Identifying Single-Input Discrete Linear System Dynamics from Reachable Sets

Taha Shafa[1] Roy Dong[2] Melkior Ornik[3]

[1]Aerospace Engineering Ph.D. Candidate
University of Illinois Urbana-Champaign

[2]Industrial and Systems Engineering Faculty
University of Illinois Urbana-Champaign

[3]Aerospace Engineering Faculty
University of Illinois Urbana-Champaign

Conference on Decision and Control, December 2023

Motivation

Motivation

- We want to understand and model the behavior of systems with limited information

Motivation

- We want to understand and model the behavior of systems with limited information
  - Adversarial systems

Motivation

- We want to understand and model the behavior of systems with limited information
  - Adversarial systems
  - Large scale systems without ability to induce actuation effort

Motivation

- We want to understand and model the behavior of systems with limited information
  - Adversarial systems
  - Large scale systems without ability to induce actuation effort
- We can observe the system behavior and determine what it can achieve within a given time

Motivation

- We want to understand and model the behavior of systems with limited information
  - Adversarial systems
  - Large scale systems without ability to induce actuation effort
- We can observe the system behavior and determine what it can achieve within a given time
- Examples include:
  - Large-scale swarm robots

Motivation

- We want to understand and model the behavior of systems with limited information
  - Adversarial systems
  - Large scale systems without ability to induce actuation effort
- We can observe the system behavior and determine what it can achieve within a given time
- Examples include:
  - Large-scale swarm robots
  - Internal body processes

Motivation

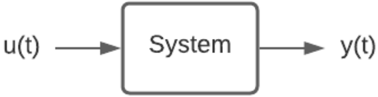
- We want to understand and model the behavior of systems with limited information
  - Adversarial systems
  - Large scale systems without ability to induce actuation effort
- We can observe the system behavior and determine what it can achieve within a given time
- Examples include:
  - Large-scale swarm robots
  - Internal body processes
  - Economic trends resulting from macro population behaviors

Motivation

- We want to understand and model the behavior of systems with limited information
  - Adversarial systems
  - Large scale systems without ability to induce actuation effort
- We can observe the system behavior and determine what it can achieve within a given time
- Examples include:
  - Large-scale swarm robots
  - Internal body processes
  - Economic trends resulting from macro population behaviors
- Existing model methods involve system identification

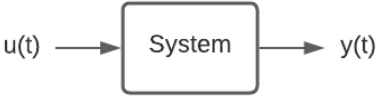
System Identification

- Create a mapping from input trajectories to observed output trajectories
  - Recursive least squares approach
  - Neural networks



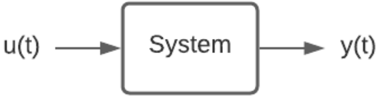
System Identification

- Create a mapping from input trajectories to observed output trajectories
  - Recursive least squares approach
  - Neural networks
- Requires knowledge of individual trajectories and control of actuators



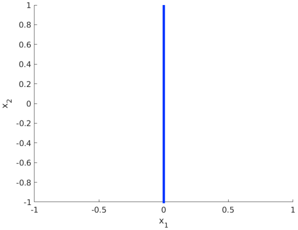
System Identification

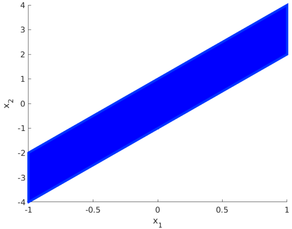
- Create a mapping from input trajectories to observed output trajectories
  - Recursive least squares approach
  - Neural networks
- Requires knowledge of individual trajectories and control of actuators
- What if you have neither?
  - What other information could you use?



Reachable Sets

- We have the system dynamics
  - $x[i+1] = Ax[i] + bu[i], \quad x[0] = 0$
- For $i \in \mathbb{Z}_{\geq 0}$, the (forward) reachable set for the system at time i is
  - $\mathcal{R}(i, x[0]) = \{\phi_u(i; x[0]) \mid u : \mathbb{Z}_{\geq 0} \to \mathcal{U}\}$
- Defines the set of states which are reachable at time i for the system using all $u \in \mathcal{U}$

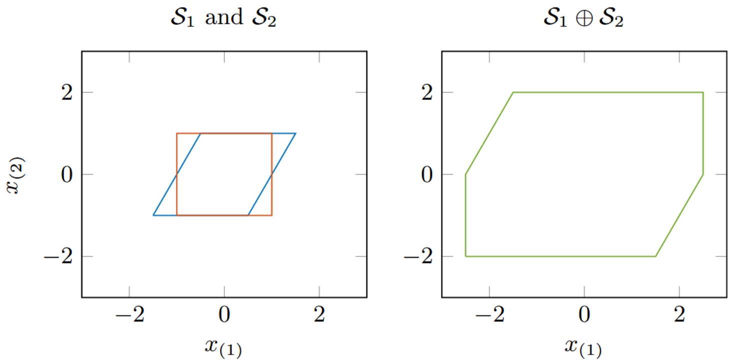




Minkowski Sum

- Given two sets $\mathcal{S}_1, \mathcal{S}_2 \in \mathbb{R}^n$ we denote

$$\mathcal{S}_1 \oplus \mathcal{S}_2 = \{s_1 + s_2 \mid s_1 \in \mathcal{S}_1,\ s_2 \in \mathcal{S}_2\}$$



Technical Assumptions

- Uniquely determine system model under the following assumptions:
  - Single-Input System
  - Generic Properties
  - Discrete Linear System
    - $x[i+1] = Ax[i] + bu[i], \quad x[0] = 0, \quad u \in \mathcal{U}$
    - Consecutive unit-length time steps
  - Fully Controllable System
  - The input set $\mathcal{U}$ is completely known

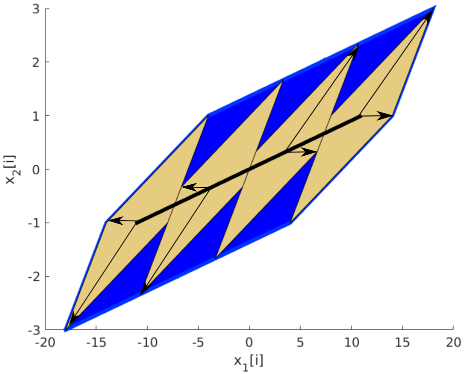
How can we solve for A and b given the assumptions above?

Reachable Sets as Minkowski Sums

- We have the following system:
  - $x[i+1] = Ax[i] + bu[i], \quad x[0] = 0, \quad u \in \mathcal{U}$
- $\mathcal{R}(i,0) =$ the reachable set of the system above at time i
- $\mathcal{R}(1,0) = b\mathcal{U}$
- $x[i] = A^i x[0] + A^{i-1}bu[0] + \ldots + bu[i-1]$
- This implies:
  - $\mathcal{R}(i,0) = A^{i-1}b\mathcal{U} \oplus \ldots \oplus b\mathcal{U}$
  - $\mathcal{R}(i,0) = A^{i-1}b\mathcal{U} \oplus \mathcal{R}(i-1,0)$
- Therefore:
  - $A^{i-1}b\mathcal{U} = \mathcal{R}(i,0) \ominus \mathcal{R}(i-1,0)$

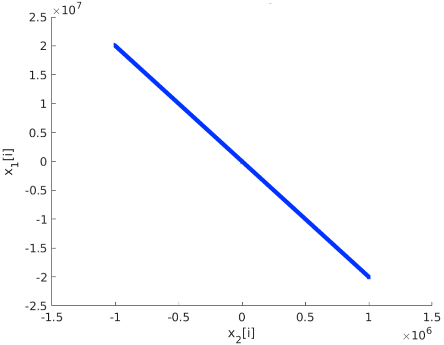
Minkowski Difference

- Given two sets $\mathcal{A}$, $\mathcal{B} \in \mathbb{R}^n$
  - $\mathcal{A} \ominus \mathcal{B} = \{c \in \mathbb{R}^n \mid c \oplus \mathcal{B} \subseteq \mathcal{A}\}$
- Let $v^{(i)} \in \mathcal{V}$ be the vertices of $\mathcal{R}(i-1, 0)$
  - $\mathcal{R}(i,0) \ominus \mathcal{R}(i-1,0) = \bigcap_{v^{(i)} \in \mathcal{V}} (\mathcal{R}(i,0) - v^{(i)})$



Solving for Dynamics

- Solving for the vector b with knowledge of $\mathcal{U}$ is trivial
  - $\mathcal{R}(1,0) = b\mathcal{U}$
- Similarly we can solve for $A^i b$ with knowledge of $\mathcal{U}$
  - $A^{i-1}b\mathcal{U} = \mathcal{R}(i,0) \ominus \mathcal{R}(i-1,0)$
- If we have a controllable system
  - $C_{A,b} = \begin{bmatrix} b & Ab & \dots & A^{i-1}b \end{bmatrix}$
  - $C_{A,b}$ is invertible
  - $A = AC_{A,b}C_{A,b}^{-1}$



Uniqueness

Can we determine if the calculated dynamics A and b are unique?

- In general cases no
  - $A = I, \quad b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$
  - $A = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

Uniqueness

Can we determine if the calculated dynamics A and b are unique?

- In general cases no
  - $A = I, \quad b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$
  - $A = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$
- Yes under the following conditions:

Uniqueness

Can we determine if the calculated dynamics A and b are unique?

- In general cases no
  - $A = I, \quad b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$
  - $A = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$
- Yes under the following conditions:
  - Generic system properties
    - A is invertible
    - A is diagonalizable
    - A contains distinct eigenvalues
    - First element of right eigenvectors of A are nonzero

Uniqueness

Can we determine if the calculated dynamics A and b are unique?

- In general cases no
  - $A = I, \quad b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$
  - $A = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$
- Yes under the following conditions:
  - Generic system properties
    - A is invertible
    - A is diagonalizable
    - A contains distinct eigenvalues
    - First element of right eigenvectors of A are nonzero
  - Any generic system with an input set $\mathcal{U}$ asymmetric around the origin

Uniqueness

Can we determine if the calculated dynamics A and b are unique?

- In general cases no
  - $A = I, \quad b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$
  - $A = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$
- Yes under the following conditions:
  - Generic system properties
    - A is invertible
    - A is diagonalizable
    - A contains distinct eigenvalues
    - First element of right eigenvectors of A are nonzero
  - Any generic system with an input set $\mathcal{U}$ asymmetric around the origin
  - Generic 2D system with an input set $\mathcal{U}$ symmetric around the origin

Asymmetric Input Set Proof

If there exist sets of dynamics A, b and A', b' which satisfy $\mathcal{R}(i, 0)$ for all i, then $A = A'$ and $b = b'$

- With change of basis, $A^k e_1 = A'^k e_1$ for all $k \in \mathbb{Z}_{\geq 0}$

Asymmetric Input Set Proof

If there exist sets of dynamics A, b and A', b' which satisfy $\mathcal{R}(i, 0)$ for all i, then $A = A'$ and $b = b'$

- With change of basis, $A^k e_1 = A'^k e_1$ for all $k \in \mathbb{Z}_{\geq 0}$
- $A^k e_1 = A'^{(-1)} A^k A' e_1$ for all $k \in \mathbb{Z}_{\geq 0}$
  - Eigenvalues of A^k and $A'^{(-1)} A^k A'$ are equal

Asymmetric Input Set Proof

If there exist sets of dynamics A, b and A', b' which satisfy $\mathcal{R}(i, 0)$ for all i, then $A = A'$ and $b = b'$

- With change of basis, $A^k e_1 = A'^k e_1$ for all $k \in \mathbb{Z}_{\geq 0}$
- $A^k e_1 = A'^{(-1)} A^k A' e_1$ for all $k \in \mathbb{Z}_{\geq 0}$
  - Eigenvalues of A^k and $A'^{(-1)} A^k A'$ are equal
- Let v_i, η_i and v'_i, η'_i denote left and right eigenvectors of A and A' respectively
  - Subtract $A'^{(-1)} A^k A' e_1$ from both sides and perform eigenvalue decomposition of A^k and $A'^{(-1)} A^k A'$ to get
  $$\sum_i \lambda_i^k (v_i - v'_i \eta'^T_{i1}) = 0 \quad \forall\, k \in \mathbb{Z}_{\geq 0}.$$
  - η scaled so $\eta_{i1} = 1$

Asymmetric Input Set Proof

- Taking $k \in \{0, \ldots, n-1\}$ we now have a series of n equations

$$\Lambda S_j = \begin{bmatrix} 1 & \ldots & 1 \\ \lambda_1 & \ldots & \lambda_n \\ \vdots & \vdots & \vdots \\ \lambda_1^{n-1} & \ldots & \lambda_n^{n-1} \end{bmatrix} \begin{bmatrix} v_{1j} - v'_{1j}\eta'^T_{11} \\ v_{2j} - v'_{2j}\eta'^T_{21} \\ \vdots \\ v_{nj} - v'_{nj}\eta'^T_{n1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

  for any $j \in \{1, \ldots, n\}$

- Notice $\Lambda \in \mathbb{C}^{n \times n}$ is the square Vandermonde matrix
  - Λ is invertible if λ_i are distinct for all i.

Asymmetric Input Set Proof

- Taking $k \in \{0, \dots, n-1\}$ we now have a series of n equations

$$\Lambda S_j = \begin{bmatrix} 1 & \dots & 1 \\ \lambda_1 & \dots & \lambda_n \\ \vdots & \vdots & \vdots \\ \lambda_1^{n-1} & \dots & \lambda_n^{n-1} \end{bmatrix} \begin{bmatrix} v_{1j} - v'_{1j}\eta'^T_{11} \\ v_{2j} - v'_{2j}\eta'^T_{21} \\ \vdots \\ v_{nj} - v'_{nj}\eta'^T_{n1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

  for any $j \in \{1, \dots, n\}$

- Notice $\Lambda \in \mathbb{C}^{n \times n}$ is the square Vandermonde matrix
  - Λ is invertible if λ_i are distinct for all i.
  - If $\eta'_{ij} = 0$ for any i, then $v_i = 0$
    - Contradicts assumption that A is diagonalizable

Asymmetric Input Set Proof

- Taking $k \in \{0, \ldots, n-1\}$ we now have a series of n equations

$$\Lambda S_j = \begin{bmatrix} 1 & \ldots & 1 \\ \lambda_1 & \ldots & \lambda_n \\ \vdots & \vdots & \vdots \\ \lambda_1^{n-1} & \ldots & \lambda_n^{n-1} \end{bmatrix} \begin{bmatrix} v_{1j} - v'_{1j}\eta'^{T}_{11} \\ v_{2j} - v'_{2j}\eta'^{T}_{21} \\ \vdots \\ v_{nj} - v'_{nj}\eta'^{T}_{n1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

  for any $j \in \{1, \ldots, n\}$

- Notice $\Lambda \in \mathbb{C}^{n \times n}$ is the square Vandermonde matrix
  - Λ is invertible if λ_i are distinct for all i.
  - If $\eta'_{ij} = 0$ for any i, then $v_i = 0$
    - Contradicts assumption that A is diagonalizable
  - $v_{ij} = v'_{ij}$ for all i, j since Λ is invertible
    - $A = A'^{(-1)} A A'$
    - Eigenvectors of commuting matrices A and A' are equivalent

Asymmetric Input Set Proof

- Recall A and A' are diagonalizable

Asymmetric Input Set Proof

- Recall A and A' are diagonalizable
- Recall we prove
  - $A^k e_1 = A'^k e_1$ for all $k \in \mathbb{Z}_{\geq 0}$

Asymmetric Input Set Proof

- Recall A and A' are diagonalizable
- Recall we prove
  - $A^k e_1 = A'^k e_1$ for all $k \in \mathbb{Z}_{\geq 0}$
- Take eigenvalue expansion of A and A' above and multiply on left by inverse of eigenvectors to get

$$(\lambda_i^k - \lambda'^k)\eta_{i1} = 0 \quad \forall\ k \in \mathbb{Z}_{\geq 0}$$

Asymmetric Input Set Proof

- Recall A and A' are diagonalizable
- Recall we prove
  - $A^k e_1 = A'^k e_1$ for all $k \in \mathbb{Z}_{\geq 0}$
- Take eigenvalue expansion of A and A' above and multiply on left by inverse of eigenvectors to get

$$(\lambda_i^k - \lambda'^k)\eta_{i1} = 0 \quad \forall\ k \in \mathbb{Z}_{\geq 0}$$

- By assumption, $\eta_{i1} \neq 0$ for all i
  - $\lambda_i^k = \lambda_i'^k$ for all $k \in \mathbb{Z}_{\geq 0}$

Asymmetric Input Set Proof

- Recall A and A' are diagonalizable
- Recall we prove
  - $A^k e_1 = A'^k e_1$ for all $k \in \mathbb{Z}_{\geq 0}$
- Take eigenvalue expansion of A and A' above and multiply on left by inverse of eigenvectors to get

$$(\lambda_i^k - \lambda'^k)\eta_{i1} = 0 \quad \forall\ k \in \mathbb{Z}_{\geq 0}$$

- By assumption, $\eta_{i1} \neq 0$ for all i
  - $\lambda_i^k = \lambda_i'^k$ for all $k \in \mathbb{Z}_{\geq 0}$
- $A = A'$
  - Q.E.D.

Symmetric Input Set Proof

- Similar steps to previous proof result in

$$A^{k-1}e_1 = (-1)^{q(k)}A'^{(-1)}A^{k-1}A'e_1 \quad \forall\, k \in \mathbb{Z}_{\geq 0}$$

  where $q(k) \in \{0,1\}$

- We arrive at a series of equations

$$\sum_i \lambda_i^{k-1}(v_i - (-1)^{q(k)}v_i'\eta_{i1}'^T) = 0 \quad \forall\, k \in \mathbb{Z}_{\geq 0}.$$

- We cannot easily arrive at the Vandermonde matrix as before
- If $q(k)$ is constant for all k or alternating every k, then $A = A'$
  - Can be proven for $n = 2$

Bandpass Filter Circuit Example

- Fourth-order band-pass circuit
  - Controllable canonical representation

$$x[i+1] = Ax[i] + bv_c[i] = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -a_0 & -a_1 & -a_2 & -a_3 \end{bmatrix} x[i] + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} v_c[i]$$

- $v_c[i] \in [0, 1]$ for all $i \in \mathbb{Z}_{\geq 0}$
- If $a_0 \neq 0$ then assumptions are likely satisfied
  - A invertible
  - A diagonalizable
- Want to recover the true parameters
  - $a_0 = 3,\ a_1 = 2,\ a_2 = 3,\ a_3 = 6$

Bandpass Filter Circuit Example

$$\mathcal{R}(1,0) = \operatorname{conv}\left(\begin{bmatrix}0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\1\end{bmatrix}\right), \mathcal{R}(2,0) = \operatorname{conv}\left(\begin{bmatrix}0\\0\\1\\-5\end{bmatrix}, \begin{bmatrix}0\\0\\1\\-6\end{bmatrix}, \begin{bmatrix}0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\1\end{bmatrix}\right)$$

$$\mathcal{R}(3,0) = \operatorname{conv}\left(\begin{bmatrix}0\\0.86\\-6.02\\33.00\end{bmatrix}, \begin{bmatrix}0\\-0.14\\0.98\\-6.00\end{bmatrix}, \begin{bmatrix}0\\-0.14\\0.98\\-5.00\end{bmatrix}, \begin{bmatrix}0\\0.86\\-6.02\\34.00\end{bmatrix}\right)$$

$$\mathcal{R}(4,0) = \operatorname{conv}\left(\begin{bmatrix}-0.15\\1.05\\-5.99\\33.00\end{bmatrix}, \begin{bmatrix}0.85\\-5.95\\34.01\\-188\end{bmatrix}, \begin{bmatrix}0.85\\-5.95\\34.01\\-187\end{bmatrix}, \begin{bmatrix}-0.15\\1.05\\-5.99\\34.00\end{bmatrix}\right)$$

$$\mathcal{R}(5,0) = \operatorname{conv}\left(\begin{bmatrix}-5.93\\33.88\\-188.02\\1035.00\end{bmatrix}, \begin{bmatrix}1.07\\-6.12\\33.98\\-188.00\end{bmatrix}, \begin{bmatrix}1.07\\-6.12\\33.98\\-187.00\end{bmatrix}, \begin{bmatrix}-5.93\\33.88\\-188.02\\1036.00\end{bmatrix}\right)$$

Bandpass Filter Circuit Example

- $\mathcal{R}(1,0) = b\mathcal{U}$ where $\mathcal{U} = [0,1]$
  - b can be trivially computed
  - $b = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^T$

Bandpass Filter Circuit Example

- $\mathcal{R}(1,0) = b\mathcal{U}$ where $\mathcal{U} = [0,1]$
  - b can be trivially computed
  - $b = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^T$
- Recall $A^{i-1}b\mathcal{U} = \mathcal{R}(i,0) \ominus \mathcal{R}(i-1,0)$

$$Ab = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -6 \end{bmatrix}, \; A^2b = \begin{bmatrix} 0 \\ 1 \\ -6 \\ 33 \end{bmatrix}, \; A^3b = \begin{bmatrix} 1 \\ -6 \\ 33 \\ -182 \end{bmatrix}, \; A^4b = \begin{bmatrix} -6 \\ 33 \\ -182 \\ 1002 \end{bmatrix}$$

Bandpass Filter Circuit Example

- $\mathcal{R}(1,0) = b\mathcal{U}$ where $\mathcal{U} = [0,1]$
  - b can be trivially computed
  - $b = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^T$
- Recall $A^{i-1}b\mathcal{U} = \mathcal{R}(i,0) \ominus \mathcal{R}(i-1,0)$

$$Ab = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -6 \end{bmatrix}, \; A^2b = \begin{bmatrix} 0 \\ 1 \\ -6 \\ 33 \end{bmatrix}, \; A^3b = \begin{bmatrix} 1 \\ -6 \\ 33 \\ -182 \end{bmatrix}, \; A^4b = \begin{bmatrix} -6 \\ 33 \\ -182 \\ 1002 \end{bmatrix}$$

- $A = AC_{A,b}C_{A,b}^{-1} = \begin{bmatrix} A^4b & A^3b & A^2b & Ab \end{bmatrix} \begin{bmatrix} A^3b & A^2b & Ab & b \end{bmatrix}^{-1}$

Bandpass Filter Circuit Example

- $\mathcal{R}(1,0) = b\mathcal{U}$ where $\mathcal{U} = [0,1]$
  - b can be trivially computed
  - $b = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^T$
- Recall $A^{i-1}b\mathcal{U} = \mathcal{R}(i,0) \ominus \mathcal{R}(i-1,0)$

$$Ab = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -6 \end{bmatrix},\ A^2b = \begin{bmatrix} 0 \\ 1 \\ -6 \\ 33 \end{bmatrix},\ A^3b = \begin{bmatrix} 1 \\ -6 \\ 33 \\ -182 \end{bmatrix},\ A^4b = \begin{bmatrix} -6 \\ 33 \\ -182 \\ 1002 \end{bmatrix}$$

- $A = AC_{A,b}C_{A,b}^{-1} = \begin{bmatrix} A^4b & A^3b & A^2b & Ab \end{bmatrix} \begin{bmatrix} A^3b & A^2b & Ab & b \end{bmatrix}^{-1}$
- If we want to find A for an n dimensional system, we can do so with $n+1$ reachable sets

Analysis

- We developed a method of system identification that does not require knowledge of individual trajectories
  - Instead utilizes reachable sets

Theorem

Under generic assumptions, we can **uniquely** identify the unknown dynamics of a discrete linear system with an input set asymmetric around the origin using $n+1$ reachable sets for unit time intervals

Remark

The theorem above holds for systems with symmetric input sets of dimension 2

Future Work

- Prove previous remark for any system of dimension n
- Solve the problem for scenarios with less information
  - $x[0] \neq 0$
  - Reachable sets are not at unit time intervals
  - Extend to continuous time systems
  - Determine models within some error bounds if we have part of the reachable sets
- Extend this work to multi-input systems
  - 2-input systems
    - Zonotopes are not closed under Minkowski difference when $\mathcal{U} \subseteq \mathbb{R}^m$ and $m > 2$
- Relax generic assumptions
  - Non-generic systems with a common structure in control systems
- Extend work to incorporate nonlinearity

Questions

Taha Shafa

tahaas2@illinois.edu