Automatic determination of numerical properties of software and systems

Eric Goubault and Sylvie Putot

Modelling and Analysis of Interacting Systems, CEA LIST

MASCOT-NUM 2012 Meeting, March 21-23, 2012

Eric Goubault and Sylvie Putot Modelling and Analysis of In Automatic determination of numerical properties of software and

Automatic validation of numerical programs

Context: safety-critical programs

- ▶ flight control or industrial install. (e.g. nuclear plants) control
- Iong-term objective: scientific computing
- guaranteed methods, that prove good behaviour or else try to give counter-examples

Starting point: bound the error due to floating-point numbers

- (deterministic) propagation of rounding errors in computations
- can also propagate any uncertainty on inputs or parameters
- ► Now: towards functional proof of algorithms, hybrid systems
- ► A few words on an extension to stochastic uncertainties

Static analysis of numerical programs

- Validate finite precision implementations: prove that the program computes something close to what is expected (in real numbers)
 - accuracy of results
 - behaviour of the program (control flow, number of iterations)
- But also validate algorithms: bound when possible the method/approximation error
 - ideally, find good match between method and implementation errors
- Automatically, given a source code, and for sets of (possibly uncertain) inputs and parameters: static analysis

Householder scheme for square root computation

Householder

Sylvie Putot MEASI, CEA LIST

Automatic determination of numerical properties of software an

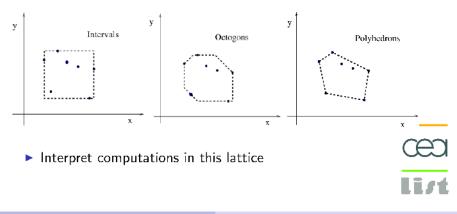
- Introduction to static analysis
- Static analysis of numerical programs
- Applications the Fluctuat static analyzer
- First steps towards imprecise probabilities

Static analysis

- Analysis of the source source, for a set of inputs and parameters, without executing it:
 - does the program always terminate; can it ever reach a bad state ?
 - is there a possibility of run-time error, such as division by zero?
 - worst case execution time?
 - synchronization errors (deadlocks, data races)?
 - does the program compute accurately?
- ▶ The ideal static analyzer (for run-time error for instance) is
 - sound: if there is an error, the analyzer reports it
 - complete: if the analyzer reports an error, it is a genuine one
- But any interesting program property is undecidable in general:
 - in general choose sound (but not complete)

Static analysis by abstract interpretation (Cousot 77)

- Choose properties of interest (for instance values of variables)
- Over-approximate them in an abstract lattice (partially ordered structure with least upper bounds/greatest lower bounds)



Example in intervals

- Program seen as a discrete dynamical system $X^{n+1} = F(Xn)$
 - based on a notion of control points in the program
 - equations describe how values of variables are collected at each control point, for all possible executions (collecting semantics)

void main() {
int x=[-100,50]; [1]
while [2] (x < 100) {
[3] x=x+1; [4]
} [5]
} (5)

$$x_0 = \top$$

 $x_1 = [-100,50]$
 $x_2 = x_1 \cup x_4$
 $x_3 =] -\infty,99] \cap x_2$
 $x_4 = x_3 + [1,1]$
 $x_5 = [100, +\infty[\cap x_2]$

- The sets of possible values of variables at control points are invariants of F, computed as the least fixpoint of the system
 - ► F monotonic on a complete lattice, least fixpoint exists

- Invariants allow to conclude about the safety (for instance absence of run-time errors) of programs
- Computing the least fixpoint X = F(X):
 - ▶ limit of the Kleene iteration (Jacobi/Gauss-Seidel like method) $X^0 = \bot, X^1 = F(X^0), ..., X^{k+1} = X^k \cup F(X^k)$
 - with convergence acceleration to terminate in finite time
 - or policy iteration (Newton-like method)

Outline of the talk

- Introduction to static analysis
- Static analysis of numerical programs
 - zonotopic abstract domains
 - modelling finite precision
- Applications the Fluctuat static analyzer
- First steps towards imprecise probabilities

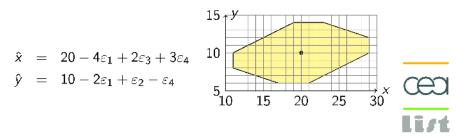
Abstraction based on Affine Arithmetic (Stolfi 93)

• The real value of variable x is represented by an affine form \hat{x} :

$$\hat{\mathbf{x}} = \mathbf{x}_0 + \mathbf{x}_1 \varepsilon_1 + \ldots + \mathbf{x}_n \varepsilon_n,$$

where $x_i \in \mathbb{R}$ and the ε_i are independent symbolic variables with unknown value in [-1, 1].

Sharing ε_i between variables expresses implicit dependency: concretization as a zonotope



Affine arithmetic

Assignment of a variable x whose value is given in a range [a, b] introduces a noise symbol ε_i :

$$\hat{x} = \frac{(a+b)}{2} + \frac{(b-a)}{2}\varepsilon_i.$$

 functional abstraction: link to the inputs via the noise symbols, allowing sensitivity analysis and worst case generation

Addition is computed componentwise (no new noise symbol):

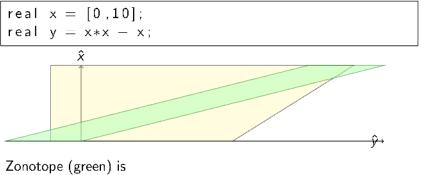
$$\hat{x} + \hat{y} = (\alpha_0^x + \alpha_0^y) + (\alpha_1^x + \alpha_1^y)\varepsilon_1 + \ldots + (\alpha_n^x + \alpha_n^y)\varepsilon_n$$

Multiplication : we select an approximate linear form, the approximation error creates a new noise term :

$$\hat{x} \times \hat{y} = \alpha_0^x \alpha_0^y + \sum_{i=1}^n (\alpha_i^x \alpha_0^y + \alpha_i^y \alpha_0^x) \varepsilon_i + (\sum_{i=1}^n |\alpha_i^x| | \sum_{i=1}^n |\alpha_i^y|) \varepsilon_{n+1}.$$

Automatic determination of numerical properties of software an

Example



$$x = 5 + 5\varepsilon_{1}$$

$$y = (5 + 5\varepsilon_{1})(5 + 5\varepsilon_{1}) - 5 - 5\varepsilon_{1}$$

$$= 20 + 45\varepsilon_{1} + 25\varepsilon_{1}^{2} = 20 + 45\varepsilon_{1} + 25(0.5 + 0.5\varepsilon_{2})$$

$$= 32.5 + 45\varepsilon_{1} + 12.5\varepsilon_{2}$$

Polyhedron (yellow) is

$$-x + 10 \ge 0; \ y + 10 \ge 0; \ x \ge 0; \ 4x - y + 50 \ge 0$$

Sylvie Putot MEASI, CEA LIST

Automatic determination of numerical properties of software a

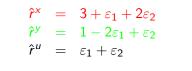
Affine arithmetic and static analysis

- Non linear operations: approximate linear form (Taylor expansion), new noise term for the approximation error
- close to Taylor models of low degree: because large ranges of values in general for static analysis
- possibly other polynomial approximations in the future

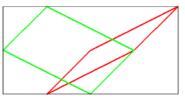
Set-theoretical operations

Partial order:

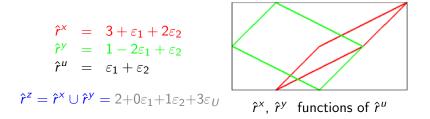
- geometric inclusion on zonotopes augmented by slack variables for initial values of variables
- functional order: preserves input-output relations
- Weaker structure than lattice: no least upper bound (join) or greatest lower bound (meet)
 - Join (to collect abstract values from different control flows): any upper bound of the union of zonotopes is sound
 - Meet (to interpret conditionals): we must over-approximate the intersection of zonotopes



 $\hat{r}^{z} = \hat{r}^{x} \cup \hat{r}^{y} = 2 + 0\varepsilon_{1} + 1\varepsilon_{2} + 3\varepsilon_{U}$

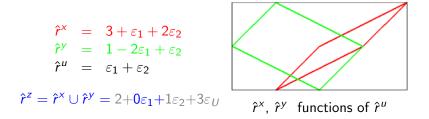


 \hat{r}^x , \hat{r}^y functions of \hat{r}^u



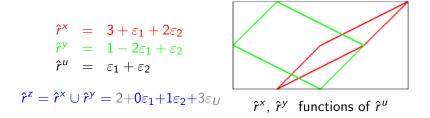
Keep "minimal common dependencies":

$$r_i^z = \underset{r_i^x \wedge r_i^y \le \alpha \le r_i^x \vee r_i^y}{\operatorname{argmin}} |\alpha|, \ \forall i \ge 1$$



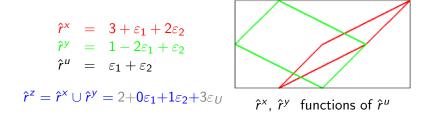
Keep "minimal common dependencies":

$$r_i^z = \underset{r_i^x \wedge r_i^y \le \alpha \le r_i^x \lor r_i^y}{\operatorname{argmin}} |\alpha|, \ \forall i \ge 1$$



Keep "minimal common dependencies":

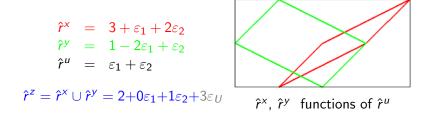
$$r_i^z = \underset{r_i^x \land r_i^y \le \alpha \le r_i^x \lor r_i^y}{\operatorname{argmin}} |\alpha|, \ \forall i \ge 1$$



For each var, the concretization is the interval union of the concretizations ($\gamma(\hat{z}) = [-2, 6]$) :

i)
$$r_0^z = mid(\gamma(\hat{r}^x) \cup \gamma(\hat{r}^y))$$

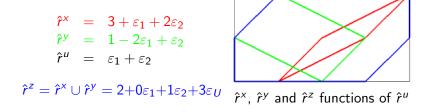
ii) $\beta^z = \sup \gamma(\hat{r}^x) \cup \gamma(\hat{r}^y) - r_0^z - \|\hat{r}^z\|_1$



For each var, the concretization is the interval union of the concretizations ($\gamma(\hat{z}) = [-2, 6]$) :

i)
$$r_0^z = mid(\gamma(\hat{r}^x) \cup \gamma(\hat{r}^y))$$

ii) $\beta^z = \sup \gamma(\hat{r}^x) \cup \gamma(\hat{r}^y) - r_0^z - \|\hat{r}^z\|_1$



For each var, the concretization is the interval union of the concretizations ($\gamma(\hat{z}) = [-2, 6]$) :

i)
$$r_0^z = mid(\gamma(\hat{r}^x) \cup \gamma(\hat{r}^y))$$

ii) $\beta^z = \sup \gamma(\hat{r}^x) \cup \gamma(\hat{r}^y) - r_0^z - \|\hat{r}^z\|_1$

Intersections: main idea, informally

real
$$x = [0, 10];$$

real $y = 2*x;$
if $(y \ge 10)$
 $y = x;$

- Affine forms before tests: $x = 5 + 5\varepsilon_1$, $y = 10 + 10\varepsilon_1$
- Main idea to interpret tests:
 - translate the condition in the space of noise symbols: $\varepsilon_1 \ge 0$
 - abstract domain for the noise symbols: intervals, octagons, etc.
 - result may no longer be a zonotope
- Here ε₁ ∈ [0, 1] in the then branch, implies y ∈ [0, 10] at the end of the program

Floating-point numbers (IEEE 754 norm)

- Limited range and precision : potentially inaccurate results or run-time errors
- A few figures for simple precision normalized f.p. numbers :
 - largest $\sim 3.40282347 * 10^{38}$
 - smallest positive $\sim 1.17549435 * 10^{-38}$
 - max relative rounding error $= 2^{-23} \sim 1.19200928955 * 10^{-7}$
- Consequences:
 - potentially non intuitive representation error:
 - $\frac{1}{10} = 0.00011001100110011001100 \cdots$ (binary)
 - \Rightarrow float $(\frac{1}{10}) = 0.100000014901161194 \cdots$ (decimal)
 - absorption : $1 + 10^{-8} = 1$ in simple precision float
 - ▶ associative law not true : $(-1+1) + 10^{-8} \neq -1 + (1+10^{-8})$
 - cancellation: loss of relative accuracy if subtracting close nb

In real world : costly or catastrophic examples

- 25/02/91: a Patriot missile misses a Scud and crashes on an american building : 28 deads.
 - the missile program had been running for 100 hours, incrementing an integer every 0.1 second
 - but 0.1 not representable in a finite number of digits in base 2
 - Drift on 100 hours ~ 0.34s : location error ~ 500m
- Explosion of Ariane 5 in 1996 (conversion of a 64 bits float into a 16 bits integer : overflow)
- ▶ An index of the Vancouver stock exchange in 1982
 - truncated at each transaction : errors all have same sign
 - within a few months : lost half of its correct value
- Sinking of an offshore oil platform in 1992 : inaccurate finite element approximation

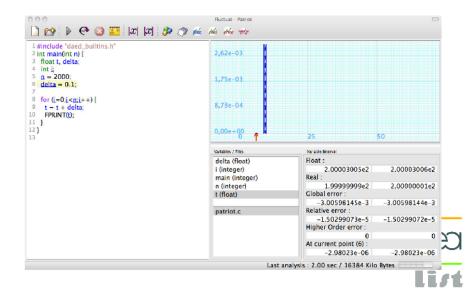
Our model for the analysis of numerical computations

- Aim: compute rounding errors and their propagation
 - for each variable, we compute (f^x, r^x, e^x)
 - then we will abstract each term (real value and errors)

float x,y,z;
x = 0.1; // [1]
y = 0.5; // [2]
z = x+y; // [3]
t = x*z; // [4]

$$f^{x} = 0.1 + 1.49e^{-9}$$
 [1]
 $f^{y} = 0.5$
 $f^{z} = 0.6 + 1.49e^{-9}$ [1] + 2.23 e^{-8} [3]
 $f^{t} = 0.06 + 1.04e^{-9}$ [1] + 2.23 e^{-9} [3] - 8.94 e^{-10} [4] - 3.55 e^{-17} [ho]

Example



Evolution of the error on t



Towards an abstract model

- ▶ IEEE norm on f.p. numbers allows to prove properties:
 - elementary rounding error when rounding r^{x} to $\uparrow_{\circ} r^{x}$:

 $\exists (\delta_r > 0, \delta_a > 0), \ |\mathbf{r}^{\mathsf{x}} - \uparrow_{\circ} \mathbf{r}^{\mathsf{x}}| \leq \max(\delta_r |\uparrow_{\circ} \mathbf{r}^{\mathsf{x}}|, \delta_a)$

- ▶ the f.p. result of arithmetic elementary operations $+, -, \times, /, \sqrt{}$ is the rounded value of the real result
- ► For each variable x, a triplet (f^x, r^x, e^x):
 - r^x is an abstraction of the real (ideal) value of x
 - f^x is an abstraction of the machine (finite precision) value of x
 - *e*^x is an abstraction of the difference
- Abstraction of r^{x} , e^{x} and f^{x} using set-membership methods
 - the simplest: intervals
 - affine forms (zonotopes)

Full abstraction

Abstract value: an interval and two affine forms $x = (\mathbf{f}^x, \hat{\mathbf{r}}^x, \hat{\mathbf{e}}^x)$:

$$\hat{r}^{x} = r_{0}^{x} + \sum_{i} r_{i}^{x} \varepsilon_{i}^{r}$$
$$\hat{e}^{x} = e_{0}^{x} + \sum_{i} e_{i}^{x} \varepsilon_{i}^{r} + \sum_{l} e_{l}^{x} \varepsilon_{l}^{e}$$

• $\mathbf{f}^{x} = [\underline{f^{x}}, \overline{f^{x}}]$ bounds the finite prec value, $(\underline{f^{x}}, \overline{f^{x}}) \in \mathbb{F} \times \mathbb{F}$,

- e^x_l ε^e_l: uncertainty on the rounding error committed at point *l* of the program (its center being in e^x₀), and its propagation through further computations,
- e^x_i e^r_i: propagation of the uncertainty on value at point i, on the error term,
- dependency between errors and values partially modelled

Outline of the talk

- Introduction to static analysis
- Static analysis of numerical programs

Applications - the Fluctuat static analyzer

First steps towards imprecise probabilities

FLUCTUAT

- Takes source C code (most of ANSI C, except union types and malloc most notably), with assertions (for instance range of values and imprecision on input, but also range of gradient of evolution of values)
- Gives, fully automatically, characterization of ranges/errors, and describe the origins of errors: identification of pieces of code with numerical difficulties
- Is/has been used for a wide variety of codes (automotive, nuclear industry, aeronautics, aerospace) of size up to about 50000 LOCs (on laptop PCs 1Gb)
- Academic version available upon request

Second order filters

Filters

Sylvie Putot MEASI, CEA LIST

Automatic determination of numerical properties of software a

Back to the Householder scheme

Householder

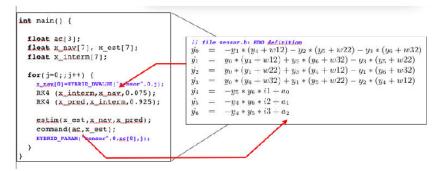
Sylvie Putot MEASI, CEA LIST

Automatic determination of numerical properties of software an

First extension to hybrid systems analysis

- Classical program analysis: inputs given in ranges, possibly with bounds on the gradient between two values
 - Behaviour is often not realistic
- Hybrid systems analysis: analyze both physical environment and control software for better precision
 - Environment modelled by switched ODE systems
 - abstraction by guaranteed integration (the solver is guaranteed to over-approximate the real solution)
 - Interaction between program and environment modelled by assertions in the program
 - sensor reads a variable value at time t from the environment,
 - actuator sends a variable value at time t to the environment,
- Other possible use of guaranteed integration in program analysis: bound method error of ODE solvers

Example: the ATV escape mechanism



- Time is controlled by the program (j)
- Program changes parameters (HYBRID_PARAM: actuators) or mode (not here) of the ODE system
- Program reads from the environment(HYBRID_DVALUE: sensors) by calling the ODE guaranteed solver

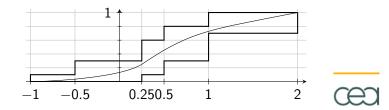
Could demonstrate convergence towards the safe escape state.

Outline of the talk

- Introduction to static analysis
- Static analysis of numerical programs
- Applications the Fluctuat static analyzer
- First steps towards imprecise probabilities

Combining deterministic and probabilistic methods

- Inputs known in intervals or by probability distribution:
 - e.g. temperature distribution, but only a range for pressure
- More generally, inputs given by uncertain probabilities
 - Discrete p-boxes or Dempster-Shafer structures
 - Generalize probabilities and interval computations
 - ▶ Represent sets of probability distributions: between an upper and a lower Cumulative Distribution Function P(X ≤ x)



Less pessimistic but still guaranteed abstractions

Arithmetic on Dempster-Shafer structures

- Starting point: arithmetic on discrete p-boxes or Dempster-Shafer structures not very efficient
 - rely on either independence or unknown dependency between variables
 - arithmetic for unknown dependency is costly and inaccurate
- $\begin{array}{l|l} X = input(); \\ Y = input(); \\ Z = X+Y; \\ T = Z+X; \end{array} \begin{array}{l} \text{New DS structure} \\ \text{New (independent) DS structure} \\ \text{Independent arithmetic} \\ \text{Must use unknown dependency(!)} \end{array}$

Probabilistic affine forms

- Encode as much deterministic dependencies as possible by affine arithmetic
- Associate a DS structure to each noise symbol
- Only the non affine operations create DS structures (noise symbols) with dependency to some other symbols

 $\begin{array}{l|l} \mathsf{X} = \mathsf{input}(); & \mathsf{X} = \varepsilon_1 : \mathsf{new} \; \mathsf{DS} \; \mathsf{structure} \\ \mathsf{Y} = \mathsf{input}(); & \mathsf{Y} = \varepsilon_2 : \mathsf{new} \; (\mathsf{independent}) \; \mathsf{DS} \; \mathsf{structure} \\ \mathsf{Z} = \mathsf{X} + \mathsf{Y}; & \mathsf{Z} = \varepsilon_1 + \varepsilon_2 \\ \mathsf{T} = \mathsf{Z} + \mathsf{X}; & \mathsf{T} = 2\varepsilon_1 + \varepsilon_2 : \; \mathsf{DS} \; \mathsf{of} \; \mathsf{T} \; \mathsf{uses} \; \mathsf{independent} \; \mathsf{arithmetic} \end{array}$

▶ Both more accurate and faster than direct DS arithmetic:

- more accurate: relation exactly encoded
- faster: operations on DS structures (the costly part) performed only when the DS of a variable is needed (end of the program or non linear operations)

References: abstract domains, Fluctuat & cases studies

- Static Analysis of Numerical Algorithms, SAS 2006 (Static Analysis Symposium)
- The Zonotope Abstract Domain Taylor1+, CAV 2009 (Computer Aided Verification) and its extension in CAV 2010
- Static Analysis of Finite Precision Computations, VMCAI 2011 (Verification, Model-Checking and Abstract Interpretation)
- A generalization of P-boxes to affine arithmetic, *Computing*, 2012
- HybridFluctuat: A Static Analyzer of Numerical Programs within a Continuous Environment, CAV 2009
- Validation using Abstract Interpretation (with ESA, ASTRIUM SAS, ENS), DASIA 2009 (DAta Systems In Aerospace Space Software)
- Towards an industrial use of FLUCTUAT on safety-critical avionics software (with Airbus), FMICS 2009

Automatic determination of numerical properties of software an

Some related work and tools for f.p.programs

Tools dedicated to the analysis of f.p. behavior:

- CADNA: roundoff error estmation by stochastic testing http://www-anp.lip6.fr/cadna/
- GAPPA: automatic proof generation of arithmetic properties http://lipforge.ens-lyon.fr/www/gappa/

Static analysis tools that take into account f.p. semantics:

- ASTREE: static analysis of run-time error http://www.astree.ens.fr/
- FRAMA-C: platform for source-code analysis of C software http://frama-c.cea.fr/index.html
- POLYSPACE: static analysis of run-time error www.mathworks.com/products/polyspace/

Contact: {Eric.Goubault,Sylvie.Putot}@cea.fr