Computing Reachable States for Nonlinear Biological Models

Thao Dang   Colas Le Guernic   Oded Maler

CNRS - VERIMAG
Grenoble, France

2009

Summary

- We propose a computer-aided methodology to help analyzing certain biological models
- Domain of applicability: **biochemical reactions** modeled as **differential equations**. State variables denote **concentrations**
- We propose **reachability computation**, a kind of **set-based simulation**, that may replace uncountably-many simulations
- The continuous analogue of **algorithmic verification** (model-checking), emerged from more than a decade of research on **hybrid systems**
- Since this is not part of the local culture, we first introduce the domain and only later move to the contribution of this paper

Outline

- **Under-determined** dynamical models and their biological relevance
- **Continuous dynamical systems** and abstract reahcability
- **Effective representation** of sets and concrete algorithms for **linear** systems
- Treating **nonlinear** systems via **hybridization**
- **Dynamic hybridization**: idea and preliminary results
- Conclusions

Dynamical Models with Nondeterminism

- Dynamical system: state space X and a rule $x' = f(x, v)$
- The **next state** as a function of the **current state** and some **external influence** (or unknown parameters) $v \in V$
- In discrete domains: a transition system with input (alphabet)
- System becomes nondeterministic if input is projected away
- Given initial state, many possible evolutions ("runs")
- **Simulation**: picking **one** input and generating **one** behavior
- **Symbolic verification**: magically computing **all** runs in parallel
- **Reachability computation**: adapting these ideas to systems defined by **differential equations** or **hybrid automata** (differential equations with mode switching)

Why Bother?

- Differential models of biochemical reactions are **very** imprecise for many reasons:
- They are obtained by measuring populations, not individuals
- Kinetic parameters are based on isolated experiments not always under same conditions
- Etc.
- It is nice to match an experimentally-observed behavior by a deterministic model, but can we do better?
- After all, biological systems are supposed to be **robust** under **variations** in environmental conditions and parameters
- Showing that **all** trajectories corresponding to a **range** of parameters exhibit the same **qualitative behavior** is much stronger

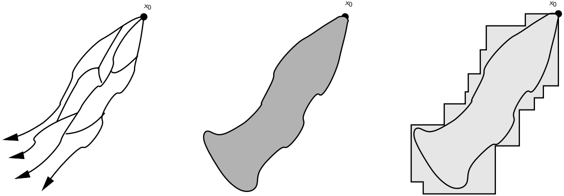
Preliminary Definitions and Notations

- A time domain $T = \mathbb{R}_+$, state space $X \subseteq \mathbb{R}^n$, input space $V \subseteq \mathbb{R}^m$
- **Trajectory**: partial function $\xi : T \rightarrow X$, **Input signal**: $\zeta : T \rightarrow V$ both defined over an interval $[0, t] \subset T$
- A continuous dynamical system $S = (X, V, f)$
- Trajectory ξ with endpoints x and x' is the **response** of S to input signal ζ if
- ξ is the solution of $\dot{x} = f(x, v)$ for initial condition x and $v(\cdot) = \zeta$, denoted by $x \xrightarrow{\zeta/\xi} x'$
- $R(x, \zeta, t) = \{x'\}$ denote the fact that x' is reachable from x by ζ within t time, that is, $x \xrightarrow{\zeta/\xi} x'$ and $|\zeta| = |\xi| = t$

Reachability

- $R(x, \zeta, t) = \{x'\}$ speaks of **one** initial state, **one** input signal and **one** time instant
- Generalizing to a **set** X_0 of initial states, to **all** time instants in an interval $I = [0, t]$ and **all** admissible input signals:

$$R_I(X_0) = \bigcup_{x \in X_0} \bigcup_{t \in I} \bigcup_{\zeta} R(x, \zeta, t)$$



- Depth-first vs. breadth-first

$$\bigcup_{\zeta} \bigcup_{t \in I} R(x, \zeta, t) = \bigcup_{t \in I} \bigcup_{\zeta} R(x, \zeta, t)$$

Abstract Reachability Algorithm

- The reachability operator satisfies the semigroup property:

$$R_{[0,t_1+t_2]}(X_0) = R_{[0,t_2]}(R_{[0,t_1]}(X_0))$$

- We can choose a time step r and apply the following iterative algorithm:

Input: A set $X_0 \subset X$
Output: $Q = R_{[0,L]}(X_0)$

$P := Q := X_0$
repeat $i = 1, 2 \ldots$
  $P := R_{[0,r]}(P)$
  $Q := Q \cup P$
until $i = L/r$

- Remark: we look at bounded time horizon and do not mind about reaching a fixpoint

From Abstract to Concrete Algorithms

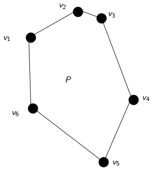
- The algorithm performs operations on **subsets** of $\mathbb{R}^n$ which, mathematically speaking, can be weird objects
- Like any **computational** geometry we restrict ourselves to classes of subsets (boxes, polytopes, ellipsoids, zonotopes) having nice properties:
- **Finite** syntactic representation
- Effective decision procedure for membership
- **Closure** (or approximate closure) under the reachability operator
- In this talk we use **convex polytopes** and their finite unions

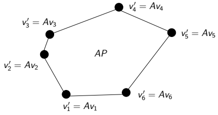
Convex Polytopes

- **Halfspace**: all points x satisfying a linear inequality $a \cdot x \leq b$
- **Convex polyhedron**: intersection of finitely many halfspaces; **Polytope**: bounded convex polyhedron
- **Convex combination** of a set of points $\{x_1, \ldots, x_l\}$ is any $x = \lambda_1 x_1 + \cdots + \lambda_l x_l$ such that $\sum_{i=1}^{l} \lambda_i = 1$
- The **convex hull** $conv(\tilde{P})$ of a set $\tilde{P}$ of points is the set of all convex combinations of elements in $\tilde{P}$
- Polytope representations:
  - **Vertices**: a polytope P admits a finite minimal set $\tilde{P}$ (vertices) such that $P = conv(\tilde{P})$.
  - **Inequalities**: a polytope P admits a canonical set of halfspaces/inequalities such that $P = \bigwedge_{i=1}^{k} a^i \cdot x \leq b^i$

Autonomous (Closed, Deterministic) Linear Systems

- Systems defined by linear differential equations of the form $\dot{x} = Ax$ where A is a matrix are the most well-studied
- There is a standard technique to fix a time step r and work in discrete time, a **recurrence equation** of the form $x_{i+1} = Ax_i$
- The image of a set P by the linear transformation A is $AP = \{Ax : x \in P\}$ (one-step successors)
- It is easy to compute, for example, for polytopes represented by vertices:
  $P = conv(\{x_1, \ldots, x_l\}) \;\Rightarrow\; AP = conv(\{Ax_1, \ldots, Ax_l\})$





Algorithm 1: Discrete-Time Linear Reachability

- **Input**: A set $X_0 \subset X$ represented as $conv(\tilde{P}_0)$
- **Output**: $Q = R_{[0..L]}(X_0)$ represented as a list $\{conv(\tilde{P}_0), \ldots, conv(\tilde{P}_L)\}$

$$
\begin{aligned}
&P := Q := \tilde{P}_0 \\
&\textbf{repeat } i = 1, 2 \ldots \\
&\quad P := AP \\
&\quad Q := Q \cup P \\
&\textbf{until } i = L
\end{aligned}
$$

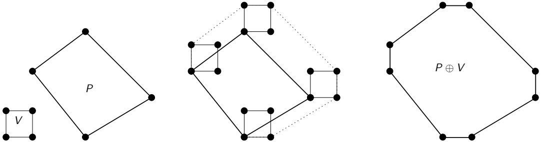
- Complexity assuming $|\tilde{P}_0| = m_0$ is $O(m_0 L M(n))$ where $M(n)$ is the complexity of matrix-vector multiplication in n dimensions: $\sim O(n^3)$
- Can be applied to other representations of objects closed under linear transformations

Linear Systems with Input

- Systems define by $x_{i+1} = Ax_i + v_i$ where the v_i's range over a bounded convex set V
- The one-step successor of P is defined as

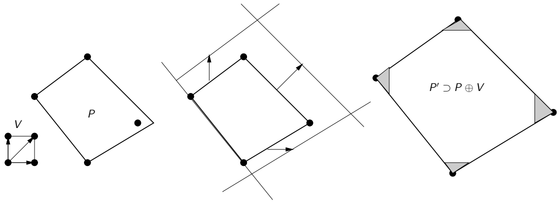
$$P' = \{Ax + v : x \in P, v \in V\} = AP \oplus V$$

- Minkowski sum $A \oplus B = \{a + b : a \in A \wedge b \in b\}$
- Same algorithm can be applied but the Minkowski sum increases the number of vertices in every step



Alternative: Pushing Facets

- Over-approximating the reachable set while keeping its complexity more or less fixed
- Assume P represented as intersection of halfspaces
- For each halfspace $H^i : a^i x \leq b^i$, let $v^i \in V$ be the input vector which pushes it in the "outermost" way
- Apply $Ax + Bv^i$ to H^i and the intersection of the pushed halfspaces over-approximates $AP \oplus V$



- The problem: over-approximation errors accumulate (the "wrapping effect")

Linear Reachability: State of the Art

- New algorithmics by C. Le Guernic and A. Girard
- Efficient computations: linear transformation applied to fixed number of points in each iteration
- No accumulation of over-approximation errors
- Initially used **zonotopes**, a class of sets closed under both linear operations and Minkowski sum; Can be applied to any "lazy" representation of the sequence of the computed sets
- Based on the observation that two consecutive sets

$$
\begin{aligned}
P_k &= A^k P_0 \oplus A^{k-1} V \oplus A^{k-2} V \oplus \ldots \oplus V \\
P_{k+1} &= A^{k+1} P_0 \oplus A^k V \oplus A^{k-1} V \oplus \ldots \oplus V
\end{aligned}
$$

  share a lot of terms
- Can compute within few minutes the reachable set after 1000 steps for linear systems with 200 (!) state variables

Linear Reachability: Some Credits

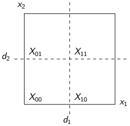
- Algorithmic analysis of hybrid systems started with tools like **Kronos** and **HyTech** for timed automata and "linear" hybrid automata: **HenzingerSifakisYovine** and **HenzingerHoWongtoi** - very simple continuous dynamics, summarized in **ACH^{+}95**
- Verifying differential equations: **Greenstreet96**
- Reachability for linear differential equations and hybrid systems: **ChutinanKrogh99**, **AsarinBournezDangMaler00** (polytopes) **KurzhanskiVaraiya00**, **BotchkarevTripakis00** (ellipsoids), **MitchellTomlin00** (level sets)
- Pushing faces and treating inputs: **DangMaler98**, **Varaiya98**
- Using zonotopes: **Girard05**
- New algorithmic scheme **Girard LeGuernic06-09**

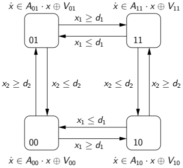
The Nonlinear Challenge

- Ok, bravo, but linear systems were studied to death by everybody. Real interesting models, biological included, are **nonlinear**
- What about systems of the form $x_{i+1} = f(x_i, u_i)$ or even $x_{i+1} = f(x_i)$ where f is an arbitrary continuous function, say a polynomial ?
- Convexity-preservation property of linear maps doesn't hold
- You can make small time steps, use a **local linear approximation** and bloat the obtained set to be safe
- This approach will either accumulate **large errors** or require expensive computation in **every step**

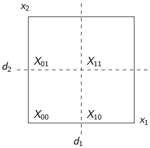
Hybridization: Asarin, Dang and Girard 2003

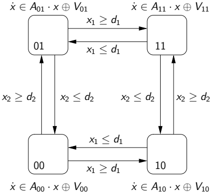
- Take a nonlinear system $x_{i+1} = f(x_i)$ and partition the state space into boxes (linearization domains)
- In each box X_q find a matrix A_q and a convex polytope V_q s.t. $f(x) \in A_q x \oplus V_q$ for every $x \in X_q$
- A_q is a **local linearization** of f with error bounded by V_q
- The new dynamics is $\quad x_{i+1} \in A_q x \oplus V_q \;$ iff $\; x \in X_q$
- A piecewise-(linear-with-input) systems, a restricted type of a hybrid automaton, which over-approximate f in terms of inclusion of trajectories





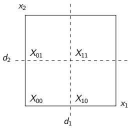
Hybridization (cont.)

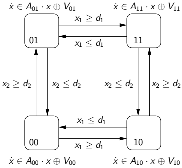




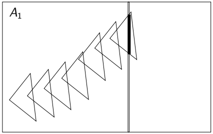
- In the hybrid automaton, x evolves according to the linear dynamics $A_q x \oplus V_q$ as long as it **remains** in X_q
- Reaching the **boundary** between X_q and $X_{q'}$, it takes a **transition** to q' and evolves according to $A_{q'} x \oplus V_{q'}$
- Linearization and error are computed only in the passage between blocks, **not** in every step
- Quality can be improved by making boxes smaller

Hybrid Reachability

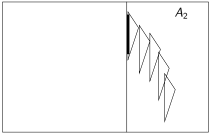




- Compute in one domain a sequences of sets using linear techniques until a set intersects with a boundary
- Take the intersection as initial set in next domain with the next linearization



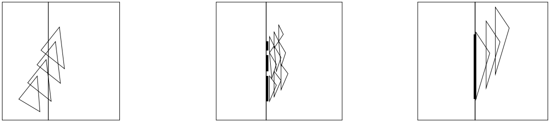
(a)



(b)

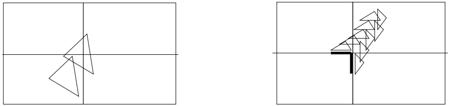
Between Theory and Practice

- First problem: intersection may be spread over many steps:



(a) (b) (c)

- Either **explosion** or union of intersections, **error** accumulation
- Major problem: a set may leave a box via **many facets**:

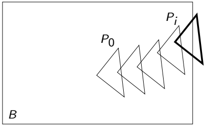


(a) (b)

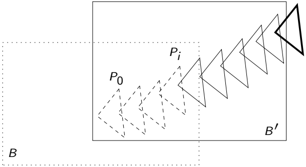
- Splitting is an **artifact** of the **fixed grid** imposed on the system
- Consequently, static hybridization is practically impossible beyond 3 dimensions

Our Contribution (at Last!)

- A **dynamic** hybridization scheme **not** based on a fixed grid
- In this scheme we do not need intersection **at all** and we allow the linearization domains to **overlap**
- When we leave a domain, we backtrack one step and define a new linearization domain around the previous set and continue with the new linearized dynamics from there



(a)

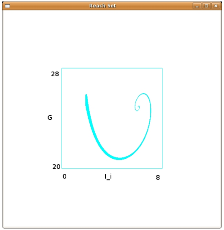


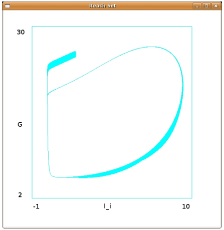
(b)

- And it works!

Example: E. Coli Lac Operon

$$
\begin{aligned}
\dot{R}_a &= \tau - \mu * R_a - k_2 R_a O_f + k_{-2}(\chi - O_f) - k_3 R_a I_i^2 + k_8 R_i G^2 \\
\dot{O}_f &= -k_2 r_a O_f + k_{-2}(\chi - O_f) \\
\dot{E} &= \nu k_4 O_f - k_7 E \\
\dot{M} &= \nu k_4 O_f - k_6 M \\
\dot{I}_i &= -2k_3 R_a I_i^2 + 2k_{-3} F_1 + k_5 I_r M - k_{-5} I_i M - k_9 I_i E \\
\dot{G} &= -2k_8 R_i G^2 + 2k_{-8} R_a + k_9 I_i E
\end{aligned}
$$





- We can also do a 9-dimensional highly-nonlinear **aging** model

Conclusions

- Disclaimer: we do **not** bring any new biological insight on any concrete system at this point
- Our goal is to develop **tools**, as **general-purpose** as possible, that can aid in the analysis of **many** non-trivial systems
- **Problem specificity** cannot be avoided of course: it will come up at the particular modeling and exploration phases
- Current version is a **prototype**:
  - Fixed-size boxes as linearizarization domains and other heuristics. Can be improved in efficiency and accuracy;
  - It is based on the old algorithmics for linear systems;
  - Improving all these aspects is on our immediate agenda
- We also explore alternative approaches for parameter synthesis based on simulation and sensitivity analysis **Donze et al09**
- Methodological aspects of the use of such tools in the **biological context** should be worked out

Thank You