Symbolic Computation
Algebraic Biology I

Bud Mishra

Courant Inst, NYU

NYU SoM, TIFR, MSSM

Systems Biology

- Introduction to Biology
- Regulatory & Metabolic Processes
- Algebraic Models in Biology

Symbolic Computation Algebraic Biology II

Bud Mishra

Courant Inst, NYU

NYU SoM, TIFR, MSSM

Model Checking

- Temporal Logic
- Kripke Models
- Model Checking
- Biologically Faithful Models

Symbolic Computation
Algebraic Biology III

Bud Mishra
Courant Inst, NYU
NYU SoM, TIFR, MSSM

Semi-Algebraic Geometry

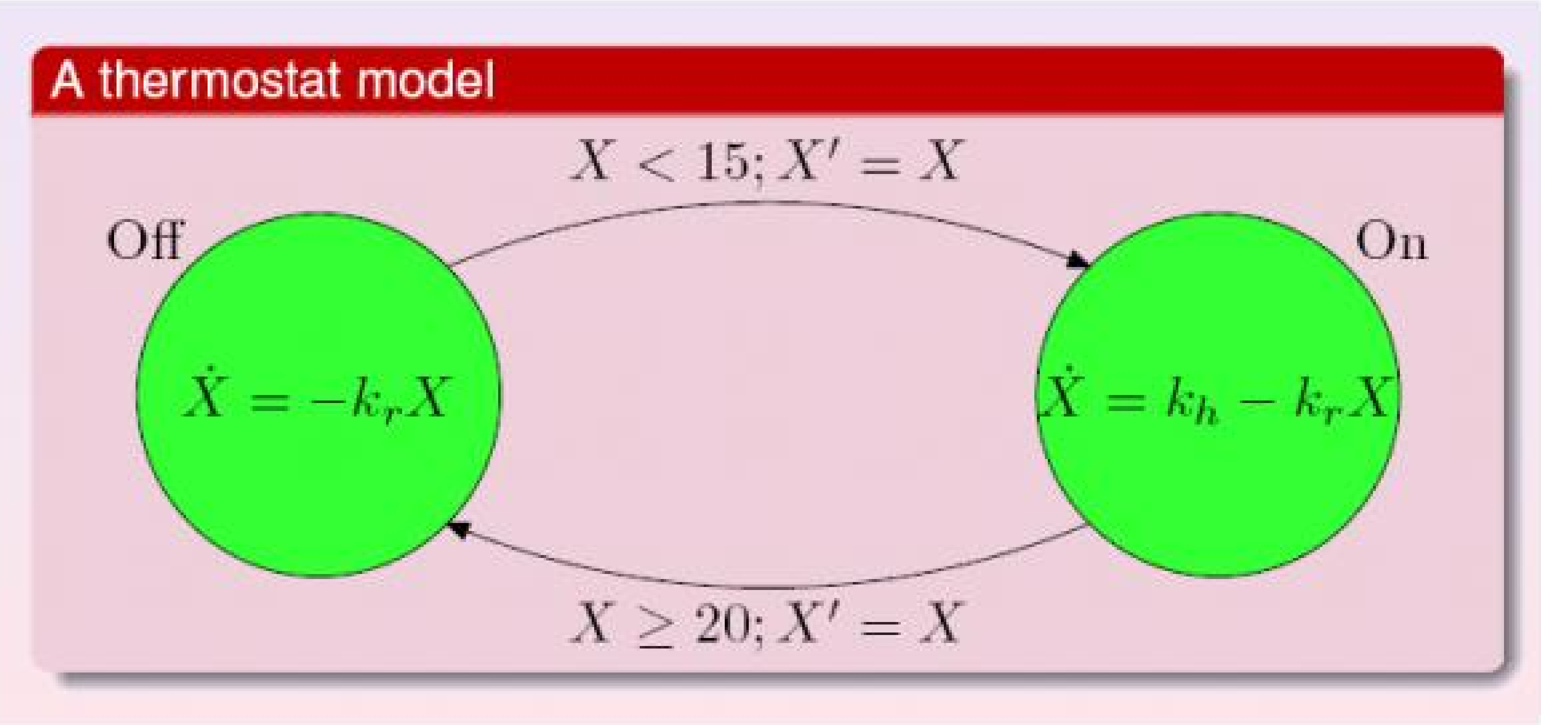
- Real Closed Field
- Tarski Algebra
- Decision Theories
- Hybrid Models
- Algorithmic Algebraic Model

Hybrid Automaton

- A hybrid automaton (of dimension k) H = $\langle$ Z, Z′, V, E, Inv, Dyn, Act , Reset $\rangle$ (over M), consists of the following components:

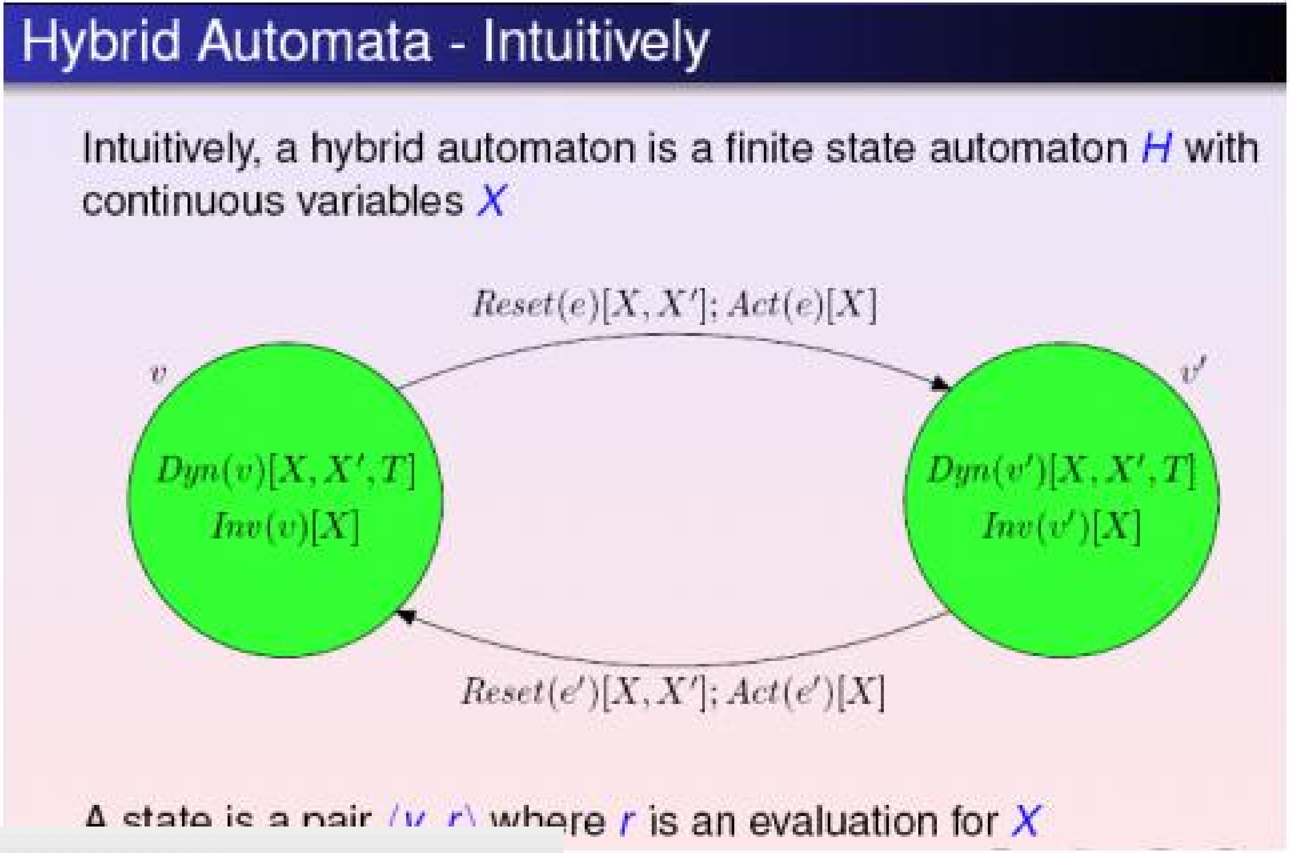
  1. $Z = (Z_1, \dots, Z_k)$ and $Z' = (Z'_1, \dots, Z'_k)$ are two vectors of variables ranging over the reals, $\mathbb{R}$;
  2. $\langle$ **V**, **E** $\rangle$ is a finite directed graph; the vertices of V are called locations, or control modes, the directed edges in E, control switches;
  3. Each v ∈ **V** is labeled by the two formulæ Inv(v)[Z] and Dyn(v)[Z,Z′, T] such that if Inv(v)[p] holds (in M), then Dyn(v)[p, p, 0] holds as well;
  4. Each e ∈ **E** is labeled by the formulæ Act(e)[Z] and Reset(e)[Z,Z′].

Thermostat



A thermostat model

Intuition

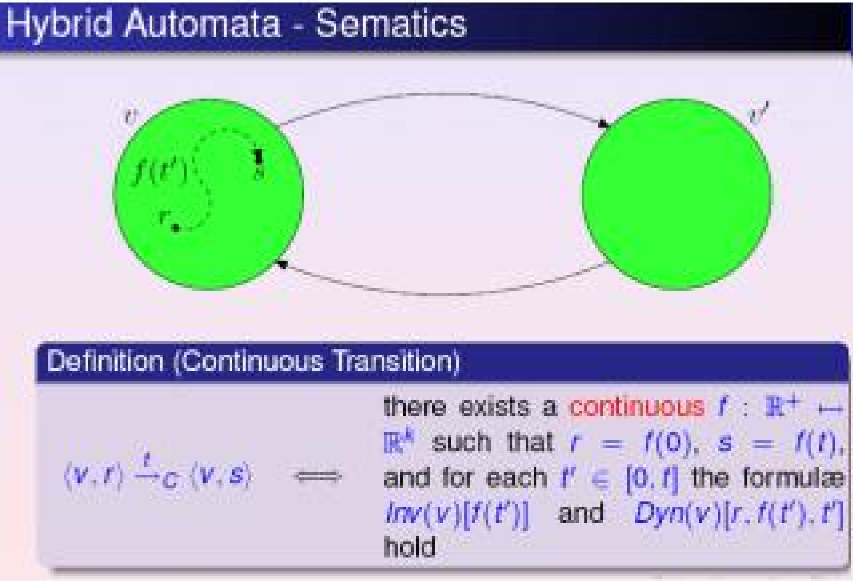


Hybrid Automata - Intuitively

Intuitively, a hybrid automaton is a finite state automaton *H* with continuous variables *X*

A state is a pair $\langle v, r \rangle$ where *r* is an evaluation for *X*

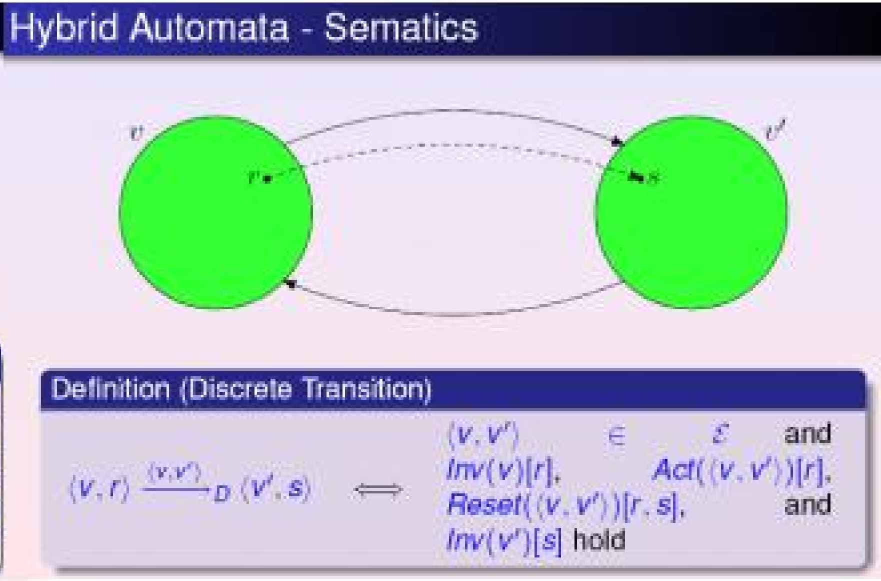
Semantics



Hybrid Automata - Sematics

Definition (Continuous Transition)

$\langle v, r\rangle \xrightarrow{t}_C \langle v, s\rangle \iff$ there exists a continuous $f : \mathbb{R}^+ \mapsto \mathbb{R}^k$ such that $r = f(0)$, $s = f(t)$, and for each $t' \in [0, t]$ the formulae $Inv(v)[f(t')]$ and $Dyn(v)[r, f(t'), t']$ hold

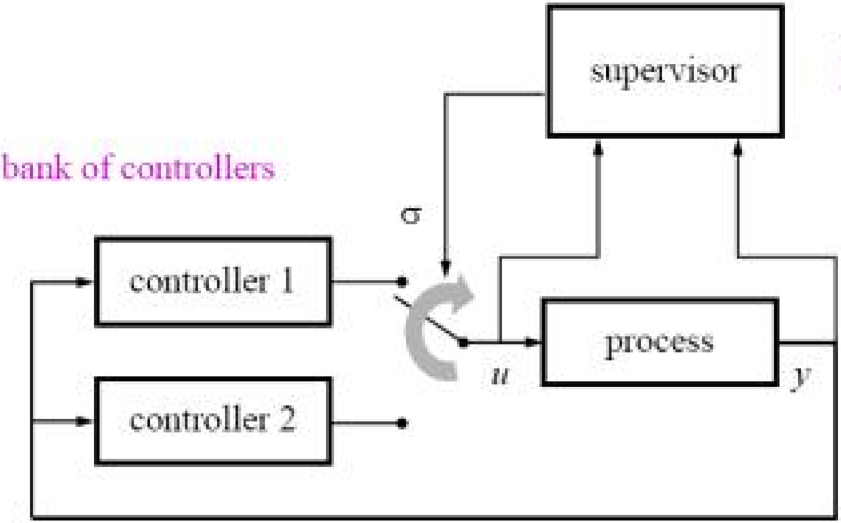


Hybrid Automata - Sematics

Definition (Discrete Transition)

$\langle v, r\rangle \xrightarrow{\langle v,v'\rangle}_D \langle v', s\rangle \iff \langle v, v'\rangle \in \mathcal{E}$ and $Inv(v)[r]$, $Act(\langle v, v'\rangle)[r]$, $Reset(\langle v, v'\rangle)[r, s]$, and $Inv(v')[s]$ hold

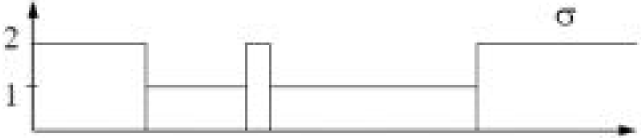
Engineered Systems



bank of controllers

logic that selects which controller to use

σ ≡ switching signal taking values in the set {1,2}



Reachability

- Let H be a hybrid automaton of dimension k. A point $r \in \mathbb{R}^k$ reaches a point $s \in \mathbb{R}^k$ (in time t) if there exists a trace $tr = \langle v, r \rangle, \ldots, \langle u, s \rangle$, for some $v, u \in V$ (and t is simply the sum of the elapsed times in continuous transitions).
  - We use ReachSet (r) to denote the set of points reachable from r. Moreover, given a region $R \subseteq R^k$ we use ReachSet (R) to denote the set $\cup_{r \in R}$ ReachSet (r). □

Decidability

- It has been shown that "**hybrid automata reachability problem**" is not decidable.
- Characterizing subclasses of hybrid automata over which reachability is decidable
- A common approach for deciding reachability of hybrid automata employs the technique of discretizing the automata using
  - **bisimulation**: equivalence relations which strongly preserve reachability
  - **abstractions** (e.g., predicate abstraction).

Examples

- Examples: timed automata, multirate automata, rectangular automata, and o-minimal automata…
- Rectangular automata are special cases of **linear hybrid automata**
- For a linear hybrid automata, its dynamics, invariants, and activation relations are all defined by linear expressions over the set Z of variables.

Linear Hybrid Automata

- For the control modes
  - The dynamics is defined by a differential equation of the form dz/dt = k, where k is a constant, one for each variable in Z
  - The invariants are defined by linear equalities and inequalities (corresponding to a convex polyhedron) in Z.
- For each transition, the set of reset assignments consists of linear formulæ in Z.
- Its trajectory is a piecewise linear function whose values at the points of discontinuity are finite sequences of discrete changes.

Nonlinear Hybrid Automata

- Changing linear descriptions to higher order algebraic descriptions…
- Semialgebraic Geometry
- Decidability through finite description via Tarski Algebra…

Computational Semialgebraic Geometry

- Study of various algorithmic questions dealing with the real solutions of **a system of equalities, inequalities, and inequations of polynomials over the real numbers**.
  - It is largely motivated by its applications to biology, robotics, vision, computer-aided design, geometric theorem proving, etc.

Tarski Formulas & Tarski Sentences

- **Tarski formulas** are formulas in a first-order language (***defined by Tarski in 1930***) *constructed from equalities, inequalities, and inequations of polynomials over the reals*.
- Such formulas may be *constructed by introducing logical connectives and universal and existential quantifiers to the atomic formulas*.
- **Tarski sentences** are Tarski formulas in which all variables are bound by quantification.

Theorem

- *Let Ψ be a Tarski sentence. There is an effective decision procedure for Ψ.*

  Let Ψ be a Tarski formula. There is a quantifier-free formula Φ logically equivalent to Ψ.
- *If Ψ involves only polynomials with rational coefficients, then so does the sentence Φ.* □

Glossary

- **Term**: A constant, variable, or term combining two terms by an arithmetic operator: $\{+, -, \cdot, /\}$. A constant is a real number. A variable assumes a real number as its value. A term contains finitely many such algebraic variables: $x_1, x_2, \ldots, x_n$.
- **Atomic formula**: A formula comparing two terms by a binary relational operator: $\{=, \neq, >, <, \geq, \leq\}$.

Glossary

- **Quantifier-free formula**: An atomic formula, a negation of a quantifier-free formula given by the unary Boolean connective {¬}, or a formula combining two quantifier-free formulas by a binary Boolean connective: {⇒, ∧, ∨}.
  - Example: The formula $(x^2 - 2 = 0) \wedge (x > 0)$ defines the (real algebraic) number $+\sqrt{2}$.

Glossary

- **Tarski formula:** If $\Phi(y_1, \ldots, y_r)$ is a quantifier-free formula, then it is also a Tarski formula. All the variables y_i are free in Φ. Let $(y_1, \ldots, y_r)$ and $(z_1, \ldots, z_s)$ be two Tarski formulas (with free variables y_i and z_i, respectively), then a formula combining and by a Boolean connective is a Tarski formula with free variables $\{y_i\} \cup \{z_i\}$. Lastly, if Q stands for a quantifier (either universal $\forall$ or existential $\exists$) and if $(y_1, \ldots, y_r, x)$ is a Tarski formula (with free variables x and y's), then

$$(Q\ x)[\Phi(y_1, \ldots, y_r, x)]$$

is a Tarski formula with only the y's as free variables. The variable x is bound in $(Q\ x)[\Phi]$.

Glossary

- **Tarski sentence:** A Tarski formula with no free variable.
  - Example: $(\exists x)\ (\forall y)\ [y^2 - x < 0]$. This Tarski sentence is false.
- **Prenex Tarski formula:** A Tarski formula of the form

  $$(Q\ x_1)\ (Q\ x_2) \cdots (Q\ x_n)\ [\Phi(y_1, y_2, \ldots, y_r, x_1, \ldots, x_n)],$$

  where ϕ is quantifier-free. The string of quantifiers $(Q\ x_1)\ (Q\ x_2) \cdots (Q\ x_n)$ is called the ***prefix*** and Φ is called the ***matrix***.
- **Prenex form of a Tarski formula, Ψ:** A prenex Tarski formula logically equivalent to Ψ.

Glossary

- For every Tarski formula, one can find its prenex form using a simple procedure that works in four steps: (1) eliminate redundant quantifiers, (2) rename variables so that the same variable does not occur as free and bound, (3) move negations inward; and finally, (4) push quantifiers to the left.
- **Extension of a Tarski formula, $\Phi(y_1, \ldots, y_r)$** with free variables $\{y_1, \ldots, y_r\}$: The set of all $\langle \zeta_1, \ldots, \zeta_r \rangle \in \mathbb{R}^r$ such that

$$\boldsymbol{\Phi(\zeta_1, \ldots, \zeta_r) = \textbf{True}}.$$

General Decision Problem for the First-order Theory of Reals

- **The general decision problem for the first-order theory of reals:** is to determine if a given Tarski sentence is true or false.
- **The existential problem for the first-order theory of reals:** An interesting special case of the problem is when all the quantifiers are existential.
- The general decision problem was shown to be decidable by Tarski [1930; published 1951].

Complexity Issues

- Tarski's original algorithm has a high complexity: a very rapidly-growing function of the input size
  - (e.g., it could not be expressed as a bounded tower of exponents of the input size).
- The first substantial improvement over Tarski's algorithm was due to Collins [1975]
  - doubly-exponential time complexity in the input size—the number of variables appearing in the sentence.
- Further improvements
  - (Grigor'ev-Vorobjov [1988], Canny [1988-93], Heintz et al. [1989-90], Renegar [1992])
  - Basu et al. [1994].

Algorithmic Complexity

- Assume that a Tarski sentence is presented in its prenex form:

$$(Q_1 x^{[1]})\ (Q_2 x^{[2]}) \cdots (Q_{\varpi}\ x^{[\varpi]})\ [\Psi(x^{[1]}, \ldots, x^{[\varpi]})],$$

where the Q_i's form a sequence of alternating quantifiers (i.e., ∀ or ∃, with every pair of consecutive quantifiers distinct), with $x^{[i]}$ a partition of the variables

$$\cup_{i=0}^{\varpi} x^{[i]} = \{x_1, x_2, \ldots, x_n\},\ x, \text{ and } |x^{[i]}| = n_i,$$

and where Ψ is a quantifier-free formula with atomic predicates consisting of polynomial equalities and inequalities of the form

$$g_i\ (x^{[1]}, \ldots, x^{[\varpi]} \overset{\geq}{\underset{\leq}{=}} 0,\ i = 1, \ldots, m.$$

Bit-complexity of the Decision Problem

- Here, g_i is a multivariate polynomial (over $\mathbb{R}$ or $\mathbb{Q}$, as the case may be) of total degree bounded by d.
- There are a total of m such polynomials.
- The special case $\omega = 1$ reduces the problem to that of the existential problem for the first-order theory of reals.
- If the polynomials of the basic equalities, inequalities, inequations, etc., are over the rationals, then we assume that their coefficients can be stored with at most L bits. Thus the arithmetic complexity can be described in terms of n, n_i, ω, m, and d, and the bit complexity will involve L as well.

Bit-complexity of the Decision Problem

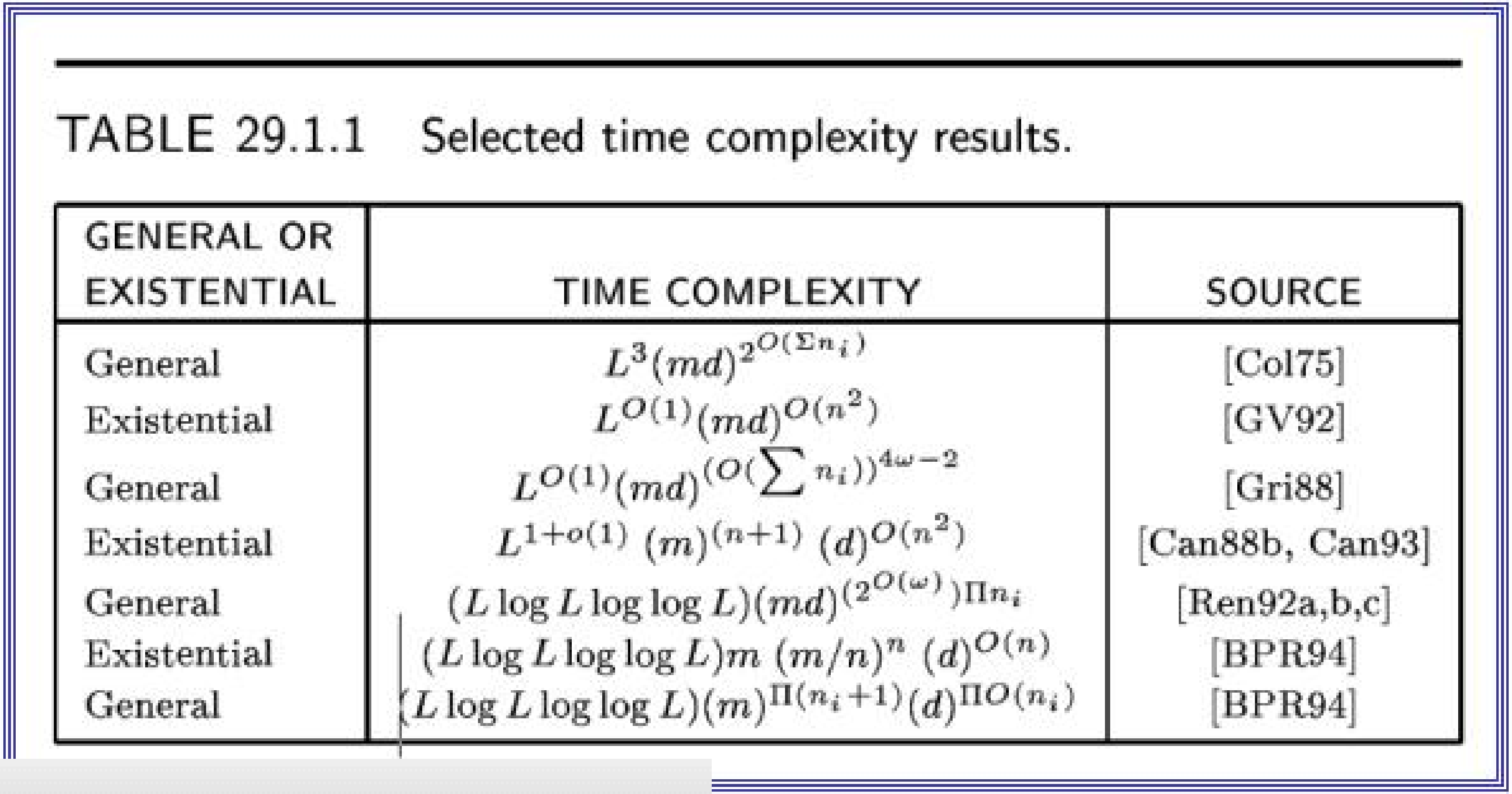


TABLE 29.1.1 Selected time complexity results.

GENERAL OR EXISTENTIAL	TIME COMPLEXITY	SOURCE
General	$L^3(md)^{2^{O(\Sigma n_i)}}$	[Col75]
Existential	$L^{O(1)}(md)^{O(n^2)}$	[GV92]
General	$L^{O(1)}(md)^{(O(\sum n_i))^{4\omega-2}}$	[Gri88]
Existential	$L^{1+o(1)}\ (m)^{(n+1)}\ (d)^{O(n^2)}$	[Can88b, Can93]
General	$(L \log L \log\log L)(md)^{(2^{O(\omega)})\Pi n_i}$	[Ren92a,b,c]
Existential	$(L \log L \log\log L)m\ (m/n)^n\ (d)^{O(n)}$	[BPR94]
General	$(L \log L \log\log L)(m)^{\Pi(n_i+1)}(d)^{\Pi O(n_i)}$	[BPR94]

Quantifier Elimination Problem

- Given a Tarski formula of the form,

$$\Psi(x^{[0]}) = (Q_1\, x^{[1]})\, (Q_2\, x^{[2]}) \cdots (Q_\omega\, x^{[\omega]})\, [\psi(x^{[0]}, x^{[1]}, \ldots, x^{[\omega]})] ,$$

  where ψ is a quantifier-free formula, the quantifier elimination problem is to construct another quantifier-free formula, $\phi(x^{[0]})$, such that $\phi(x^{[0]})$ holds if and only if $\phi(x^{[0]})$ holds.

Quantifier-Free Formula

- Such a quantifier-free formula takes the form

$$\phi(x^{[0]}) \equiv \vee_{i=1}^{I} \wedge_{j=1}^{J_i} f_{i,j}(x^{[0]}) \overset{>}{\underset{<}{=}} 0,$$

where $f_{i,j} \in \mathbb{R}[x^{[0]}]$ is a multivariate polynomial with real coefficients.

- Significantly improved bounds were given by Basu, Polack & and are summarized next

$$I \leq (m)^{\Pi\ (n_i+1)}(d)^{\Pi\ O(n_i)}$$

$$J_i \leq (m)^{\Pi_{i>0}\ (n_i+1)}(d)^{\Pi_{i>0}\ O(n_i)}.$$

- The total degrees of the polynomials $f_{i,j}(x^{[0]})$ are bounded by

$$(d)^{\Pi_{i>0}\ O(n_i)}.$$

Quantifier-Free Formula

- The best bound for the size of the equivalent quantifier-free formula is now

$$I, J_i \leq (m)^{\Pi_{i>0}\,(n_i+1)}(d)^{n'_0\,\Pi_{i>0}\,O(n_i)},$$

- where $n'_0 = \min(n_0, \tau \Pi_{i>0}(n_i+1))$ and τ is a bound on the number of free-variables occurring in any polynomial in the original Tarski formula. The total degrees of the polynomials $f_{i,j}(x^{[0]})$ are still bounded by

$$(d)^{\Pi_{i>0}\,O(ni)}.$$

Quantifier-Free Formula

- Furthermore, the algorithmic complexity of the new procedure involves only

$$(m)^{\Pi_{i>0}\,(n_i+1)}(d)^{n'_0\,\Pi_{i>0}\,O(n_i)}$$

arithmetic operations.

Glossary

- **Semialgebraic Set:** A subset $S \subseteq \mathbb{R}^n$ defined by a set-theoretic expression involving a system of polynomial inequalities

  $$S = \cup_{i=1}^{I} \cap_{j=1}^{J_i} \{ \langle \xi_1, \ldots, \xi_n \rangle \in \mathbb{R}^n \mid \mathrm{sgn}(f_{i,j}(\xi_1, \ldots, \xi_n)) = s_{i,j} \},$$

  - where the $f_{i,j}$'s are multivariate polynomials over R and the $s_{i,j}$'s are corresponding sets of signs in $\{-1, 0, +1\}$.

- **Real algebraic set:** A subset $Z \subseteq \mathbb{R}^n$ defined by a system of algebraic equations.

  $$Z = \{ \langle \xi_1, \ldots, \xi_n \rangle \in \mathbb{R}^n \mid f_1(\xi_1, \ldots, \xi_n) = \cdots = f_m(\xi_1, \ldots, \xi_n) = 0 \},$$

  - where the f_i's are multivariate polynomials over $\mathbb{R}$.

Glossary

- **Semialgebraic decomposition of a semialgebraic set S:** A finite collection $\mathcal{K}$ of disjoint connected semialgebraic subsets of S whose union is S. The collection of connected components of a semialgebraic set forms a semialgebraic decomposition. Thus, every semialgebraic set admits a semialgebraic decomposition.
- **Set of sample points for S:** A finite number of points meeting every nonempty connected component of S.

Glossary

- **Sign assignment:** A vector of sign values of a set of polynomials at a point p. More formally, let F be a set of real multivariate polynomials in n variables. Any point $p = \langle \xi_1, \ldots, \xi_n \rangle \in \mathbb{R}^n$ has a sign assignment with respect to $\mathcal{F}$ as follows:

$$\mathbf{sgn}_{\mathcal{F}}(\mathbf{p}) = \langle\ \mathbf{sgn}(\mathbf{f}(\xi_1, \ldots, \xi_n))\ |\ \mathbf{f} \in \mathcal{F}\ \rangle.$$

- *A sign assignment induces an equivalence relation:* Given two points $p, q \in \mathbb{R}^n$, we say $p \sim_{\mathcal{F}} q$, if and only if $\mathrm{sgn}_{\mathcal{F}}(p) = \mathrm{sgn}_{\mathcal{F}}(q)$.

Glossary

- **Sign class of $\mathcal{F}$:** An equivalence class in the partition of $\mathbb{R}^n$ defined by the equivalence relation $\sim_{\mathcal{F}}$.
- **Semialgebraic decomposition for $\mathcal{F}$:** A finite collection of disjoint connected semialgebraic subsets $\{C_i\}$ such that each C_i is contained in some semialgebraic sign class of $\mathcal{F}$. That is, the sign of each $f \in \mathcal{F}$ is invariant in each C_i. The collection of connected components of the sign-invariant sets for $\mathcal{F}$ forms a semialgebraic decomposition for $\mathcal{F}$.

Glossary

- **Cell decomposition for $\mathcal{F}$:** A semialgebraic decomposition for $\mathcal{F}$ into finitely many disjoint semialgebraic subsets $\{C_i\}$ called cells, such that each cell C_i is homeomorphic to $\mathbb{R}^{\delta(i)}$, $0 \leq \delta(i) \leq n$. $\delta(i)$ is called the dimension of the cell C_i, and C_i is called a $\delta(i)$-cell.
- **Cellular decomposition for $\mathcal{F}$:** A cell decomposition for $\mathcal{F}$ such that the closure C_i of each cell C_i is a union of cells C_j : $C^*_i = \cup_j C_j$.

Univariate Decomposition

- One-dimensional case: A semialgebraic set is the union of finitely many intervals whose endpoints are real algebraic numbers.
- Given a set of univariate defining polynomials:

$$\mathcal{F} = \{ \mathbf{f}_i(x) \in \mathbb{Q}[x] \mid \mathbf{i} = \mathbf{1}, \ldots, \mathbf{m}\},$$

  we may enumerate all the real roots of the fi's (i.e., the real roots of the single polynomial $\mathcal{F} = \prod f_i$) as

$$-\infty < \xi_1 < \xi_2 < \cdots < \xi_{i-1} < \xi_i < \xi_{i+1} < \cdots < \xi_s < +\infty,$$

- Consider the following finite set $\mathcal{K}$ of elementary intervals defined by these roots:

$$[-\infty, \xi_1), [\xi_1, \xi_1], (\xi_1, \xi_2), \ldots, (\xi_{i-1}, \xi_i), [\xi_i, \xi_i], (\xi_i, \xi_{i+1}), \ldots, [\xi_s, \xi_s], (\xi_s, +\infty].$$

Univariate Decomposition

- Note that $\mathcal{K}$ is, in fact, a cellular decomposition for $\mathcal{F}$. Any semialgebraic set S defined by $\mathcal{F}$ is simply the union of a subset of elementary intervals in $\mathcal{K}$. Furthermore, for each interval $C \in \mathcal{K}$, we can compute a sample point α_C as follows:

$$\alpha_C = \begin{cases} \xi_1^{I} - 1, & \text{if } C = [-\infty, \xi_1); \\ \xi_i, & \text{if } C = [\xi_i, \xi_i]; \\ (\xi_i + \xi_{i+1})/2, & \text{if } C = (\xi_i, \xi_{i+1}); \\ \xi_s + 1, & \text{if } C = (\xi_s, +\infty]. \end{cases}$$

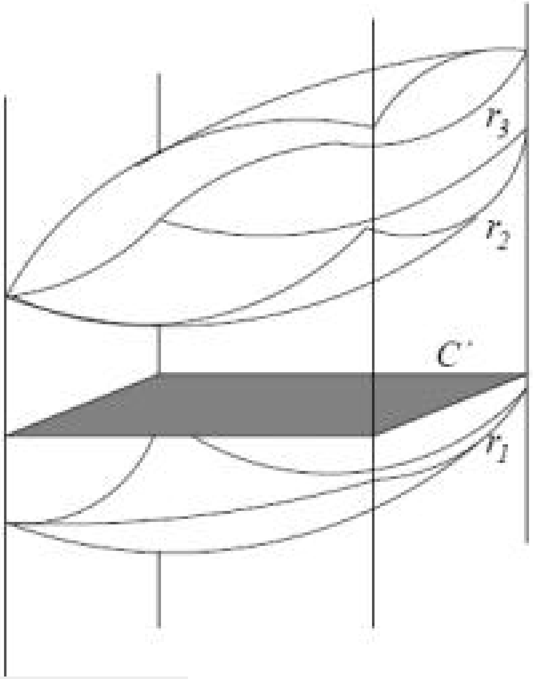
Multivariate Decomposition

- A generalization of the univariate decomposition to higher dimensions
- **Collins's cylindrical algebraic decomposition.**
- To represent a semialgebraic set $S \subseteq \mathbb{R}^n$, assume recursively that we can construct a cell decomposition of its projection $\pi(S) \subseteq \mathbb{R}^{n-1}$ (also a semialgebraic set); … then decompose S as a union of the sectors and sections in the cylinders above each cell of the projection, $\pi(S)$. This also leads to a cell decomposition of S.

Multivariate Decomposition

- One can further assign an algebraic sample point in each cell of S recursively in a straightforward manner.
- If $\mathcal{F}$ is a set of polynomials defining the semialgebraic set $S \subseteq \mathbb{R}^n$, then at no additional cost, we may in fact compute a cell decomposition for $\mathcal{F}$ using the procedure described above.
- Such a decomposition leads to a cylindrical algebraic decomposition for $\mathcal{F}$.

Cylindrical Algebraic Decomposition



Cylindrical Algebraic Decomposition (CAD)

- A recursively defined cell decomposition of $\mathbb{R}^n$ for $\mathcal{F}$. The decomposition is a cellular decomposition if the set of defining polynomials $\mathcal{F}$ satisfies certain nondegeneracy conditions.
- In the recursive definition, the cells of n-dimensional CAD are constructed from an (n−1)-dimensional CAD: Every (n−1)-dimensional CAD cell C' has the property that the distinct real roots of F over C' vary continuously as a function of the points of C'.□

CAD

- Moreover, the following quantities remain invariant over a $(n-1)$-dimensional cell:

1. the total number of complex roots of each polynomial of F;
2. the number of distinct complex roots of each polynomial of F; and
3. the total number of common complex roots of every distinct pair of polynomials of F.

- These conditions can be expressed by a set $\Phi(F)$ of at most $O(md)^2$ polynomials in $(n - 1)$ variables, obtained by considering principal subresultant coefficients (PSC's)..

CAD

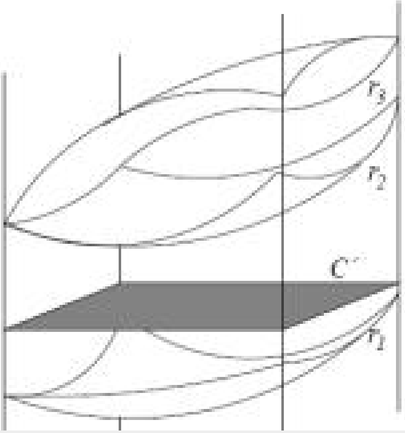
- Thus, the conditions encoded by $\Phi(\mathcal{F})$ correspond roughly to resultants and discriminants, and ensure that the polynomials of $\mathcal{F}$ do not intersect or “fold” in a cylinder over an $(n-1)$-dimensional cell
- The polynomials in $\Phi(F)$ are each of degree no more than d^2.
- More formally, an $\mathcal{F}$-sign-invariant cylindrical algebraic decomposition of $\mathbb{R}^n$ is:
- **Base Case: n = 1**. A univariate cellular decomposition of R1 as shown earlier

CAD

- **Inductive Case: n > 1.** Let K' be a $\Phi(\mathcal{F})$-sign-invariant CAD of $\mathbb{R}^{n-1}$. For each cell $C' \in \mathcal{K}'$, define an auxiliary polynomial $g_{C'}(x_1, \ldots, x_{n-1}, x_n)$ as the product of those polynomials of $\mathcal{F}$ that do not vanish over the $(n-1)$-dimensional cell, C'. The real roots of the auxiliary polynomial g'_C over C' give rise to a finite number (perhaps zero) of semialgebraic continuous functions, which partition the cylinder $C' \times (\mathbb{R} \cup \{\pm \infty)$ into finitely many $\mathcal{F}$-sign-invariant "slices." The auxiliary polynomials are of degree no larger than md.

CAD

- Assume that the polynomial $g_{C'}(p', x_n)$ has l distinct real roots for each $p' \in C'$: $r_1(p')$, $r_2(p')$, . . . , $r_l(p')$, each r_i being a continuous function of p′.
- The following sectors and sections are cylindrical over C′



$$\begin{aligned}
C_0^* &= \Big\{\langle p', x_n\rangle \mid p' \in C' \wedge x_n \in [-\infty, r_1(p'))\Big\},\\
C_1 &= \Big\{\langle p', x_n\rangle \mid p' \in C' \wedge x_n \in [r_1(p'), r_1(p')]\Big\},\\
C_1^* &= \Big\{\langle p', x_n\rangle \mid p' \in C' \wedge x_n \in (r_1(p'), r_2(p'))\Big\},\\
&\vdots\\
C_l^* &= \Big\{\langle p', x_n\rangle \mid p' \in C' \wedge x_n \in (r_l(p'), +\infty]\Big\}.
\end{aligned}$$

Sample Points

- Cylindrical algebraic decomposition (CAD) provides a sample point in every sign-invariant connected component for $\mathcal{F}$
- However, the total number of sample points generated is doubly-exponential, while the number of connected components of all sign conditions is only singly-exponential.
- In order to avoid this high complexity (both algebraic and combinatorial) of a CAD, many efficient techniques have been proposed recently.

Decision Process

- In the general case, the decision procedure follows a search process that proceeds only on the coordinates of the sample points in the CAD
- This follows because a sample point in a cell acts as a representative for any point in the cell as far as the sign conditions are concerned.
- Consider a Tarski sentence

$$(Q_1 x^{[1]})\ (Q_2 x^{[2]}) \cdots (Q_\omega\ x^{[\omega]})\ [\psi(x^{[1]}, \ldots, x^{[\omega]}],$$

with $\mathcal{F}$ the set of polynomials appearing in the matrix ψ. Let $\mathcal{K}$ be a cylindrical algebraic decomposition of $\mathbb{R}^n$ for $\mathcal{F}$.

Decision Process

- Since the cylindrical algebraic decomposition produces a sequence of decompositions:

$$\mathcal{K}_1 \text{ of } \mathbb{R}^1, \mathcal{K}_2 \text{ of } \mathbb{R}^2, \ldots, \mathcal{K}_n \text{ of } \mathbb{R}^n,$$

- such that the each cell $C_{i-1,j}$ of $\mathcal{K}_i$ is cylindrical over some cell C_{i-1} of $\mathcal{K}_{i-1}$, the search progresses by first finding cells C_1 of $\mathcal{K}_1$ such that

$$(Q_2 x_2) \cdots (Q_n\ x_n)\ [\psi(\alpha_{C_1}, x_2, \ldots, x_n)] = \textbf{True}.$$

- For each C_1, the search continues over cells C_{12} of $\mathcal{K}_2$ cylindrical over C_1 such that

$$(Q_3 x_3) \cdots (Q_n x_n)\ [\psi(\alpha_{C_1}, \alpha_{C_{12}}, x_3, \ldots, x_n)] = \textbf{True},$$

etc.

Decision Process

- Finally, at the bottom level the truth properties of the matrix ψ are determined by evaluating at all the coordinates of the sample points.
- This produces a tree structure, where each node at the (i−1)-th level corresponds to a cell $C_{i-1} \in \mathcal{K}_{i-1}$ and its children correspond to the cells $C_{i-1,j} \in \mathcal{K}_i$ that are cylindrical over C_{i-1}. The leaves of the tree correspond to the cells of the final decomposition $\mathcal{K} = \mathcal{K}_n$. Because we only have finitely many sample points, the universal quantifiers can be replaced by finitely many conjunctions and the existential quantifiers by disjunctions.

Decision Process

- Thus, we label every node at the (i−1)-th level "AND" (respectively, "OR") if Q_i is a universal quantifier ∀ (respectively, ∃) to produce a so-called AND-OR tree. The truth of the Tarski sentence is thus determined by simply evaluating this AND-OR tree.
- A quantifier elimination algorithm can be devised by a similar reasoning and a slight modification of the CAD algorithm described earlier.

Next Step

- Explore possible confluence of the theory of **hybrid automata** and the techniques of **algorithmic algebra** and **model checking** to create a computational basis for **systems biology**.
- **Simplest Scenario:**
- Devise a method to compute bounded reachability by combining Taylor polynomials and cylindric algebraic decomposition algorithms.
- What are the power and limitations of this framework .

Algorithmic Algebraic Model Checking

- Replacing numerical integration by a symbolic step:
- Generalizing Euler forward Numerical integration:

$$f(X,t+h) \sim f(X,t) + c_1.f'(X,t)\ h + \cdots + c_k.f''(X,t)\ h^k$$

- Expression in "X", "t" and "h"
- Error: integration discretization approximation
- Model Checking = iterative process of checking what is true now and at "next" time
- Possible over "semi-algebraic sets" using "quantifier elimination"

Symbolic Model Checking

- Take the following question: Is a semi-algebraic formula Φ an invariant of the system?
- Given Φ is true at t, is it true at t+h?

$$\forall_t\ \Phi(s(t)) \Rightarrow \Phi(s(t+h))?$$

The above statement can be expressed as a Tarski sentence…

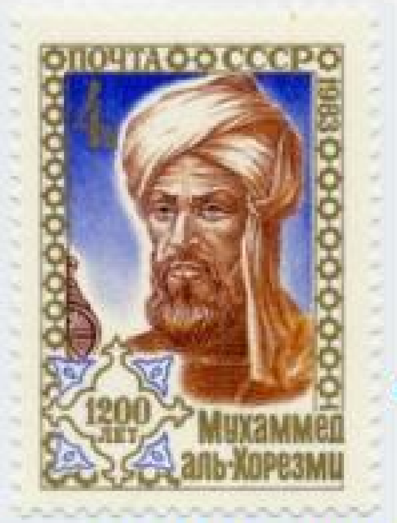
Topics in Semi-Algebraic Hybrid Systems

- Algorithmic Algebraic Model Checking
- Semi-Algebraic subclass & TCTL
- Undecidability in the “real” Turing Machine
- Approximate Methods: Extended Bisimulation Partitioning, Polytopes, Grids, Time Discretization

History

Algorithmic Algebra

- A mathematician in the court of Caliph Harun Al Rasid of Abassid Dynasty
- Two of his books:
  - Al-Kitab al-Mukhtasar fi-hisab al-Jabr al_Muqabalah (**Algebra**)
  - Kitab al-Jam'a wal-Tafreeq bil-Hisab al-Hindi (**Algorithm**)
  - Translated into Latin in the twelfth century, as Algoritmi de numero Indorum
  - Translated Aryabhatta's Siddhanta into Arabic (**SindHind**)
- Amalgamation of Indian & Greek mathematics



ūsā al-Khwārizmī (780-850 AD)

أبو عبد الله محمد

Some Milestones in the History of Algebra

- **820**: The word algebra is derived from operations described in the treatise of **al-Khwārizmī** titled **Al-Kitab al-Jabr wa-l-Muqabala**
- **Circa 850**: Persian mathematician **al-Mahani** conceived the idea of reducing geometrical problems such as duplicating the cube to problems in algebra.
- **Circa 850**: Indian mathematician **Mahavira** solves various quadratic, cubic, quartic, quintic and higher-order equations, as well as indeterminate quadratic, cubic and higher-order equations.

Some Milestones in the History of Algebra

- **Circa 990**: Persian **Abu Bakr al-Karaji**, in his treatise al-Fakhri, further develops algebra …He replaces geometrical operations of algebra with modern arithmetical operations, and defines the **monomials** x, x_2, x_3, ... and $1/x$, $1/x_2$, $1/x_3$, ... and gives rules for the products of any two of these.
- **Circa 1050**: Chinese mathematician **Jia Xian** finds numerical solutions of polynomial equations.
- **1072**: Persian mathematician **Omar Khayyam** develops algebraic geometry and, in the Treatise on Demonstration of Problems of Algebra, gives a complete classification of cubic equations

Some Milestones in the History of Algebra

- **1114**: Indian mathematician **Bhaskara**, in his Bijaganita (Algebra), solves various cubic, quartic and higher-order polynomial equations, as well as the general quadratic indeterminant equation.
- **1202**: Algebra is introduced to Europe largely through the work of **Leonardo Fibonacci of Pisa** in his work **Liber Abaci**.
- **Circa 1300**: Chinese mathematician **Zhu Shijie** deals with ***polynomial algebra***, solves simultaneous equations etc.
- **Circa 1400**: Indian mathematician **Madhava of Sangamagramma** finds **iterative methods** for approximate solution of non-linear equations.

Some Milestones in the History of Algebra

- **1545**: **Girolamo Cardano** publishes Ars magna -The great art which gives Fontana's solution to the general quartic equation.
- **1591**: **Francois Viete** develops improved symbolic notation *In artem analyticam isagoge*.
- **1682**: **Gottfried Wilhelm Leibniz** develops his notion of symbolic manipulation with formal rules which he calls **characteristica generalis**.

Some Milestones in the History of Algebra

- **1750**: **Gabriel Cramer**, in his treatise Introduction to the analysis of algebraic curves, states Cramer's rule and studies algebraic curves, matrices and determinants.
- **1824**: **Niels Henrik Abel** proved that the general quintic equation is insoluble by radicals.
- **1832**: Galois theory is developed by **Évariste Galois** in his work on abstract algebra.

Semialgebraic Geomtery

- **1950: Tarski's** work on a decision method for elementary algebra and geometry
  - Tarski's method is rather prohibitive, as its complexity cannot be bound by a tower of exponential functions, i.e. is not even elementary recursive.
  - This asymptotic complexity is also the one of the methods described by **Seidenberg** and **Cohen**.
- The first elementary recursive method was found by **Collins** using the technique of Cylindrical Algebraic Decomposition (CAD), whose complexity is doubly exponential.

Practicality

- For purely existentially or universally quantified problems methods of single exponential complexity was described first by **Renegar**.
- A practically working quantifier-elimination methods have been the so called "virtual substitution" methods. Based on ideas of Ferrante and Rackoff for decision problems, virtual substitution methods for quantifier elimination was created by **Weispfenning**.
- Implemented in **Redlog**

Quantifier Elimination (QE)

- Hong implemented Qepcad
- Other Tools: **Redlog**, **Maple**, **Mathematica**, AQCS
  - Input: $(\exists\, x)\ [\ x^2 + b\,x + c = 0\ ]$
  - Output: $[\ b^2 - 4\,c >= 0]$

..to be continued…

Symbolic Computation Algebraic Biology IV

Bud Mishra

Courant Inst, NYU

NYU SoM, TIFR, MSSM

Hybrid Systems

- Hybrid Models
- Algorithmic Algebraic Models & Model Checking
- O-minimal Systems & SaCoRe
- IDA
- Open Problems

The End