Lecture 1 Synthesis, Analysis, and Verification

Viktor Kuncak

Today

See web site http://lara.epfl.ch/w/sav for logistics What

- verification
- analysis
- synthesis

Demo of http://leon.epfl.ch Verification industry ((r)/TM/(c))

- Coverity (press release)
- AbsInt and ASTREE
- Jasper
- Monoidics

Countdown Program

```
object Even {
  def countdown(acc: BigInt, x: BigInt): BigInt = {
    require(0 <= x)
    if (x == 0) acc
    else countdown(acc + 2, x - 1)
    } ensuring(_ == acc + 2*x)</pre>
```

```
def count(x: BigInt): BigInt = {
  require(0 <= x)
  countdown(0, x)
} ensuring(_ == 2*x)</pre>
```

def ex1 = count(100)

Deductive Verification

Three-step approach:

- 1. Compile program meaning to logical formulas (verification-condition generator, symbolic execution)
- 2. Express properties in logic or code (assertions, preconditions, post-conditions, invariants, run-time error conditions)
- 3. Develop and use an automated theorem prover for generated conditions (SAT solving, SMT solving, resolution-based theorem proving, rewriting, interactive provers)

Which logic to use? Today: integer linear arithmetic

Presburger arithmetic

Integer arithmetic with logical operations (and, or, not), quantifiers, only addition as an arithmetic operation, and < and = as a relation.

► minimalistically one can define a variant of it over non-negative natural numbers as having ∧, ¬, ∀, +, = as the only symbols

One of the earliest theories shown decidable. Mojżesz Presburger gave an algorithm for quantifier elimination in 1929.

- a student of a famous logician Alfred Tarski
- Tarski gave him this question for his MSc thesis

The result at this time was of interest to mathematical logic and foundations of mathematics

 only much later it found applications in automated reasoning (Cooper 1972, Derek C. Oppen - STOC 1973)

Presburger Arithmetic for Verification

Verification condition (VC) for preservation of loop invariant:

$$\left[\textit{I}(\textit{res},i) \land i' = i - 1 \land \textit{res}' = \textit{res} + 2 \land 0 < i\right] \rightarrow \textit{I}(\textit{res}',i')$$

To prove that this VC is valid, we check whether its negation

$$I(res, i) \land i' = i - 1 \land res' = res + 2 \land 0 < i \land \neg I(res', i')$$

is satisfiable, i.e. whether this PA formula is true:

$$\exists x, res, i, res', i'. [res + 2i = 2x \land 0 \le i \land 0 < i \land$$
$$i' = i - 1 \land res' = res + 2 \land$$
$$\neg (res' + 2i' = 2x \land 0 \le i')]$$

Introducing: One-Point Rule

If \bar{y} is a tuple of variables not containing x, then

$$\exists x.(x = t(\bar{y}) \land F(x, \bar{y})) \iff F(t(\bar{y}), \bar{y})$$

Proof:

- → : Consider the values of \bar{y} such that there exists x, say x_1 , for which $x_1 = t(\bar{y}) \land F(x_1, \bar{y})$. Because $F(x_1, \bar{y})$ evaluates to true and the values of x_1 and $t(\bar{y})$ are the same, $F(t, \bar{y})$ also evaluates to true.
- $\leftarrow : \text{Let } \bar{y} \text{ be such that } F(t, \bar{y}) \text{ holds. Let } x \text{ be the value of } t(\bar{y}).$ Then of course $x = t(\bar{y})$ evaluates to true and so does $F(x, \bar{y})$. So there exists x for which $x = t(\bar{y}) \wedge F(x, \bar{y})$ holds.

One point rule:

replaces left side (LHS) of equivalence by the right side (RHS). *Flattening*, used when t is complex, replaces RHS by LHS.

Dual One-Point Rule for \forall

$$\forall x.(x = t(\bar{y}) \rightarrow F(x, \bar{y})) \iff F(t(\bar{y}), \bar{y})$$

To prove it, negate both sides:

$$\exists x.(x = t(\bar{y}) \land \neg F(x, \bar{y})) \iff \neg F(t(\bar{y}), \bar{y})$$

so it reduces to the rule for \exists .

$$\exists x, res, i, res', \underline{\mathbf{i}'}. [res + 2i = 2x \land 0 \le i \land 0 < i \land \\ \underline{\mathbf{i'} = \mathbf{i} - 1} \land res' = res + 2 \land \\ \neg (res' + 2i' = 2x \land 0 \le i')]$$

$$\exists x, res, i, res', \underline{i'}. [res + 2i = 2x \land 0 \le i \land 0 < i \land \\ \underline{i' = i - 1} \land res' = res + 2 \land \\ \neg (res' + 2i' = 2x \land 0 \le i')]$$

$$\exists x, res, i, \underline{res'}. [res + 2i = 2x \land 0 \le i \land 0 < i \land$$
$$\underbrace{res' = res + 2}_{\neg (res' + 2(i-1)) = 2x \land 0 \le i - 1)]$$

$$\exists x, res, i, res', \underline{i'}. [res + 2i = 2x \land 0 \le i \land 0 < i \land \\ \underline{i' = i - 1} \land res' = res + 2 \land \\ \neg (res' + 2i' = 2x \land 0 \le i')]$$

$$\exists x, res, i, \underline{res'}. [res + 2i = 2x \land 0 \le i \land 0 < i \land$$
$$\underbrace{\frac{res' = res + 2}{\neg (res' + 2(i - 1))}}_{2x \land 0 \le i - 1)]$$

$$\exists x, res, i. \ \left[\frac{res + 2i = 2x}{res + 2i + 2(i-1)} \land 0 \le i \land 0 < i \land \right. \\ \neg (res + 2 + 2(i-1) = 2x \land 0 \le i - 1) \right]$$

$$\exists x, res, i, res', \underline{i'}. [res + 2i = 2x \land 0 \le i \land 0 < i \land \\ \underline{i' = i - 1} \land res' = res + 2 \land \\ \neg (res' + 2i' = 2x \land 0 \le i')]$$

$$\exists x, res, i, \underline{res'}. [res + 2i = 2x \land 0 \le i \land 0 < i \land$$
$$\underbrace{\frac{res' = res + 2}{\neg (res' + 2(i-1))}}_{= 2x \land 0 \le i - 1)]$$

$$\exists x, res, i. \ \left[\underline{res + 2i = 2x} \land 0 \le i \land 0 < i \land \\ \neg (res + 2 + 2(i-1) = 2x \land 0 \le i - 1) \right]$$

$$\exists x, \underline{\text{res}}, i. \ \left[\underline{\text{res} = 2x - 2i} \land 0 \le i \land 0 < i \land \\ \neg(\text{res} + 2 + 2(i - 1) = 2x \land 0 \le i - 1)\right]$$

$$\exists x, res, i, res', \underline{i'}. [res + 2i = 2x \land 0 \le i \land 0 < i \land \\ \underline{i' = i - 1} \land res' = res + 2 \land \\ \neg (res' + 2i' = 2x \land 0 \le i')]$$

$$\exists x, res, i, \underline{res'}. [res + 2i = 2x \land 0 \le i \land 0 < i \land$$
$$\underbrace{\frac{res' = res + 2}{\neg (res' + 2(i-1))}}_{= 2x \land 0 \le i - 1)]$$

$$\exists x, res, i. \ \left[\underline{res + 2i = 2x} \land 0 \le i \land 0 < i \land \\ \neg (res + 2 + 2(i-1) = 2x \land 0 \le i-1) \right]$$

$$\exists x, \underline{res}, i. \ \left[\underline{res = 2x - 2i} \land 0 \le i \land 0 < i \land \\ \neg (res + 2 + 2(i - 1)) = 2x \land 0 \le i - 1)\right]$$

$$\exists x, i. \ \begin{bmatrix} 0 \le i \land 0 < i \land \\ \neg (2x - 2i + 2 + 2(i - 1)) = 2x \land 0 \le i - 1) \end{bmatrix}$$

Simplifies to $\exists x, i. \ 0 < i \land \neg (0 \leq i - 1)$ and then to false.

But there is more

One-point rule is one of the many steps used in **quantifier elimination** procedures.

Quantifier Elimination (QE)



Given a formula $F(\bar{y})$ containing quantifiers find a formula $G(\bar{y})$

- equivalent to $F(\bar{y})$
- that has no quantifiers
- ► and has a **subset (or equal set) of free variables** of *F* Note
 - ► Equivalence: For all y
 , F(y
) and G(y
) have same truth value → we can use G(y
) instead of F(y)
 - No quantifiers: easier to check satisfiability of $G(\bar{y})$
- \bar{y} is a possibly empty tuple of variables

We are lucky when a theory has ("admits") QE

Suppose F has no free variables (all variables are quantified). What is the result of applying QE to F?

We are lucky when a theory has ("admits") QE

Suppose F has no free variables (all variables are quantified). What is the result of applying QE to F? Are there any variables in the resulting formula?

We are lucky when a theory has ("admits") QE

Suppose F has no free variables (all variables are quantified). What is the result of applying QE to F? Are there any variables in the resulting formula?

- ▶ No free variables: they are a subset of the original, empty set
- ► No quantified variables: because it has no quantifiers ☺

Formula without any variables! Example:

$$(2+4=7) \lor (1+1=2)$$

We check the truth value of such formula by simply evaluating it!

Using QE for Deciding Satisfiability/Validity

- To check satisfiability of $H(\bar{y})$: eliminate the quantifiers from $\exists \bar{y}.H(\bar{y})$ and evaluate.
- ▶ Validity: eliminate quantifiers from $\forall \bar{y}.H(\bar{y})$ and evaluate

We can even check formulas like this:

$$\forall x, y, r. \exists z. (5 \leq r \land x + r \leq y) \rightarrow (x < z \land z < y \land 3|z)$$

Here 3|z denotes that z is divisible by 3.

Does Presburger Arithmetic admit QE?

Does Presburger Arithmetic admit QE?

Depends on the particular set of symbols! (Recall objective: given $F(\bar{y})$ containing quantifiers find a formula $G(\bar{y})$

- equivalent to $F(\bar{y})$
- that has no quantifiers
- and has a subset (or equal set) of free variables of F)

If we lack some operations that can be expressed using quantifiers, there may be no equivalent formula without quantifiers.

• $\exists y.x = y + y + y$, so we better have divisibility

Quantifier elimination says: if you can define some relationship between variables using an arbitrary, possibly quantified, formula F,

$$r \stackrel{def}{=} \{ (x, y) \mid F(x, y) \}$$

then you can also define same r using another quantifier-free formula G.

Presburger Arithmetic (PA)

We look at the theory of integers with addition.

- introduce constant for each integer constant
- to be able to restrict values to natural numbers when needed, and to compare them, we introduce <</p>
- introduce not only addition but also subtraction
- ► to conveniently express certain expressions, introduce function m_K for each K ∈ Z, to be interpreted as multiplication by a constant, m_K(x) = K ⋅ x. We write m_K as K ⋅ x. Note: there is no multiplication between variables in PA
- ► to enable quantifier elimination from ∃x.y = K · x introduce for each K predicate K|y (divisibility, y%K = 0)

The resulting language has these function and relation symbols: $\{+, -, =, <\} \cup \{K \mid K \in \mathbb{Z}\} \cup \{(K \cdot _) \mid K \in \mathbb{Z}\} \cup \{(K|_) \mid K \in \mathbb{Z}\}$ We also have, as usual: $\land, \lor, \neg, \rightarrow$ and also: \exists, \forall

Example

Eliminate *y* from this formula:

$$\exists y. \ 3y - 2w + 1 > -w \land 2y - 6 < z \land 4 \mid 5y + 1$$

What should we do first?

Example

Eliminate *y* from this formula:

$$\exists y. \ 3y - 2w + 1 > -w \land 2y - 6 < z \land 4 \mid 5y + 1$$

What should we do first?

Simplify/normalize what we can using properties of integer operations:

$$\exists y. \ 0 < -w + 3y + 1 \ \land \ 0 < -2y + z + 6 \ \land \ 4 \mid 5y + 1$$

Example

Eliminate *y* from this formula:

$$\exists y. \ 3y - 2w + 1 > -w \land 2y - 6 < z \land 4 \mid 5y + 1$$

What should we do first?

Simplify/normalize what we can using properties of integer operations:

$$\exists y. \ 0 < -w + 3y + 1 \ \land \ 0 < -2y + z + 6 \ \land \ 4 \mid 5y + 1$$

First we will consider only eliminating existential from a **conjunction of literals**.

Conjunctions of Literals

Atomic formula: a relation applied to argument. Here, relations are: =, <, $K|_{-}$. So, atomic formulas are: $t_1 = t_2$, $t_1 < t_2$, $K \mid t$

Conjunctions of Literals

Atomic formula: a relation applied to argument.

Here, relations are: =, <, $K|_{-}$. So, atomic formulas are:

 $t_1 = t_2, \quad t_1 < t_2, \quad K \mid t$

Literal: Atomic formula or its negation. Example: $\neg(x = y + 1)$ Conjunction of literals: $L_1 \land \ldots \land L_n$

- no disjunctions, no implications
- negation only applies to atomic formulas

We first consider the quantifier elimination problem of the form:

$$\exists y. L_1 \land \ldots \land L_n$$

This will prove to be sufficient to eliminate all quantifiers!

Can we eliminate \exists from any **quantifier-free formula**?

 $\exists x.F(x,\bar{y})$

where F is quantifier-free?

Can we eliminate \exists from any **quantifier-free formula**?

 $\exists x.F(x,\bar{y})$

where F is quantifier-free?

Formula without quantifiers has \land,\lor,\neg applied to atomic formulas.

Can we eliminate \exists from any **quantifier-free formula**?

 $\exists x.F(x,\bar{y})$

where *F* is quantifier-free?

Formula without quantifiers has \land,\lor,\neg applied to atomic formulas. Convert F to **disjunctive normal form**:

$$F \iff \bigvee_{i=1}^m C_i$$

each C_i is a **conjunction of literals**.

Can we eliminate \exists from any **quantifier-free formula**?

 $\exists x.F(x,\bar{y})$

where *F* is quantifier-free?

Formula without quantifiers has \land, \lor, \neg applied to atomic formulas. Convert F to **disjunctive normal form**:

$$\vdash \iff \bigvee_{i=1}^m C_i$$

each C_i is a **conjunction of literals**.

$$\left[\exists x.\bigvee_{i=1}^{m}C_{i}\right]\iff\bigvee_{i=1}^{m}(\exists x.C_{i})$$

Which steps should we use?

Which steps should we use? **Negation propagation:**

$$egreentcolor \neg (p \land q) \rightsquigarrow (\neg p) \lor (\neg q)$$

 $egreence \neg (p \lor q) \rightsquigarrow (\neg p) \land (\neg q)$
 $egreence \neg \neg p \rightsquigarrow p$

Result is **negation-normal form**, NNF NNF transformation is polynomial (exercise!)

Which steps should we use? **Negation propagation:**

$$eglinetized \neg (p \land q) \rightsquigarrow (\neg p) \lor (\neg q)
eglinetized \neg (p \lor q) \rightsquigarrow (\neg p) \land (\neg q)
eglinetized \neg \neg p \rightsquigarrow p$$

Result is **negation-normal form**, NNF NNF transformation is polynomial (exercise!) **Distributivity**

$$a \wedge (b_1 \vee b_2) \rightsquigarrow (a \wedge b_1) \vee (a \wedge b_2)$$

This can lead to exponential explosion. Can we obtain equivalent DNF formula without explosion?

Which steps should we use? **Negation propagation:**

$$eglinetized \neg (p \land q) \rightsquigarrow (\neg p) \lor (\neg q)
eglinetized \neg (p \lor q) \rightsquigarrow (\neg p) \land (\neg q)
eglinetized \neg \neg p \rightsquigarrow p$$

Result is **negation-normal form**, NNF NNF transformation is polynomial (exercise!) **Distributivity**

$$a \wedge (b_1 \vee b_2) \rightsquigarrow (a \wedge b_1) \vee (a \wedge b_2)$$

This can lead to exponential explosion. Can we obtain equivalent DNF formula without explosion? No! See exercise. Eliminating from quantifier free formulas

$$\exists x.F \iff \left[\exists x.\bigvee_{i=1}^m C_i\right] \iff \bigvee_{i=1}^m (\exists x.C_i)$$

$\exists x_1.\exists x_2.\underline{\exists \mathbf{x_3}}.F_0(x_1,x_2,x_3,\bar{y})$

$\exists x_1.\exists x_2.\underline{\exists \mathbf{x_3}}.F_0(x_1,x_2,x_3,\bar{y})$

$\exists x_1. \exists \mathbf{x_2}. F_1(x_1, x_2, \bar{y})$

$\exists x_1.\exists x_2.\underline{\exists \mathbf{x_3}}.F_0(x_1,x_2,x_3,\bar{y})$

$\exists x_1. \underline{\exists x_2}. F_1(x_1, x_2, \overline{y})$

$\underline{\exists \mathbf{x_1}}.F_2(x_1, \overline{y})$

$\exists x_1.\exists x_2.\underline{\exists \mathbf{x_3}}.F_0(x_1,x_2,x_3,\bar{y})$

$\exists x_1. \underline{\exists x_2}. F_1(x_1, x_2, \overline{y})$

$\underline{\exists \mathbf{x_1}}.F_2(x_1, \overline{y})$

$F_3(\bar{y})$

$\exists x_1.\exists x_2.\underline{\exists \mathbf{x_3}}.F_0(x_1,x_2,x_3,\bar{y})$

$\exists x_1.\underline{\exists \mathbf{x_2}}.F_1(x_1,x_2,\bar{y})$

$\underline{\exists \mathbf{x_1}}.F_2(x_1, \overline{y})$

F₃(ȳ) ☺

Universal Quantifiers

If $F_0(x, \bar{y})$ is quantifier-free, how to eliminate

 $\forall y.F_0(x,\bar{y})$

Universal Quantifiers

If $F_0(x, \bar{y})$ is quantifier-free, how to eliminate

 $\forall y.F_0(x,\bar{y})$

Equivalence (property always holds if there is no counterexample):

$$\forall y.F_0(x,\bar{y}) \iff \neg \big[\exists y.\neg F_0(x,\bar{y}) \big]$$

It thus suffices to process:

$$\neg [\exists y. \neg F_0(x, \bar{y})]$$

Note that $\neg F_0(x, \bar{y})$ is quantifier-free, so we know how to handle it:

$$\exists y. \neg F_0(x, \bar{y}) \quad \rightsquigarrow \quad F_1(\bar{y})$$

Therefore, we obtain

 $\neg F_1(\bar{y})$

Removing any alternation of quantifiers: illustration

Alternation: switch between existentials and universals

$$\exists x_{1}.\forall x_{2}.\forall x_{3}.\exists x_{4}.F_{0}(x_{1}, x_{2}, x_{3}, x_{4}, \bar{y})$$

$$\exists x_{1}.\neg \exists x_{2}.\exists x_{3}.\neg \exists x_{4}.F_{0}(x_{1}, x_{2}, x_{3}, x_{4}, \bar{y})$$

$$\exists x_{1}.\neg \exists x_{2}.\exists x_{3}.\neg F_{1}(x_{1}, x_{2}, x_{3}, \bar{y})$$

$$\exists x_{1}.\neg \exists x_{2}.F_{2}(x_{1}, x_{2}, \bar{y})$$

$$\exists x_{1}.\neg F_{3}(x_{1}, \bar{y})$$

$$F_{4}(\bar{y})$$

Each quantifier alternation involves a disjunctive normal form transformation.

In practice, we do not have many alternations.

Back to Presburger Arithmetic

Consider the quantifier elimination problem of the form:

$$\exists y. L_1 \land \ldots \land L_n$$

where L_i are literals from PA. Note that, for integers:

$$\blacktriangleright \neg (x < y) \iff y \le x$$

- $\blacktriangleright x < y \iff x + 1 \le y$
- $\blacktriangleright \ x \le y \iff x < y + 1$

We use these observations below. Instead of \leq we choose to use < only. We do not write x > y but only y < x.

Normalizing Literals for PA

Normal Form of Terms: All *terms* are built from $K, +, -, K \cdot ,$ so using standard transformations they can be represented as: $K_0 + \sum_{i=1}^n K_i x_i$ We call such term a linear term. **Normal Form for Literals in PA:**

$$egic{}
egic{}
egi$$

To remove disjunctions we generated, compute DNF again.

(*) We transformed equalities just for simplicity. Usually we handle them directly.

Why one-point rule will not be enough

Note that we must handle inequalities, not merely equalities

If we have integers, we cannot always divide perfectly. Variable to eliminate can occur not as y but as, e.g. 3y

Exposing the Variable to Eliminate: Example

$$\exists y. \ 0 < -w + \underline{3y} + 1 \ \land \ 0 < -\underline{2y} + z + 6 \ \land \ 4 \mid \underline{5y} + 1$$

Least common multiple of coefficients next to y, M = lcm(3, 2, 5) = 30Make all occurrences of y in the body have this coefficient:

$$\exists y. \ 0 < -10w + 30y + 10 \land 0 < -30y + 15z + 90 \land 24 \mid 30y + 6$$

Now we are quantifying over y and using 30y everywhere. Let x denote 30y. It is **not an arbitrary** x. It is divisible by 30.

$$\exists x. \ 0 < -10w + x + 10 \land \ 0 < -x + 15z + 90 \land 24 \mid x + 6 \land 30 \mid x$$

Exposing the Variable to Eliminate in General

Eliminating y from conjunction F(y) of literals:

- ▶ 0 < t
- ▶ *K* | *t*

where t is a linear term. To eliminate $\exists y$ from such conjunction, we wish to ensure that the coefficient next to y is one or minus one.

Observation:

- 0 < t is equivalent to 0 < c t
- $K \mid t$ is equivalent to $c K \mid c t$

for c a positive integer.

Let K_1, \ldots, K_n be all coefficients next to y in the formula. Let M be a positive integer such that $K_i \mid M$ for all $i, 1 \le i \le n$

► for example, let *M* be the **least common multiple**

$$M = lcm(K_1, \ldots, K_n)$$

Multiply each literal where y occurs in subterm $K_i y$ by constant $M/|K_i|$

• the point is, M is divisible by $|K_i|$ by construction

What is the coefficient next to y in the resulting formula?

Multiply each literal where y occurs in subterm $K_i y$ by constant $M/|K_i|$

► the point is, M is divisible by $|K_i|$ by construction What is the coefficient next to y in the resulting formula? M or -M

Multiply each literal where y occurs in subterm $K_i y$ by constant $M/|K_i|$

► the point is, M is divisible by $|K_i|$ by construction What is the coefficient next to y in the resulting formula? M or -M

We obtain a formula of the form $\exists y.F(M \cdot y)$. Letting x = My, we conclude the formula is equivalent to

 $\exists x. F(x) \land (M \mid x)$

What is the coefficient next to y in the resulting formula?

Multiply each literal where y occurs in subterm $K_i y$ by constant $M/|K_i|$

► the point is, M is divisible by $|K_i|$ by construction What is the coefficient next to y in the resulting formula? M or -M

We obtain a formula of the form $\exists y.F(M \cdot y)$. Letting x = My, we conclude the formula is equivalent to

$$\exists x. F(x) \land (M \mid x)$$

What is the coefficient next to y in the resulting formula? 1 or -1

Consider the coefficient next to x in 0 < t. If it is -1, move the term to left side. If it is 1, move the remaining terms to the left side. We obtain formula $F_1(x)$ of the form

$$\bigwedge_{i=1}^{L} \mathsf{a}_i < x \land \bigwedge_{j=1}^{U} x < \mathsf{b}_j \land \bigwedge_{i=1}^{D} \mathsf{K}_i \mid (x+t_i)$$