Comparing Forward and Backward Reachability as Tools for Safety Analysis

Ian Mitchell
Department of Computer Science
The University of British Columbia

research supported by
National Science and Engineering Research Council of Canada

Outline

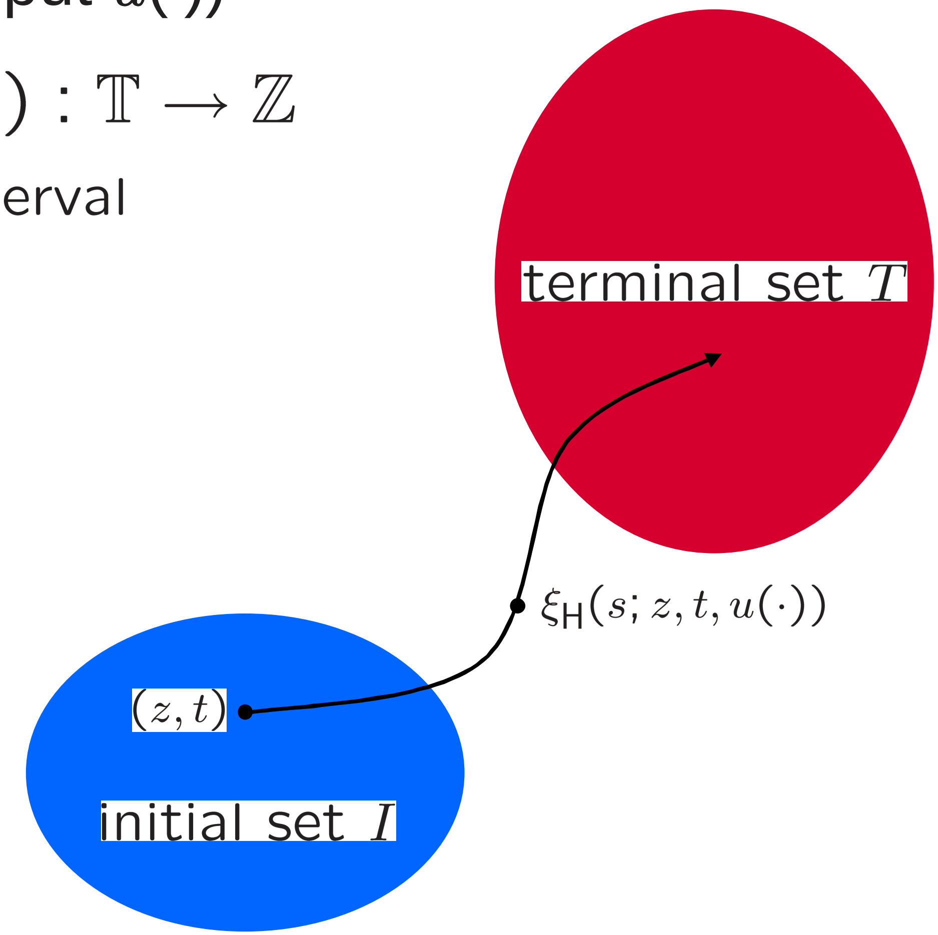
- Definitions
  - safety analysis and system models
  - forward and backward reach sets and tubes
- Exchanging algorithms by time reversal
- Safety analysis with different input policies
  - maximal reachability
  - minimal reachability
- Sensitivity of reachability operators
  - ill conditioned continuous & hybrid examples

Safety Analysis

- Does there exist a trajectory of system **H** leading from a state in initial set I to a state in terminal set T? (under some policy for input $u(\cdot)$)

Trajectory $\xi_{\mathsf{H}}(s; z, t, u(\cdot)) : \mathbb{T} \to \mathbb{Z}$

- $\mathbb{T} = [-\mathcal{T}, +\mathcal{T}]$ is time interval
- $\mathbb{Z}$ is state space of H
- $s \in \mathbb{T}$ is current time
- $z \in \mathbb{Z}$ is initial state
- $t \in \mathbb{T}$ is initial time
- $u(\cdot) \in \mathbb{U}$ is input signal



Assumption:

Given z, t and $u(\cdot)$ trajectory is unique

Typical Systems: ODEs

- Common model for continuous state spaces
- Standard existence and uniqueness

$$\dot{z}(t) = f(z(t), u(t))$$

- $f : \mathbb{Z} \times U \to \mathsf{T}\mathbb{Z}$ are dynamics
- $U \subset \mathbb{R}^{d_u}$ is convex and compact
- $\mathbb{U} = \{\phi : \mathbb{T} \to U \mid \phi(\cdot) \text{ is measurable}\}$
- Often $\mathbb{Z} \subseteq \mathbb{R}^{d_z}$

System specified by $\mathsf{H}_\mathsf{C} = (\mathbb{Z}, f, U)$

Typical Systems: Hybrid Automata

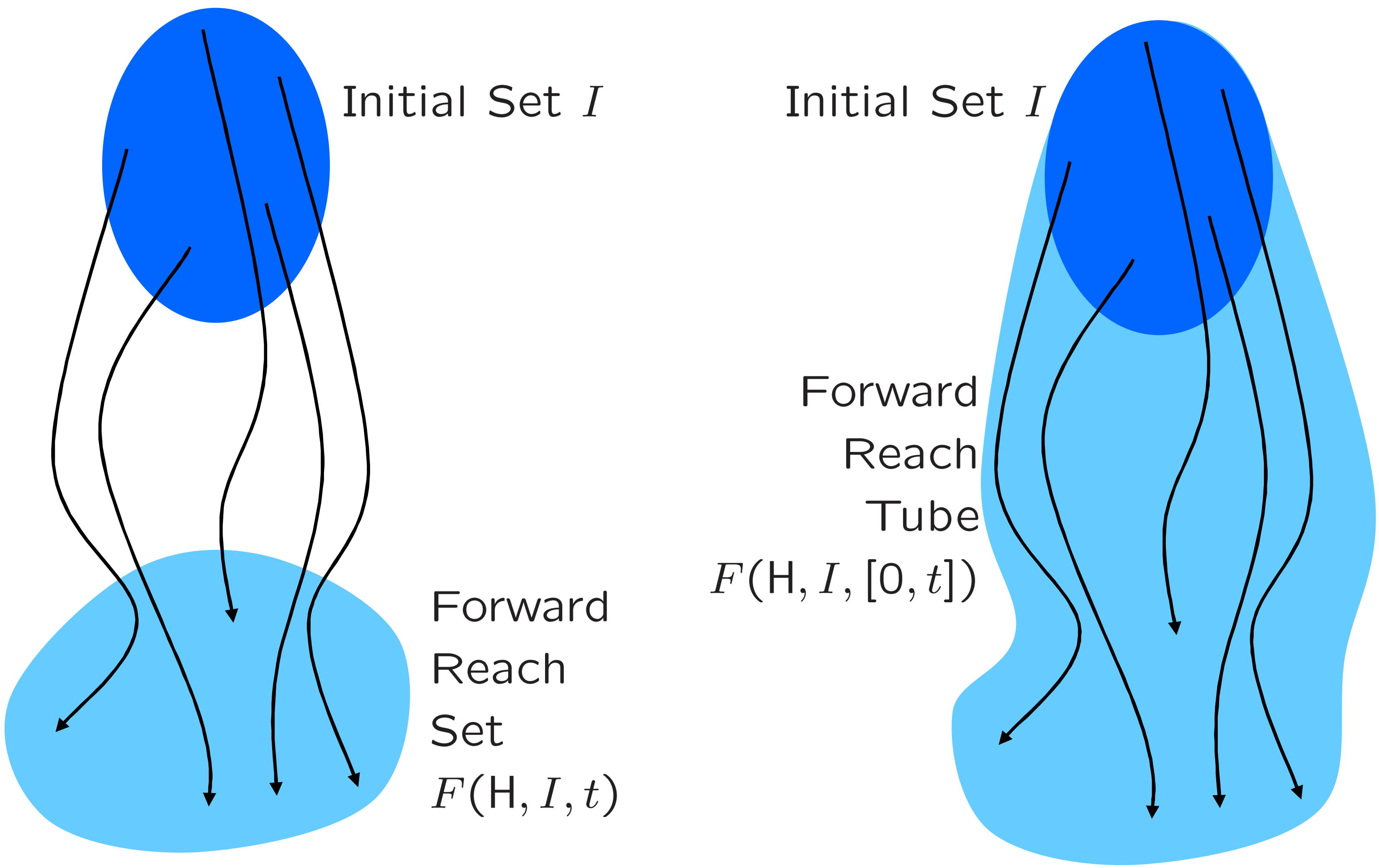
- Adapted from [Gao, Lygeros & Quincampoix 2006]
- Challenging existence and uniqueness
  - eg: [Broucke & Arapostathis, Sys. & Con. Letters 2002] or [Lygeros, Johansson, Simic, Zhang & Sastry, TAC 2003]
  - requires at least non-Zeno and non-blocking
  - all non-determinism must be expressed through input $u(\cdot)$

System specified by $\mathsf{H}_{\mathsf{H}} = (\mathbb{Q}, \mathbb{X}, f, D, G, r, U)$

$\mathbb{Q}$	discrete states;
$\mathbb{X}$	continuous states;
$f : \mathbb{Q} \times \mathbb{X} \times U_C \to \mathsf{T}\mathbb{X}$	continuous dynamics (vector field);
$D : \mathbb{Q} \times U_D \to P(\mathbb{X})$	domain of continuous evolution;
$G : \mathbb{Q} \times \mathbb{Q} \times U_D \to P(\mathbb{X})$	guards for discrete evolution;
$r : \mathbb{Q} \times \mathbb{Q} \times \mathbb{X} \times U \to \mathbb{X}$	reset function;
$U = (U_C, U_D)$	continuous and discrete input sets;

Forward Reachability

- Start at initial conditions and compute forward

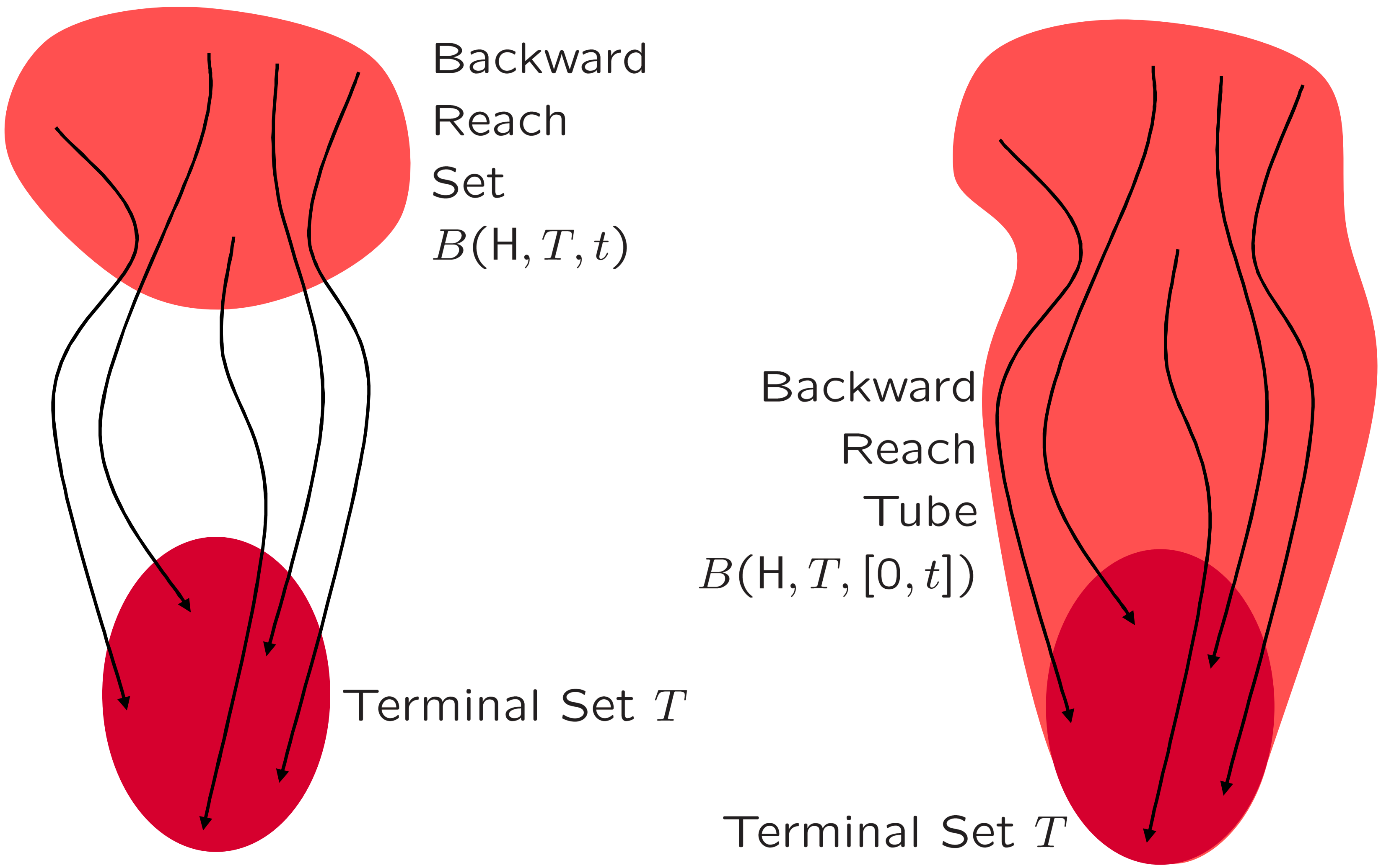


Forward Reachability Algorithms

- Forward approach typical of Lagrangian algorithms
  - Representation moves with the underlying dynamics
  - Varying ability to handle nonlinearity and/or inputs
  - Demonstrated ability to handle high dimensions
- Examples
  - [Henzinger, Ho & Wong-Toi, IEEE TAC 1998]
  - [Greenstreet & Mitchell, HSCC 1999]
  - [Bemporad, Torrisi & Morari HSCC 2000]
  - [Kurzhanski & Varaiya, HSCC 2000]
  - [Asarin, Dang & Girard, HSCC 2003]
  - [Girard, Guernic & Maler, HSCC 2006]
  - [Han & Krogh, HSCC 2006]

Backward Reachability

- Start at terminal set and compute backwards



Backward Reachability Algorithms

- Backward approach typical of Eulerian algorithms
  - Representation not moving (although it may adapt)
  - Generally handle nonlinear and multiple inputs
  - No examples beyond four dimensions?
- Examples
  - [Broucke, Benedetto, Gennaro & Sangiovanni-Vincentelli, HSCC 2001]
  - [Saint-Pierre, HSCC 2002]
  - [Sethian & Vladimirsky, HSCC 2002]
  - [Mitchell, Bayen & Tomlin, IEEE TAC 2005]
  - [Gao, Lygeros & Quincampoix, HSCC 2006]

Exchanging Algorithms

- Algorithms are (mathematically) interchangeable if system dynamics can be reversed in time

$$\text{Backward dynamic system } \overleftarrow{\mathsf{H}}$$
$$\text{such that } \forall s, t \in \mathbb{T}$$
$$\xi_{\mathsf{H}}(s; z, t, u(\cdot)) = \hat{z} \iff \xi_{\overleftarrow{\mathsf{H}}}(s; \hat{z}, t, u(\cdot)) = z.$$

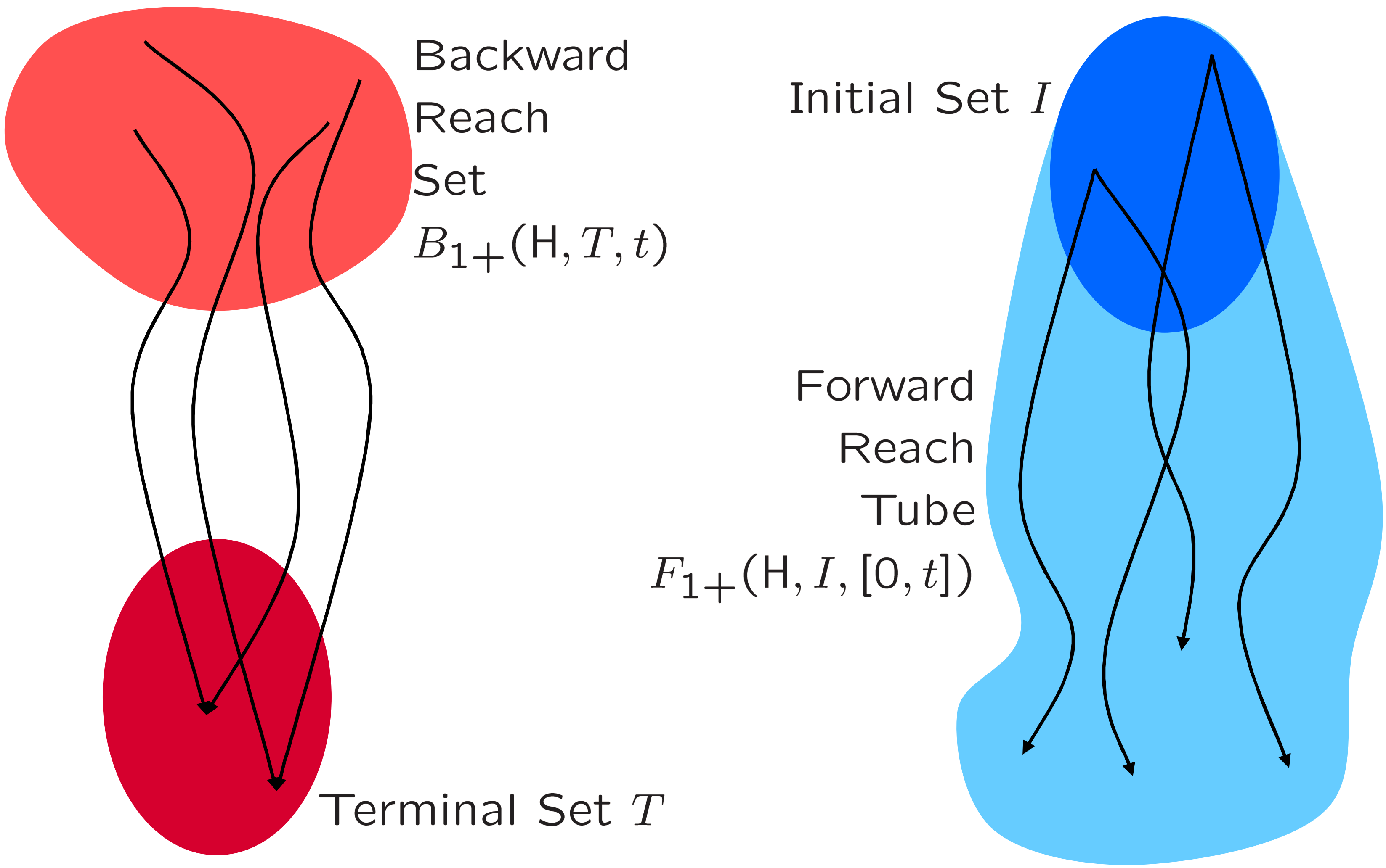
- For example: $\overleftarrow{\mathsf{H}_{\mathsf{C}}} = (\mathbb{Z}, -f, U)$
  $$\overleftarrow{\mathsf{H}_{\mathsf{H}}} = (\mathbb{Q}, \mathbb{X}, -f, D, \overleftarrow{G}, \overleftarrow{r}, U)$$

- Then

$$F(\mathsf{H}, S, [0, t]) = B(\overleftarrow{\mathsf{H}}, S, [0, t])$$
$$F(\mathsf{H}, S, t) = B(\overleftarrow{\mathsf{H}}, S, t)$$

Maximal Reachability

- Input signal $u(\cdot)$ maximizes size of the set or tube



Maximal Reachability Definition

$$F_{1+}(\mathsf{H}, S, t) \triangleq \{\hat{z} \in \mathbb{Z} \mid \exists u(\cdot) \in \mathbb{U}, \exists z \in S,$$
$$\xi_{\mathsf{H}}(t; z, 0, u(\cdot)) = \hat{z}\}$$
$$F_{1+}(\mathsf{H}, S, [0, t]) \triangleq \{\hat{z} \in \mathbb{Z} \mid \exists u(\cdot) \in \mathbb{U}, \exists z \in S, \exists s \in [0, t],$$
$$\xi_{\mathsf{H}}(s; z, 0, u(\cdot)) = \hat{z}\}$$
$$B_{1+}(\mathsf{H}, S, t) \triangleq \{z \in \mathbb{Z} \mid \exists u(\cdot) \in \mathbb{U}, \exists \hat{z} \in S,$$
$$\xi_{\mathsf{H}}(0; z, -t, u(\cdot)) = \hat{z}\}$$
$$B_{1+}(\mathsf{H}, S, [0, t]) \triangleq \{z \in \mathbb{Z} \mid \exists u(\cdot) \in \mathbb{U}, \exists \hat{z} \in S, \exists s \in [0, t],$$
$$\xi_{\mathsf{H}}(0; z, -s, u(\cdot)) = \hat{z}\}$$

Maximal results also apply to systems without input; for example: $F_0(\mathsf{H}, S, t) \triangleq \{\hat{z} \in \mathbb{Z} \mid \exists z \in S, \xi_{\mathsf{H}}(t; z, 0) = \hat{z}\}$.

Maximal Reachability Results

- Reach sets and tubes provide similar information

$$F_{1+}(\mathsf{H}, S, [0,t]) = \bigcup_{\hat{t} \in [0,t]} F_{1+}(\mathsf{H}, S, \hat{t})$$

$$B_{1+}(\mathsf{H}, S, [0,t]) = \bigcup_{\hat{t} \in [0,t]} B_{1+}(\mathsf{H}, S, \hat{t})$$

- The following properties are equivalent

1. H is safe over horizon $t \leq \mathcal{T}$ for all possible inputs $u(\cdot) \in \mathbb{U}$.
2. $F_{1+}(\mathsf{H}, I, s) \cap T = \emptyset$ for all $s \in [0, t]$.
3. $F_{1+}(\mathsf{H}, I, [0, t]) \cap T = \emptyset$.
4. $B_{1+}(\mathsf{H}, T, s) \cap I = \emptyset$ for all $s \in [0, t]$.
5. $B_{1+}(\mathsf{H}, T, [0, t]) \cap I = \emptyset$.

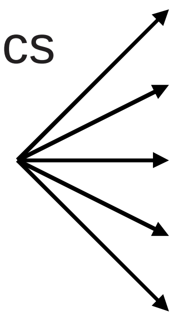
- Any maximal reachability operator can be used to demonstrate safety for all possible inputs

Maximal Reachability Demonstration

System Dynamics

$$\dot{x} = \begin{bmatrix} +1 \\ u \end{bmatrix}$$

$$|u| \leq 1$$



Initial and Terminal Sets

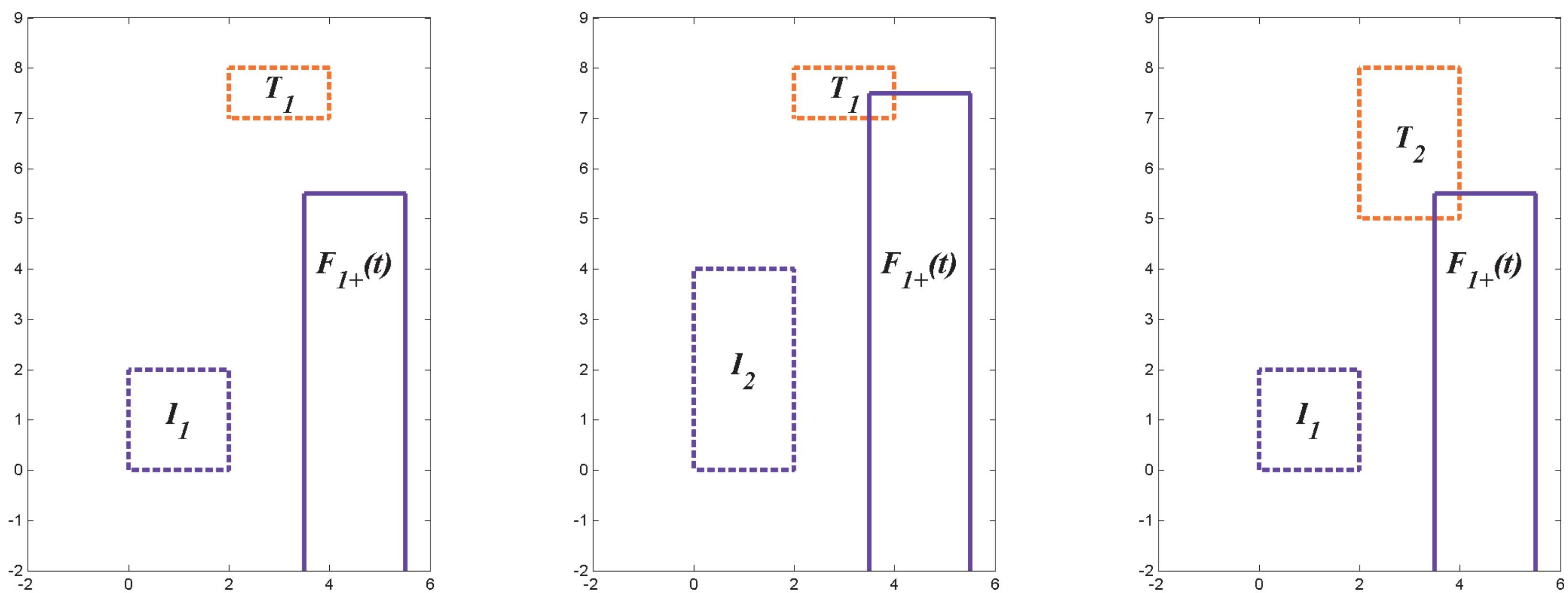
$$I_1 = [0, +2] \times [0, +2]$$

$$I_2 = [0, +2] \times [0, +4]$$

$$T_1 = [+2, +4] \times [+7, +8]$$

$$T_2 = [+2, +4] \times [+5, +8]$$

Forward Reach Set Results

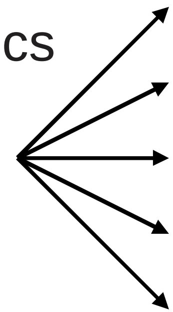


Maximal Reachability Demonstration

System Dynamics

$$\dot{x} = \begin{bmatrix} +1 \\ u \end{bmatrix}$$

$$|u| \leq 1$$



Initial and Terminal Sets

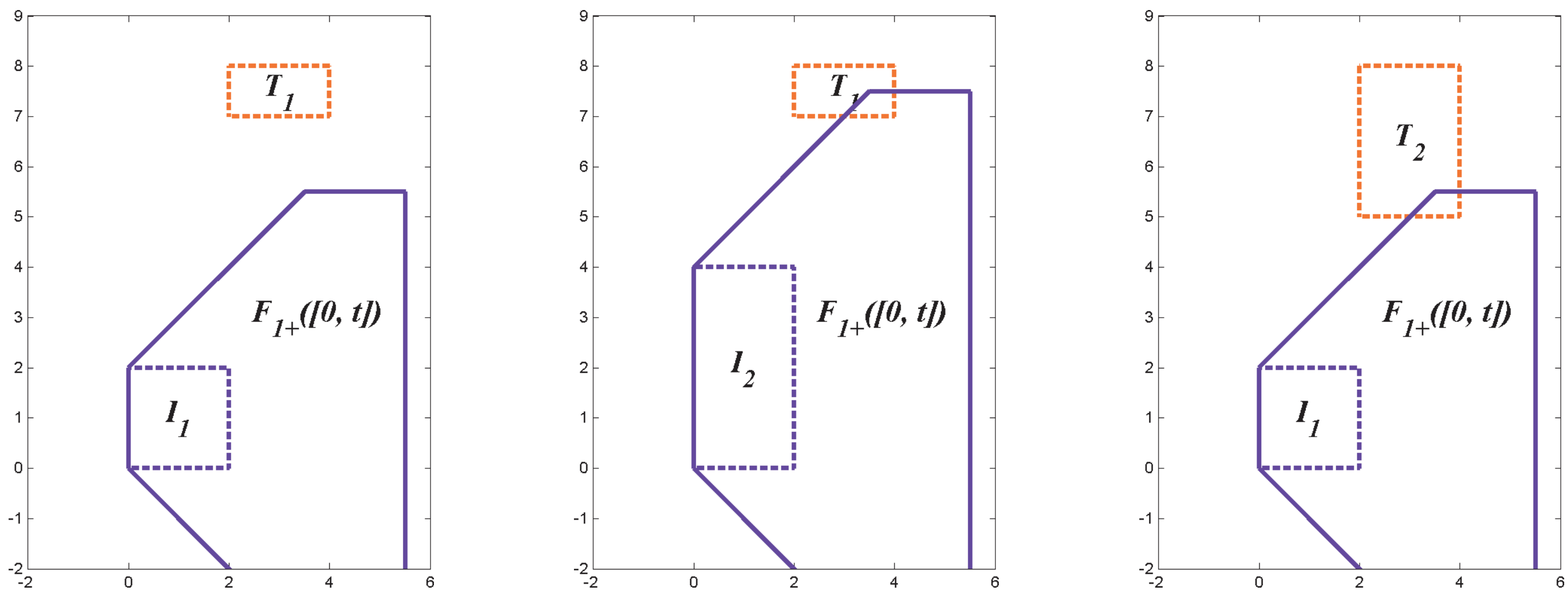
$$I_1 = [0, +2] \times [0, +2]$$

$$I_2 = [0, +2] \times [0, +4]$$

$$T_1 = [+2, +4] \times [+7, +8]$$

$$T_2 = [+2, +4] \times [+5, +8]$$

Forward Reach Tube Results

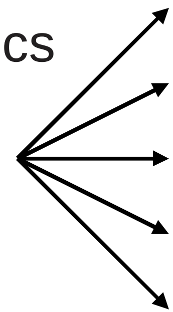


Maximal Reachability Demonstration

System Dynamics

$$\dot{x} = \begin{bmatrix} +1 \\ u \end{bmatrix}$$

$$|u| \leq 1$$



Initial and Terminal Sets

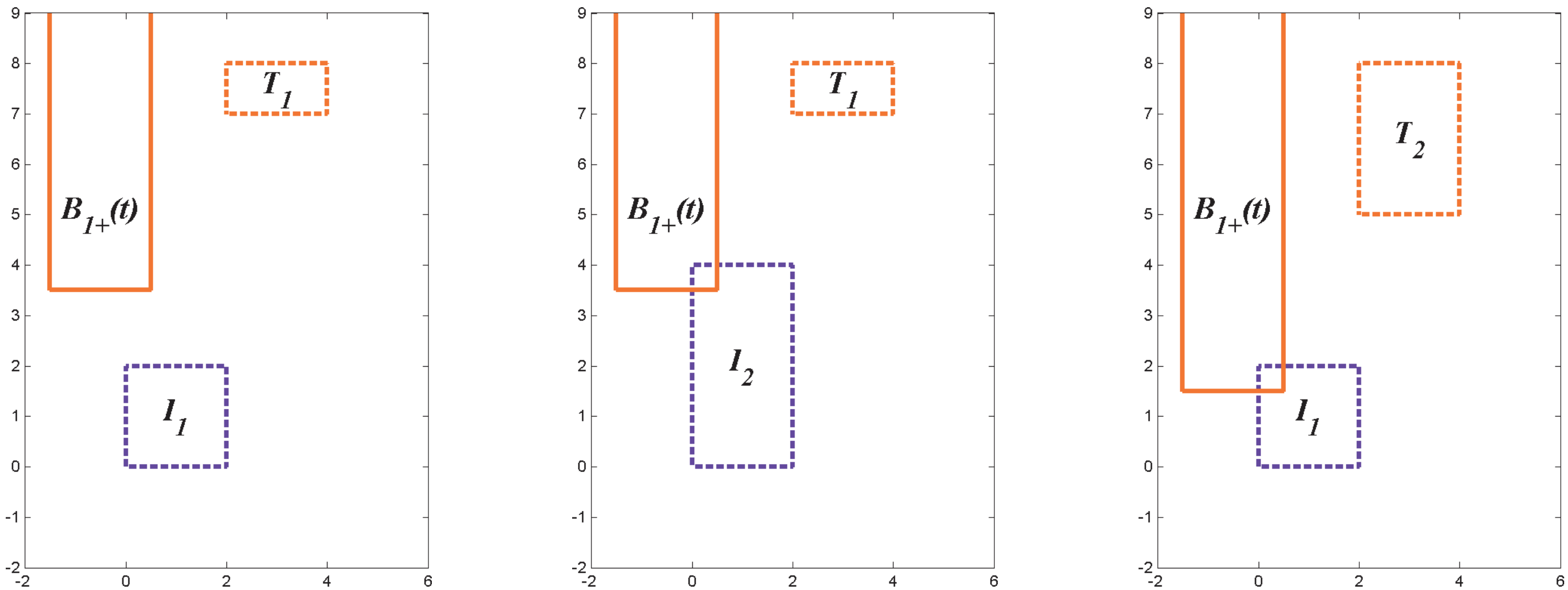
$$I_1 = [0, +2] \times [0, +2]$$

$$I_2 = [0, +2] \times [0, +4]$$

$$T_1 = [+2, +4] \times [+7, +8]$$

$$T_2 = [+2, +4] \times [+5, +8]$$

Backward Reach Set Results

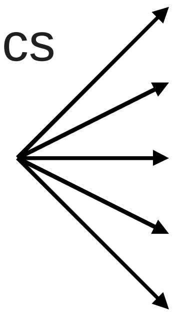


Maximal Reachability Demonstration

System Dynamics

$$\dot{x} = \begin{bmatrix} +1 \\ u \end{bmatrix}$$

$$|u| \leq 1$$



Initial and Terminal Sets

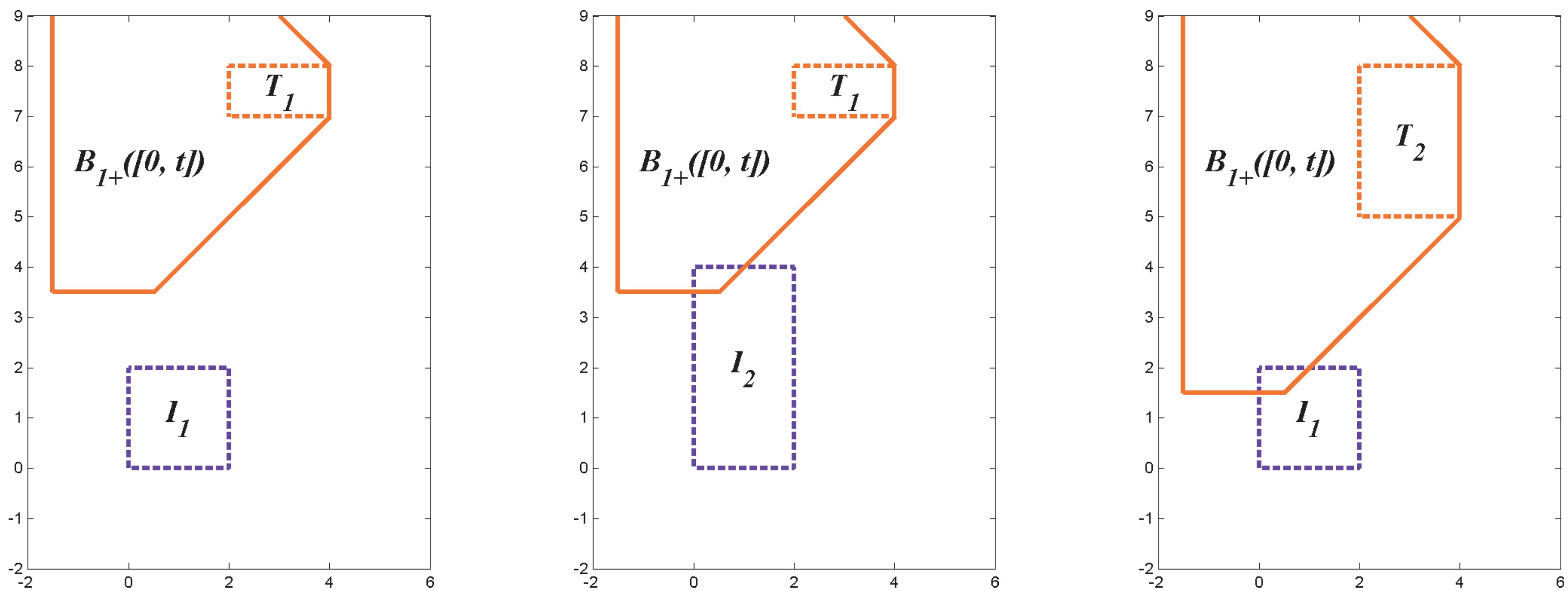
$$I_1 = [0, +2] \times [0, +2]$$

$$I_2 = [0, +2] \times [0, +4]$$

$$T_1 = [+2, +4] \times [+7, +8]$$

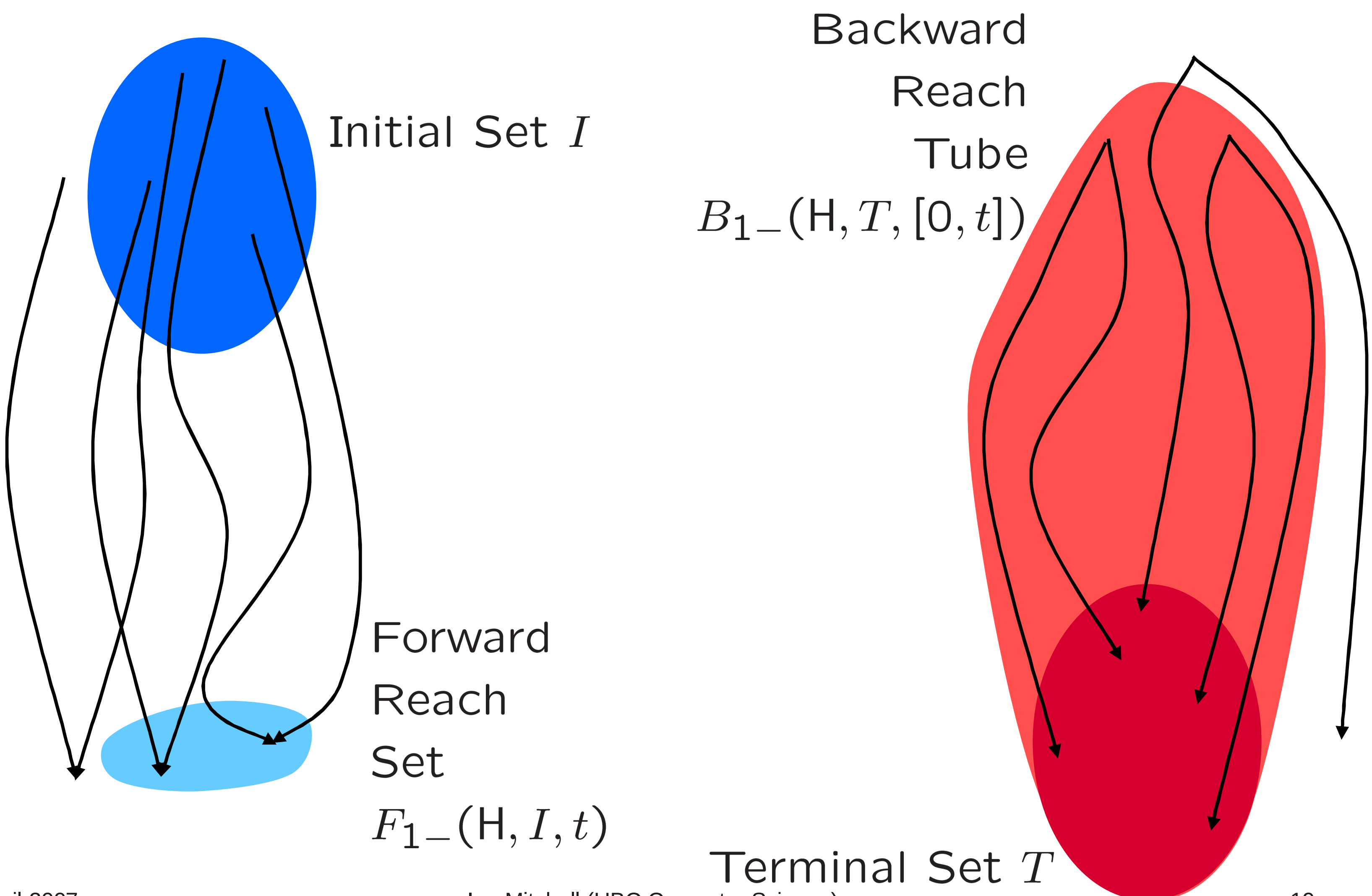
$$T_2 = [+2, +4] \times [+5, +8]$$

Backward Reach Tube Results



Minimal Reachability

- Input signal $u(\cdot)$ minimizes size of the set or tube



Minimal Reachability Definition

$$F_{1-}(\mathsf{H}, S, t) \triangleq \{\hat{z} \in \mathbb{Z} \mid \forall u(\cdot) \in \mathbb{U}, \exists z \in S, \\ \xi_{\mathsf{H}}(t; z, 0, u(\cdot)) = \hat{z}\}$$

$$F_{1-}(\mathsf{H}, S, [0, t]) \triangleq \{\hat{z} \in \mathbb{Z} \mid \forall u(\cdot) \in \mathbb{U}, \exists z \in S, \exists s \in [0, t], \\ \xi_{\mathsf{H}}(s; z, 0, u(\cdot)) = \hat{z}\}$$

$$B_{1-}(\mathsf{H}, S, t) \triangleq \{z \in \mathbb{Z} \mid \forall u(\cdot) \in \mathbb{U}, \exists \hat{z} \in S, \\ \xi_{\mathsf{H}}(0; z, -t, u(\cdot)) = \hat{z}\}$$

$$B_{1-}(\mathsf{H}, S, [0, t]) \triangleq \{z \in \mathbb{Z} \mid \forall u(\cdot) \in \mathbb{U}, \exists \hat{z} \in S, \exists s \in [0, t], \\ \xi_{\mathsf{H}}(0; z, -s, u(\cdot)) = \hat{z}\}$$

Minimal results also apply to systems with adversarial inputs; for example:

$$B_{2}(\mathsf{H}, S, [0, t]) \triangleq \{z \in \mathbb{Z} \mid \exists v(\cdot) \in \mathbb{V}, \forall u(\cdot) \in \mathbb{U}, \\ \exists \hat{z} \in S, \exists s \in [0, t], \\ \xi_{\mathsf{H}}(0; z, -s, u(\cdot), v(\cdot)) = \hat{z}\}.$$

Minimal Reachability Results

- Reach tubes provide more information

$$\bigcup_{\hat{t} \in [0,t]} F_{1-}(\mathsf{H}, S, \hat{t}) \subseteq F_{1-}(\mathsf{H}, S, [0,t])$$

$$\bigcup_{\hat{t} \in [0,t]} B_{1-}(\mathsf{H}, S, \hat{t}) \subseteq B_{1-}(\mathsf{H}, S, [0,t])$$

  - Choice of trajectory length t is quantified first for sets but last for tubes

Minimal Reachability Results

- Backward reach tubes are the only minimal reachability operator that can prove that there exists an input $u(\cdot)$ which keeps the system safe

$$B_{1-}(\mathsf{H}, T, [0,t]) \cap I = \emptyset \iff \exists u(\cdot) \in \mathbb{U} \text{ that keeps H safe}$$

$$B_{1-}(\mathsf{H}, T, s) \cap I = \emptyset, \ \forall s \leq t \Longleftarrow \exists u(\cdot) \in \mathbb{U} \text{ that keeps H safe}$$

$$\left.\begin{array}{r} F_{1-}(\mathsf{H}, I, t) \\ F_{1-}(\mathsf{H}, I, [0,t]) \end{array}\right\} \text{ provide no relevant information}$$

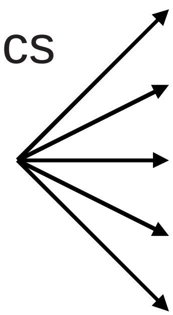
  – Basic problem with minimal forward reachability: the state lying in the terminal set is chosen before the input, while the state lying in the initial set is chosen after

Minimal Reachability Demonstration

System Dynamics

$$\dot{x} = \begin{bmatrix} +1 \\ u \end{bmatrix}$$

$$|u| \leq 1$$



Initial and Terminal Sets

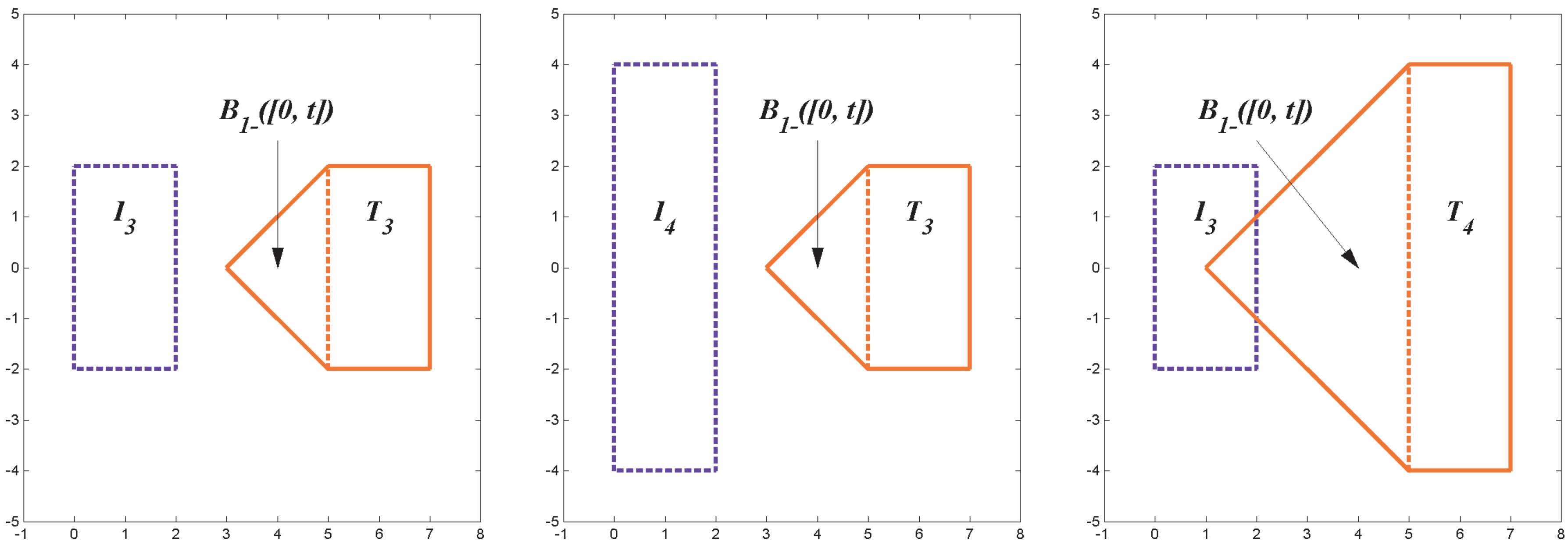
$$I_3 = [0, +2] \times [-2, +2]$$

$$I_4 = [0, +2] \times [-4, +4]$$

$$T_3 = [+5, +7] \times [-2, +2]$$

$$T_4 = [+5, +7] \times [-4, +4]$$

(Correct) Backward Reach Tube Results

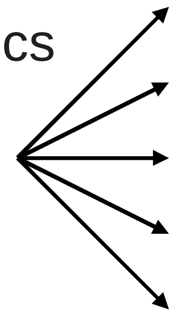


Minimal Reachability Demonstration

System Dynamics

$$\dot{x} = \begin{bmatrix} +1 \\ u \end{bmatrix}$$

$$|u| \leq 1$$



Initial and Terminal Sets

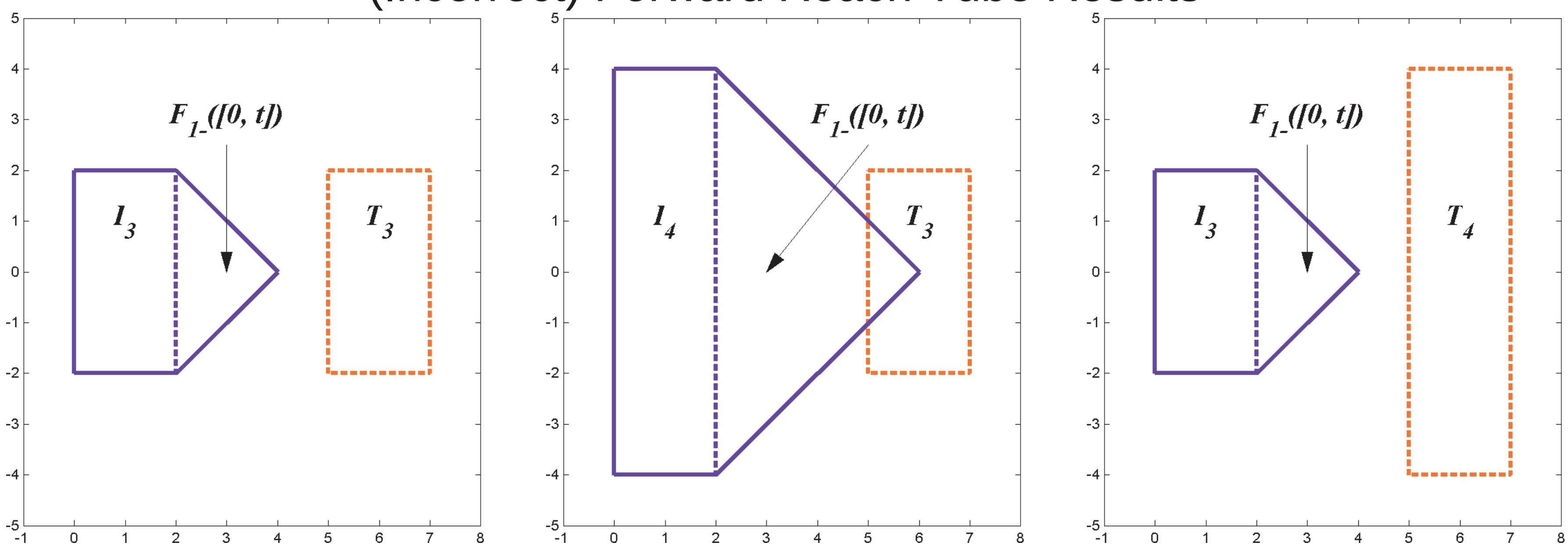
$$I_3 = [0, +2] \times [-2, +2]$$

$$I_4 = [0, +2] \times [-4, +4]$$

$$T_3 = [+5, +7] \times [-2, +2]$$

$$T_4 = [+5, +7] \times [-4, +4]$$

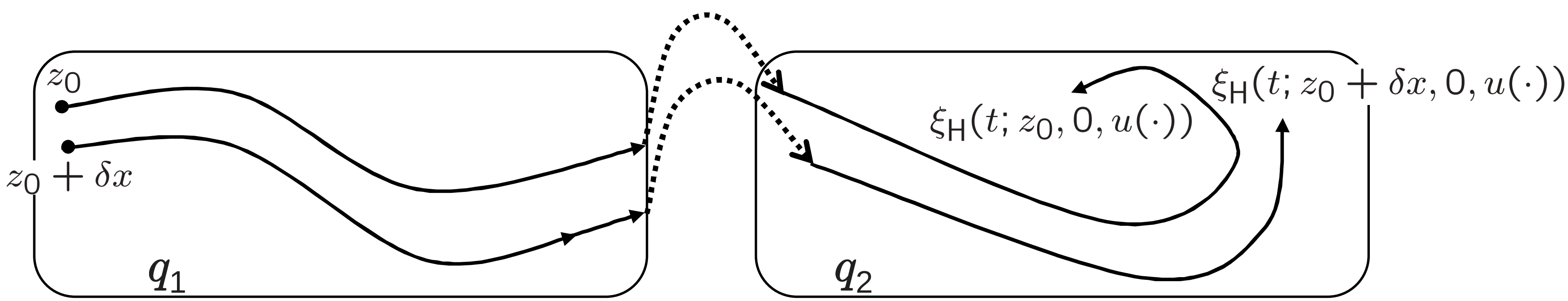
(Incorrect) Forward Reach Tube Results



Trajectory Sensitivity

- To approximate reach sets and tubes, direct algorithms integrate trajectories
- Small(?) perturbations occur in representation
  - Floating point roundoff
  - Simplified dynamics
  - Approximating the true set with a larger set from the appropriate class
- How might the interaction of perturbations and dynamics affect the quality of the approximation?

Compare $\xi_{\mathsf{H}}(t; z_0 + \delta x, 0, u(\cdot))$ to $\xi_{\mathsf{H}}(t; z_0, 0, u(\cdot))$



Sensitivity Analysis

- Focus on effect of continuous perturbation

$$\xi_H(t; z_0 + \delta x, 0, u(\cdot)) = \xi_H(t; z_0, 0, u(\cdot)) + \Xi_H(t; \xi_H(\cdot))\delta x + \mathcal{O}(\delta x^2)$$

- Sensitivity matrix

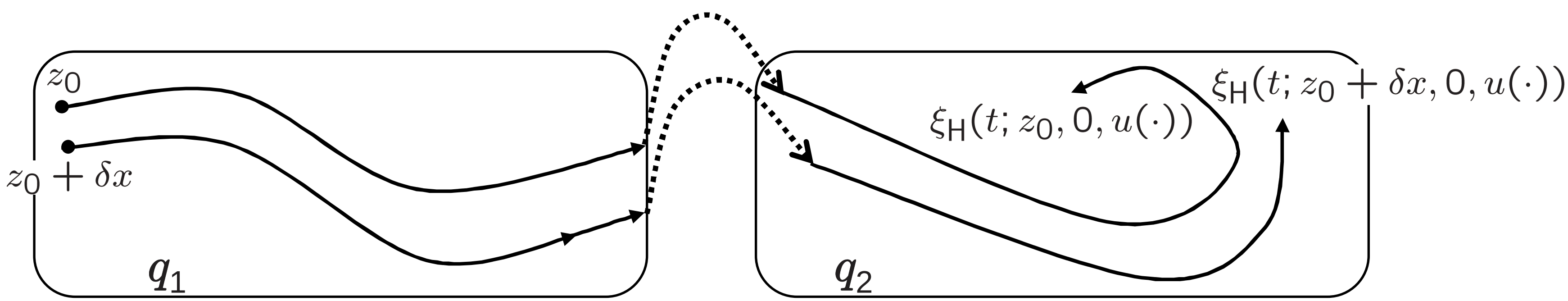
$$\Xi_H(t; \xi_H(\cdot)) \triangleq \frac{\partial \xi_H(t; z_0, 0, u(\cdot))}{\partial x_0}.$$

- Sensitivity of system dynamics

$$\mathbf{F}(q, x, u) \triangleq \mathbf{D}_x f(q, x, u) \qquad \mathbf{R}(q, \hat{q}, x, u) \triangleq \mathbf{D}_x r(q, \hat{q}, x, u)$$

- Continuous evolution of sensitivity matrix

$$\frac{d}{dt}\Xi_H(t) = \mathbf{F}(q, x, u)\Xi_H(t) \qquad \Xi_H(0) = \mathbf{I}$$



Sensitivity Analysis

- Discrete evolution of sensitivity matrix
  - Switching surfaces (guards, domains) specified implicitly

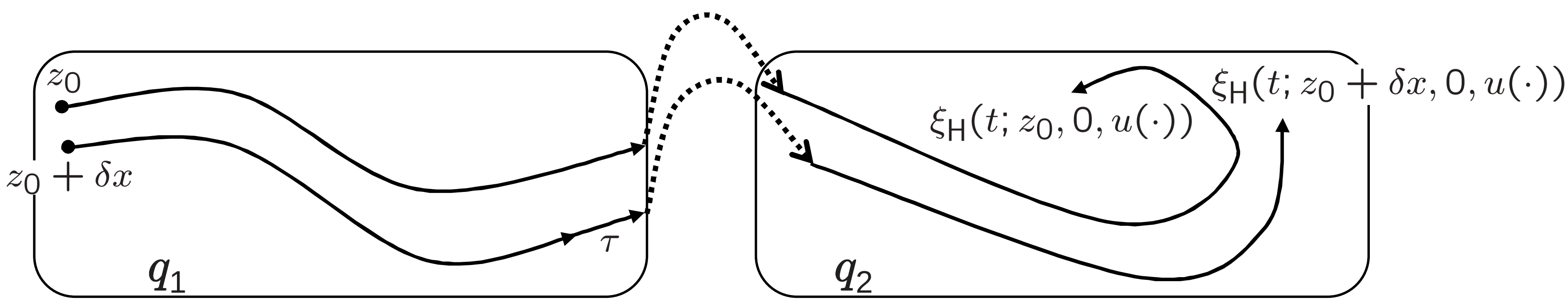
$$D(q, u_D) = \{x \in \mathbb{X} \mid \psi_D(q, x, u_D) \leq 0\}$$
$$G(q, \hat{q}, u_D) = \{x \in \mathbb{X} \mid \psi_G(q, \hat{q}, x, u_D) \leq 0\}$$

  - Difference in switching time [Hiskins & Pai, IEEE TC&S 2000]

$$\tau = \frac{\partial t(z_0)}{\partial z_0} = -\frac{\nabla \psi(x^-)^T \, \Xi_\mathsf{H}(t^-)}{\nabla \psi(x^-)^T f(q^-, x^-, u)}$$

  - Jump in sensitivity

$$\Xi_\mathsf{H}(t^+) = \mathbf{R}(q^-, q^+, x^-, u) \left(\Xi_\mathsf{H}(t^-) + f(q^-, x^-, u)\tau\right) - f(q^+, x^+, u)\tau,$$



Sensitivity of Forward Reachability

$$\|\xi_{\mathsf{H}}(s; z+\delta x, t, u(\cdot)) - \xi_{\mathsf{H}}(s; z, t, u(\cdot))\| \leq \|\Xi_{\mathsf{H}}(s; \xi_{\mathsf{H}}(\cdot))\| \|\delta x\|.$$

- Sensitivity matrix can become large via

$$\begin{aligned} \text{Real}[\lambda(\mathbf{F})] \gg 0 &\quad \text{continuous evolution} \\ |\lambda(\mathbf{R})| \gg 1 &\quad \text{discrete jumps} \\ \nabla\psi^T f^- \approx 0 &\quad \text{grazing contact with switching surface} \end{aligned}$$

- Systems satisfying these properties are inherently unpredictable
  - Deterministic models are rarely used for such systems

Sensitivity of Backward Reachability

- System dynamics are reversed

$$\overleftarrow{f} = -f \implies \overleftarrow{\mathbf{F}} = -\mathbf{F} \implies \lambda(\overleftarrow{\mathbf{F}}) = -\lambda(\mathbf{F})$$
$$\mathbf{R}\overleftarrow{\mathbf{R}} = \mathbf{I} \implies \overleftarrow{\mathbf{R}} = \mathbf{R}^{-1} \implies \lambda(\overleftarrow{\mathbf{R}}) = \lambda(\mathbf{R})^{-1}$$

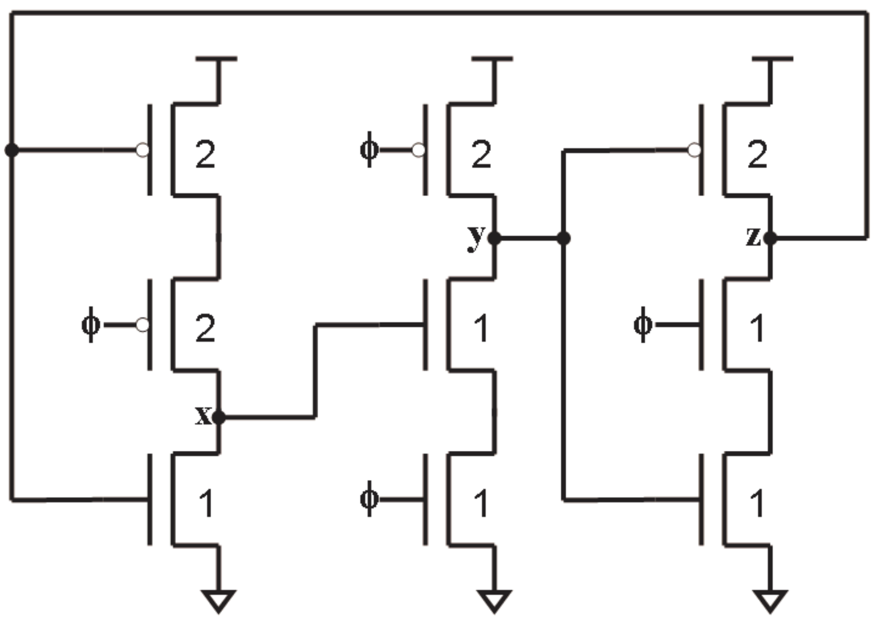
- Sensitivity matrix can become large via

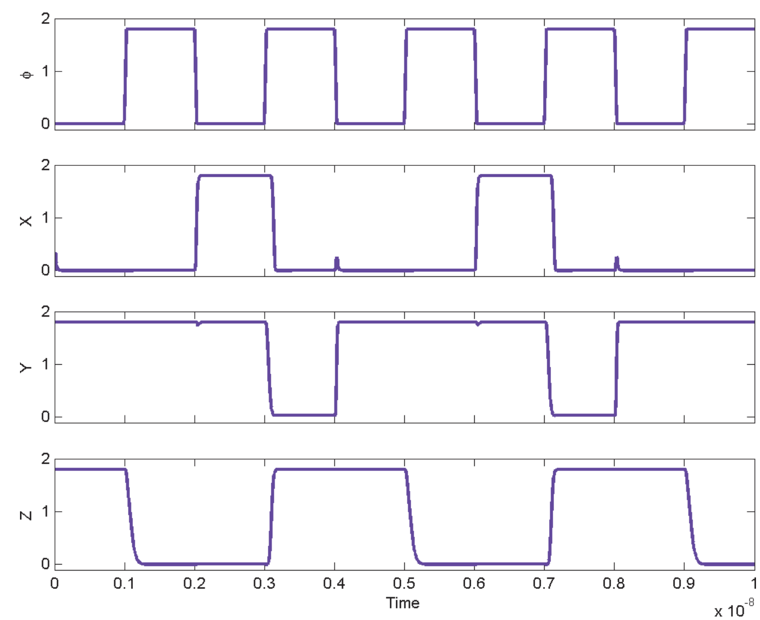
$$\begin{aligned} \text{Real}[\lambda(\mathbf{F})] \ll 0 &\quad \text{backward continuous evolution} \\ |\lambda(\mathbf{R})| \ll 1 &\quad \text{backward discrete jumps} \\ \nabla\psi^T f^+ \approx 0 &\quad \text{backward grazing contact} \\ &\quad \text{with switching surface} \end{aligned}$$

- Systems which show contraction are likely to be ill-conditioned for backward reachability
  - Such systems are commonly encountered, because their models are well-conditioned in forward time

Continuous System Sensitivity Example

- Toggle circuit [Yuan & Svensson, IEEE JSSC 1998]
  - Period of output **z** is double period of input ϕ
  - Short channel transistor model with velocity saturation, all capacitance to ground and interconnect capacitance is ignored [Hodges, Jackson & Saleh, 3rd edition 2004]
  - Forward verification that chain of toggles can operate as a counter [Greenstreet, CAV 1996]
  - Thanks: Mark Greenstreet, Chao Yan & Suwen Yang for simulation

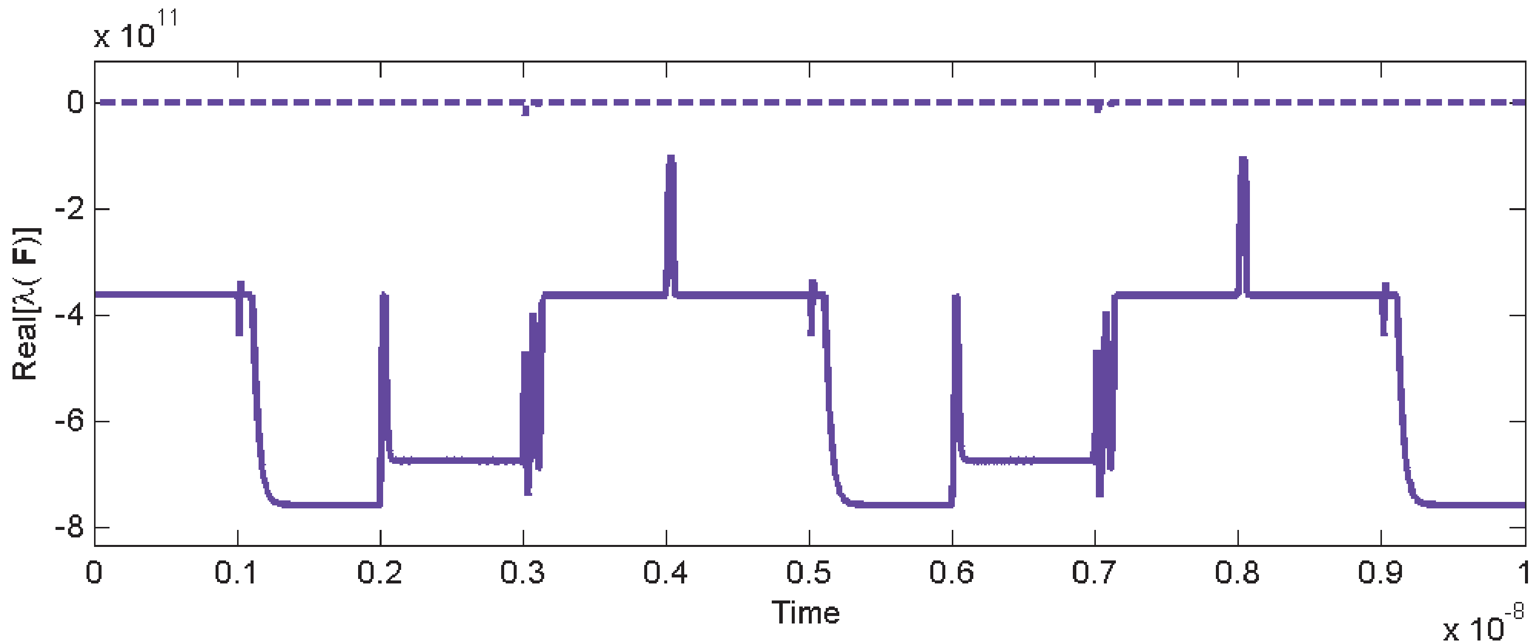




Toggle Circuit Sensitivity

- System dynamics has components which are strongly contractive
  - Sensitivity matrix of continuous dynamics has eigenvalues with large negative real component

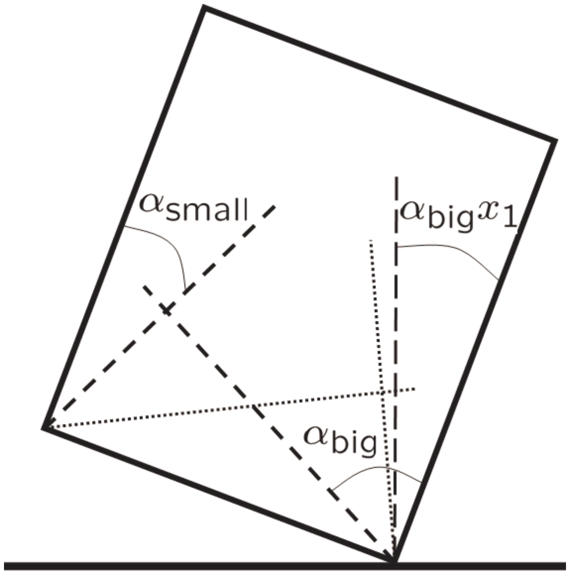
$$\mathbf{F}(q, x, u) \triangleq \mathbf{D}_x f(q, x, u)$$

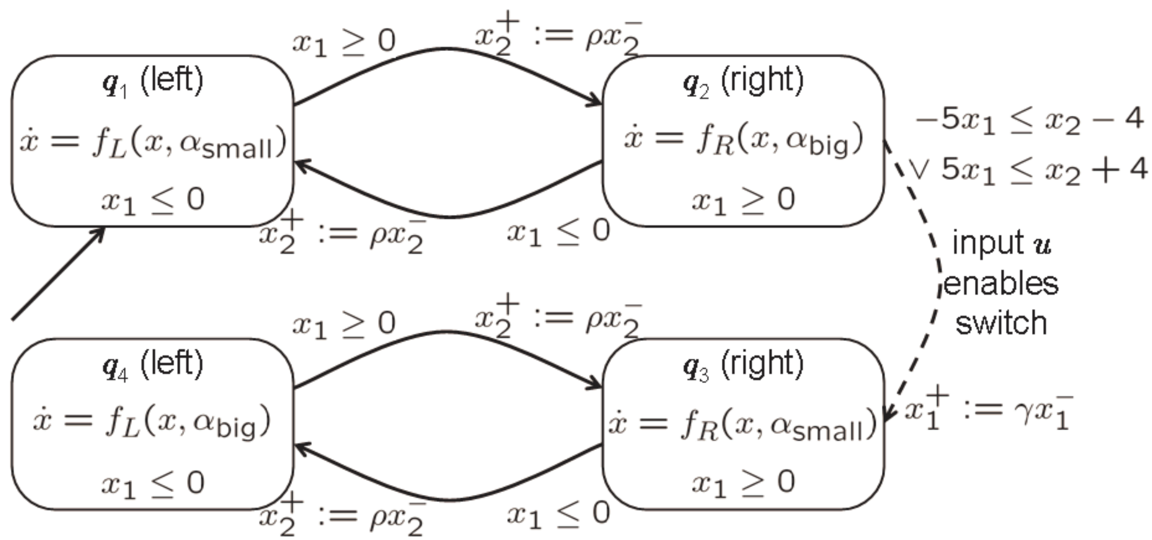


- Backward reachability will be ill-conditioned

Discrete System Sensitivity Example

- Adapted from rocking block in [Lygeros, Johansson, Simic, Zhang & Sastry, IEEE TAC 2003]
  - Discrete control input can change location of center of mass



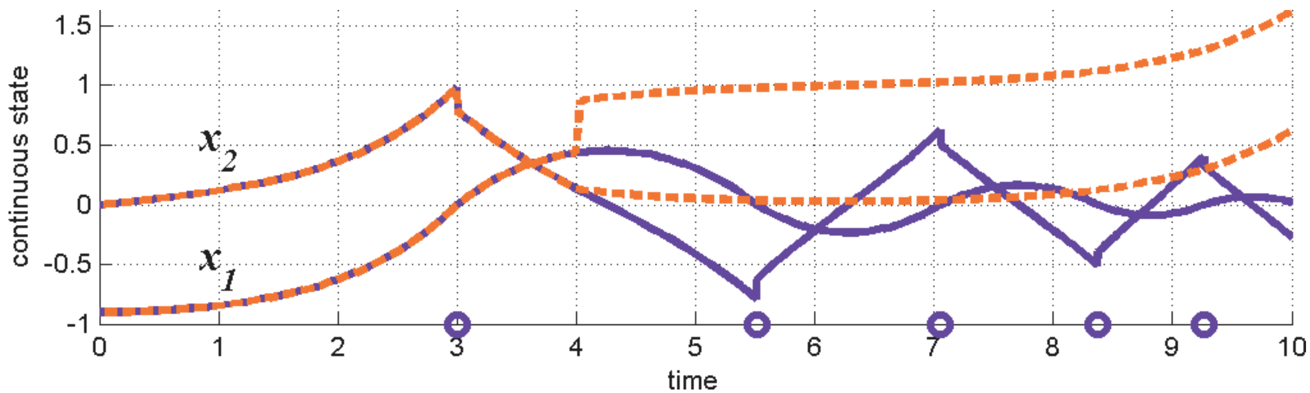


$$f_L(x, \alpha) = \begin{bmatrix} x_2 \\ \frac{1}{\alpha} \sin(\alpha(1 + x_1)) \end{bmatrix} \qquad f_R(x, \alpha) = \begin{bmatrix} x_2 \\ -\frac{1}{\alpha} \sin(\alpha(1 - x_1)) \end{bmatrix}$$

$$\alpha_{\text{big}} = \pi/3, \quad \alpha_{\text{small}} = \pi/6, \quad \rho = 0.8, \quad \gamma = \alpha_{\text{big}}/\alpha_{\text{small}} = 2.$$

Rocking Block Sensitivity

- Two typical trajectories
  - Constant center of mass (blue) or switched (red)



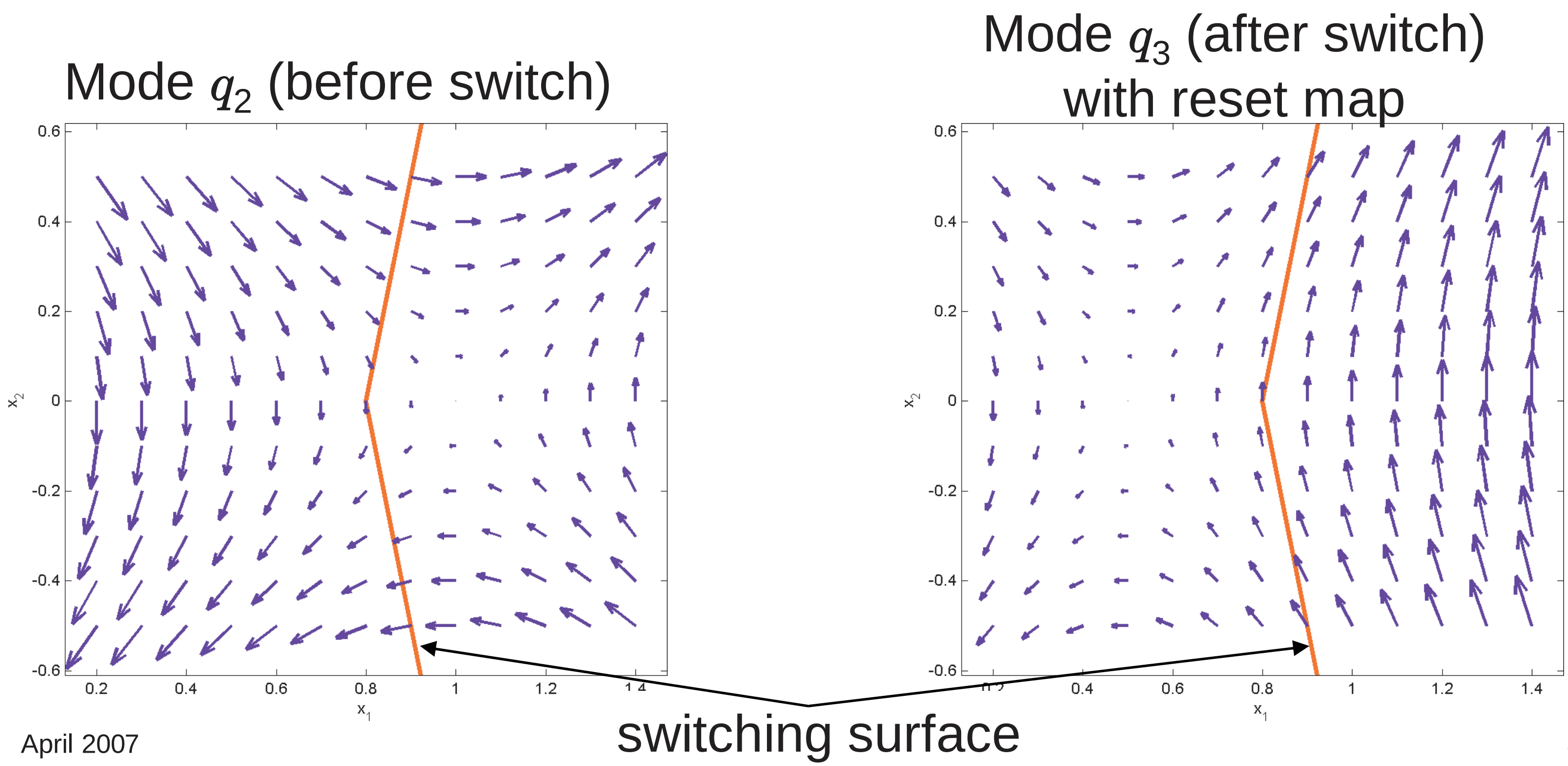
- Forward behaviour
  - Final state is sensitive to initial conditions (tipped or not)
  - Switching (controlled or autonomous) is not locally sensitive
- Backward behaviour
  - Controlled switch is sensitive through interaction with reset
  - Reset is sensitive for $\rho \ll 1$

Switching Surface Sensitivity

- Backward switching sensitivity is not obvious

$$\overleftarrow{\tau} = \frac{\partial t(z_0)}{\partial z_0} = \frac{\nabla \psi(x^+)^T \, \Xi_{\mathsf{H}}(t^+)}{\nabla \psi(x^+)^T f(q^+, x^+, u)}$$

$$\nabla \psi(x^+)^T f(q^+, x^+, u) = 0 \text{ at } x^+ \approx \frac{1}{23}\begin{bmatrix} 19 \\ 3 \end{bmatrix}$$



Conclusions

- All reachability operators are effective for proving universal safety over all input signals
- Only backward reach tube is effective for proving existence of a safe input signal
- For typical models, ill conditioning is more likely to occur for backward operators
- Results depend on the desired operator, not the algorithm

Comparing Forward and Backward Reachability as Tools for Safety Analysis

For more information contact

Ian Mitchell

Department of Computer Science

The University of British Columbia

mitchell@cs.ubc.ca

http://www.cs.ubc.ca/~mitchell