ 0 & 0 & 0 & 8 & 8 \end{bmatrix}$$

- 1 Move lost variables to the left
- 2 Howellize the matrix
- 3 Drop every row constraining a lost variable
- I Drop the lost-variable columns

- 1 Move lost variables to the left
- 2 Howellize the matrix
- 3 Drop every row constraining a lost variable
- 4 Drop the lost-variable columns

How can we adapt KS to directly model *w*-bit ints? Use Howell form for projection!

Which is more precise, KS or MOS?

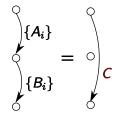
MOS element: a set of matrices of *w*-bit ints; every affine combination those matrices may transform the initial state

Example

$$\left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 6 \\ 0 & 0 & 1 \end{bmatrix} \right\} : \exists k : \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2k \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix}$$

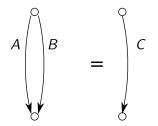
See: Müller-Olm and Seidl, TOPLAS 2007

MOS Compose



$$C = Basis \{B_i A_i\}$$

MOS Join



$$C = Basis \{A \cup B\}$$

Can use Howellize for the Basis function

Can use Howellize for the *Basis* function This allows easy equality checking! MOS can represent non-affine constraints!

MOS can represent non-affine constraints!

$$\left\{ \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\} : \exists k : \begin{bmatrix} k & 1-k & 0 \\ k & 1-k & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix}$$

MOS can represent non-affine constraints!

$$\left\{ \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\} : \exists k : x' = y' = y + k(x - y)$$

$$[assume(x = 5)]: x = x' \land y = y' \land x = 5$$

$$\llbracket assume(x = 5) \rrbracket : x = x' \land y = y' \land x = 5$$

One of the best MOS transformers is
$$\left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\}$$

Example

$$\llbracket assume(x = 5) \rrbracket : x = x' \land y = y' \land x = 5$$

One of the best MOS transformers is
$$\left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\}$$

MOS cannot represent assumes!

Which is more precise, KS or MOS? KS and MOS are incomparable

Which is more efficient, KS or MOS?

For k variables:

	KS	MOS
Element size	$O(k^2)$	$O(k^4)$
Join	$O(k^3)$	$O(k^6)$
Compose	$O(k^{3})$	$O(k^7)$

If ring matrix multiplication is $O(k^{\alpha})$, then:

	KS	MOS
Element size	$O(k^2)$	$O(k^4)$
Join	$O(k^{lpha})$	$O(k^{2lpha})$
Compose	$O(k^{lpha})$	$O(k^{4+\alpha})$

For eight small programs (500-4000 instructions):

- 1 Compute MOS and KS elements on program edges
- Perform two-phase queries at the beginning of basic blocks that end in branches
- 3 Compare MOS and KS precision at each query point

Symbolic abstraction to build KS elements Operator reinterpretation to build MOS elements Symbolic abstraction to build KS elements Operator reinterpretation to build MOS elements

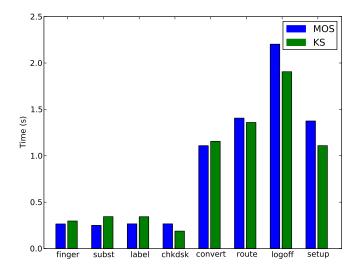
SMT is used only in devising initial KS elements Not in KS's analysis phases; nowhere in MOS KS (with symbolic abstraction) was at least as precise as MOS (with operator reinterpretation) at **every** query point KS (with symbolic abstraction) was at least as precise as MOS (with operator reinterpretation) at **every** query point

If this holds for KS with operator reinterpretation, then MOS's non-affine constraints don't help for real programs

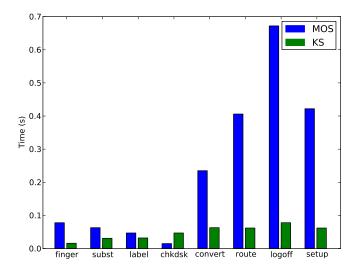
Constructing KS via symbolic abstraction took 325 times longer than constructing MOS via operator reinterpretation

- Constructing KS via symbolic abstraction took 325 times longer than constructing MOS via operator reinterpretation
- Constructing KS via operator reinterpretation: coming soon

Experimental Results: Phase 1 Time



Experimental Results: Phase 2 Time



Overall, KS analysis time was 91% of MOS analysis time Phase 1 time, KS/MOS: 94% Phase 2 time, KS/MOS: 20%

Seems that KS analysis is somewhat faster than MOS on real inputs

- Howell form allows projection in KS
- Howell form is a normal form for KS and MOS
- MOS can capture non-affine constraints

KS for w-bit ints:

- Needs no bit blasting
- Now applies to larger programs
- KS and MOS are mathematically incomparable
- KS analysis is more efficient than MOS, in theory and (provisionally) in practice