Reinforcement Learning for Characterization of Hysteresis Behavior in Shape Memory Alloys, and Application to Control of Morphing Air Vehicles

Kenton Kirkpatrick
John Valasek

Aerospace Engineering Department
Texas A&M University

Infotech@Aerospace Conference
8 May 2007
Rohnert Park, CA
Briefing Agenda

- Shape Memory Alloy Characterization and Control
- Learning
- Simulation Results
- Experimental Results
- Conclusions
- Extensions and Future Directions
Today’s Challenges:

- Develop light, strong, and structurally efficient air vehicles.
- Improved aerodynamic efficiency.
- Design fuel-efficient, low-emission propulsion systems.
- Develop safe, fault-tolerant vehicle systems.

Technology Solutions:

- Nanostructures: 100 times stronger than steel at 1/6 the weight
- Active flow control
- Distributed propulsion
- Electric propulsion, advanced fuel cells, high-efficiency electric motors
- Integrated advanced control systems and information technology
- Central “nervous system” and adaptive vehicle control
Shape Memory Alloys (SMA)

- Metallic Alloys used as actuators
- Shape Memory Effect – SMA can fully recover from a plastically deformed shape change by the addition of heat
- Electricity can be used to induce a cycle of heating, cooling, and deformation in an SMA in order to execute a dynamic task

**PROBLEM:**
Efficiently characterizing and controlling SMA behavior
Why Is Characterizing SMAs Such A Challenge?

1. Uncertain Model Parameters
2. Temperature–Strain Relationship: **HYSTERESIS**
   - Behavior based on a 2-D FIELD, not a 2-D PATH
SMA Characterization Methodologies

**CURRENT METHOD:** Experimentation and Testing to approximate constitutive model

- Mainly based on material property parameters OR system identification
  - Approximate Models:
    - Neglect hysteresis
    - Arrange antagonist SMAs to “cancel” the hysteretic effect
    - Modified plasticity model composed of averaged thermal effects
    - Disregard coupling of hysteretic and structural response
    - Other averaging techniques
Research Strategy

1. Develop Machine Learning algorithm to characterize and control SMA position states with time dependency

2. Test Machine Learning method in numerical simulation of SMA input-output (temperature-strain) behavior

3. Compare the RL unit’s deduction of SMA behavior with current mathematical models and experimental results of SMA behavior

4. Initial Scope: 1-D Ni-Ti SMA Wire Position Control

BENEFITS

- Reduction in time, labor, cost, resources
- Streamlining of modeling and simulation steps
- Better characterization information/data
Reinforcement Learning - 1

- Does not require any prior knowledge.
  - Knowledge is based on experience and interaction with the environment, not on input-output data supplied by an external supervisor

- Achieves a specific goal by learning from interactions with the environment.
  - Considers state information (s)
  - Performs sequences of actions, (a), observing the consequences
  - Attempts to maximize rewards (r) over time
    - These specify what is to be achieved, not how to achieve it

- Constructs a state value function (V)
  - Learns an optimal control policy
    \[ \pi^* = \arg \max_a [r(s,a) + \gamma V^*(\delta(s,a))] \]

- Memory is contained in the state value function
  \[ V^*(s) = \max_{\pi} V^\pi(s) \text{ for all } s \]

Kirkpatrick & Valasek - 9
Actor-critic method
- On policy, Temporal Displacement method
- **The actor selects actions**
  Preference of an action: \( p(s, a) \)

  **Greedy Policy:**
  \[ \pi_t(s, a) = \arg \max_a p(s, a) \]

  **Gibbs softmax policy:**
  \[ \pi_t(s, a) = \Pr\{a_t = a \mid s_t = s\} = \frac{e^{p(s,a)}}{\sum_a e^{p(s,a)}} \]

- **The critic criticizes the actions**
  State value functions: \( V(s_t) \)

  **TD error:**
  \[ \delta_t = r_{t+1} + \gamma V(s_{t+1}) - V(s_t) \]

  - Strengthen or weaken the tendency to select one action
  \[ p(s_t, a_t) \leftarrow p(s_t, a_t) + \beta \delta_t \]

  *Kirkpatrick & Valasek - 10*
To test the RL algorithm, a simulation was first written for a simple grid learning problem:

- **States**: Defined by a grid including 20 locations:
  - States 1-8 are considered in-bounds while 9-20 are out-of-bounds
  - **Goal State** is randomly selected at the start of each run; Always between States 1-8
- **Actions**: Up, Down, Left, and Right
- **Rewards**:
  - 1 for attaining the Goal State
  - 0 for any in-bound State that is not the Goal State
  - -1 for going out-of-bounds

![Simulation Test Graph]

<table>
<thead>
<tr>
<th>Action</th>
<th>Percent Optimal Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>1000</td>
<td>20%</td>
</tr>
<tr>
<td>2000</td>
<td>40%</td>
</tr>
<tr>
<td>3000</td>
<td>60%</td>
</tr>
<tr>
<td>4000</td>
<td>80%</td>
</tr>
<tr>
<td>5000</td>
<td>100%</td>
</tr>
</tbody>
</table>

Kirkpatrick & Valasek - 11
SMA Results

a) Simulation

b) Experiment
Simulation Model

- **Temperature-Strain Relation:**
  - Hyperbolic Tangent Model

  \[
  \text{Strain}(i) = \frac{1}{2} h \tanh((\text{Tem} - \text{ctl}) a) + s \left( \text{Tem} - \frac{1}{2} \text{ctl} - \frac{1}{2} \text{ctr} \right) + \frac{1}{2} h + cs
  \]

- **Voltage-Temperature Relation:**

  \[
  \frac{dT}{dt} = \frac{V^2}{R} - h \pi D L (T - T_\infty) \quad \frac{1}{m C_p}
  \]

- Learns input-output data
  (how to apply voltage to achieve a particular position state), not the constitutive model of an SMA
Reinforcement Learning Agent

- **Agent**: SMA wire

- **Goal**: Desired strain (position)

- **States**: Temperature and strain

- **Actions**: Increase temperature, decrease temperature
  - range of 0.01 to 0.1 volts

- **Rewards**: +1 for attaining the goal state
  - 0 for any other in-bound state
  - -1 for going out of bounds (outside hysteresis loop)

- **Optimal control policy**: The **Action** which brings the **Agent** to the **Goal** with the fewest **Actions**
Simulating Results

- Goal State: 1 ± 0.3 % Strain
- Learning Time dependent on Size of Control Policy, Q
  - Q size dependent on size of Temp-Strain Mesh
  - Strain ranges: 0.3% & Temperature ranges: 10° C → 170 States
Simulation Results

- Goal State: 3 ± 0.1 % Strain
- Temp-Strain Mesh Size: 500 States
- Control Optimization: 86%
- Max Strain Error (explored states): 0.01%
Simulation Results

Hysteresis Behavior

- Temperature (C)
- Strain (%)
SMA Results

a) Simulation

b) Experiment
Experimental Setup - 1

- LVDT: Position Sensor
- Thermocouple: Temperature Sensor
- Voltage Input: 0-2.3 Volts
- Tensile Stress: 105 MPa
Experimental Setup - 2

[Description of the setup with labeled components: Cooling Fan, Liquid Cooler, Antifreeze Bath, DAQ Board, LVDT Power Supply, Variable Voltage Supply, Weight, Pump, Liquid Heater, etc.]
RL Algorithm: Sarsa

\[ Q(S,A,G) = Q(S,A,G) + \alpha \left[ R(S',A',G) + \gamma Q(S',A',G) - Q(S,A,G) \right] \]

3-D Control Policy Matrix

- States: Temperature and Strain Dependent
- Actions: Change Temperature (Voltage Application)
- Goal: Desired Strain
- \( \alpha \): Repetition Penalty
- \( \gamma \): Future Policy Weight

Action Choice Method: \( \epsilon \)-Greedy

- Explore or Exploit: Dependent upon \( \epsilon \)
- \( \epsilon \) Varies with Episodes
Ni-Ti SMA Wire Specimen Properties

L = length = 5.118 in.
D = diameter = 0.0248 in.
\( \rho \) = density = 0.235 lb/in³
\( A_s \) = austenite start temperature = 30°C
\( A_f \) = austenite finish temperature = 75°C
\( T_m \) = melting point = 1300°C
Ni = nickel composition percentage = 55.5%
Ti = titanium composition percentage = 44.5%
Experiment Supervised Characterization: Major Hysteresis

SMA Major Hysteresis

Temperature (C) vs. Strain

-0.005 0 0.005 0.01 0.015 0.02 0.025 0.03 0.035

0 20 40 60 80 100 120

Experimental Data  Math Model

Kirkpatrick & Valasek - 23
Minor Hysteresis Loop Example 1

SMA Hysteresis

Temperature (C) vs. Strain
Minor Hysteresis Loop Example 2

SMA Hysteresis

Temperature (C)

Strain

Experimental Major
Math Model
Experimental Minor
Experiment – Goals v. Actions

Convergence Behavior for Goal of 2.7% Strain

Number of Actions

Episodes

0 5 10 15 20 25 30 35 40 45
Convergence Behavior for Goal of 2.7% Strain

early learning
Experiment – Early Learning Refinement

Episode 12 - 147 Actions to Goal Strain
Experiment – Goals v. Time

Convergence Behavior for Goal of 2.7% Strain

Time (s)

Episodes

mid learning
Experiment – Mid Learning Refinement

Episode 23 - 4 Actions to Goal Strain

Strain

Temperature (C)
Experiment – Goals v. Time

Convergence Behavior for Goal of 2.7% Strain

late learning
Experiment – Late Learning Refinement

Episode 30 - 1 Action to Goal Strain

Temperature (C)

Strain

Kirkpatrick & Valasek - 32
Experiment – Minor Hysteresis
Conclusions

- Reinforcement Learning successfully learns SMA hysteresis behavior without supervision, and learns the optimal control policy
  - Approach allows for the learning of both major and minor hysteresis behavior

- Sarsa better than Q-learning for this application
  - Prevents damage to specimens
  - Converges in “reasonable” number of episodes

- Ni-Ti SMAs exhibit more pronounced hysteresis behavior than Ni-Ti-Cu SMAs.
Future Directions

- Expand scope from wire (one spatial dimension displacement) to surface (two and three-dimensional displacements)

- Learn voltage-position relationship
  - More useful for automatic control

- Create Preisach model of SMA hysteresis for comparison to current simulation and experimentation results
  - Verification

- Use control policy matrix from learned hysteresis behavior to control wire length in this experimental apparatus.
Acknowledgements

- NASA Headquarters as part of the University Research & Education Training Institute (URETI) program

  Texas Institute for Intelligent Bio-Nanomaterials and Structures for Aerospace Vehicles

- Texas A&M University Undergraduate Summer Research Grant (USRG)

This support is gratefully acknowledged by the authors