Machine Learning for Robotics

Pieter Abbeel
UC Berkeley EECS

Outline

- Apprenticeship learning
  - Learning to perform dynamic maneuvers
    - Helicopter
  - Learning to perform tasks that require generalization
    - Rope manipulation
- Brief highlights of three other projects on machine learning for robotics
  - Inverse optimal control
    - Quadruped locomotion
  - Safe exploration
    - Martian surface exploration
  - Perception
    - Clothes manipulation

Challenges in helicopter control

- Unstable
- Nonlinear
- Complicated dynamics
  - Air flow
  - Coupling
  - Blade dynamics
- Noisy estimates of position, orientation, velocity, angular rate (and perhaps blade and engine speed)

Many success stories in hover and forward flight regime

- Just a few examples:
  - Bagnell & Schneider, 2001;
  - LaCivita, Papageorgiou, Messner & Kanade, 2002;
  - Ng, Kim, Jordan & Sastry 2004a (2001); Ng et al., 2004b;
  - Roberts, Corke & Buskey, 2003;
  - Saripalli, Montgomery & Sukhatme, 2003;
  - Shim, Chung, Kim & Sastry, 2003;
  - Doherty et al., 2004;
  - Gavrilets, Martinos, Mettler and Feron, 2002.
- Varying control techniques: inner/outer loop PID with hand or automatic tuning, H1, LQR, …

Example result

[Ng, Kim, Jordan, Sastry, 2004]

[Ng, Coates, Tse, et al, 2004]

One of our first attempts at autonomous flips [using similar methods to what worked for ihover]

Target trajectory: meticulously hand-engineered
Model: from (commonly used) frequency sweeps data

Aggressive, non-stationary regimes

- Gavrilets, Martinos, Mettler and Feron, 2002
  - → 3 maneuvers: split-S, snap axial roll, stall-turn

- This presentation
  - Wide range of aggressive maneuvers
  - Maneuvers in rapid succession
  - Starting point: human expert pilots

Stationary vs. aggressive flight

- Hover / stationary flight regimes:
  - Restrict attention to specific flight regime
  - Extensive data collection = collect control inputs, position, orientation, velocity, angular rate
  - Build model + model-based controller

→ Successful autonomous flight.

- Aggressive flight maneuvers --- additional challenges:
  - **Task description**: What is the target trajectory?
  - **Dynamics model**: How to build a dynamics model sufficiently accurate to enable feedback control through non-stationary flight regimes?

Learning to perform dynamic maneuvers: outline

- **Learning a target trajectory**
- Learning a dynamics model
- Autonomous flight results
  - Aerobatics
  - Auto-rotation landings

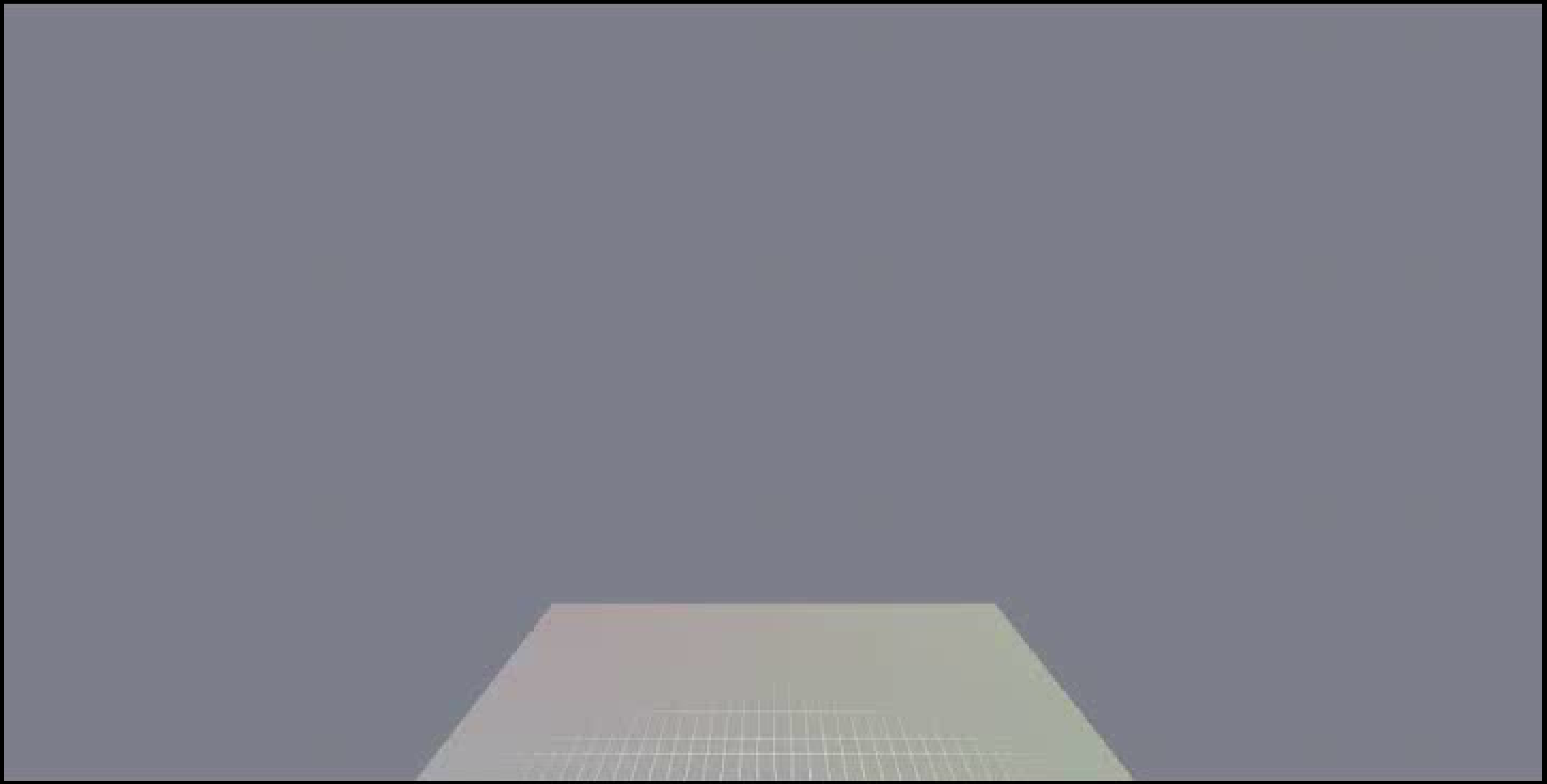
Target trajectory

- Difficult to specify by hand:
  - Required format: position + orientation over time
  - Needs to satisfy helicopter dynamics

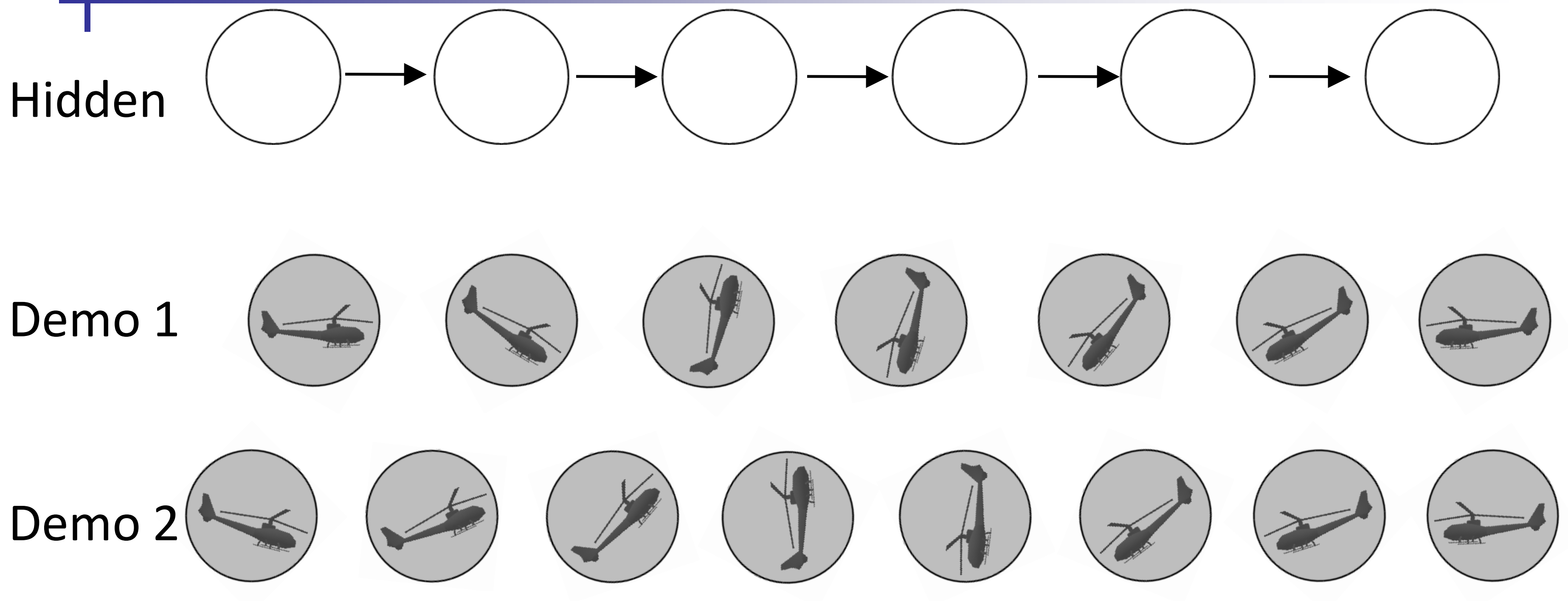
- Our solution:
  - Collect demonstrations of desired maneuvers
  - Challenge: extract a clean target trajectory from many suboptimal/noisy demonstrations

Abbeel, Coates, Ng, IJRR 2010

Expert demonstrations: Airshow



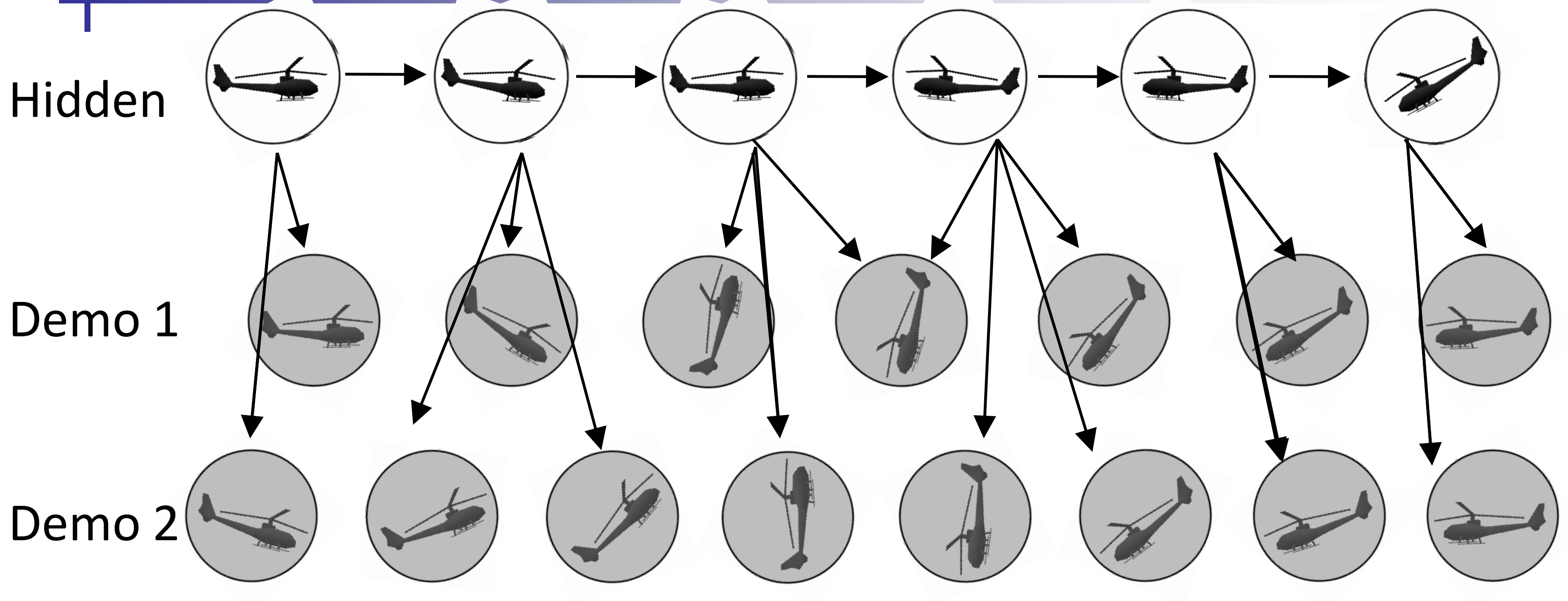
Learning Trajectory



- HMM-like generative model
  - Dynamics model used as HMM transition model
  - Demos are observations of hidden trajectory
- Problem: how do we align observations to hidden trajectory?

Abbeel, Coates, Ng, IJRR 2010

Learning Trajectory

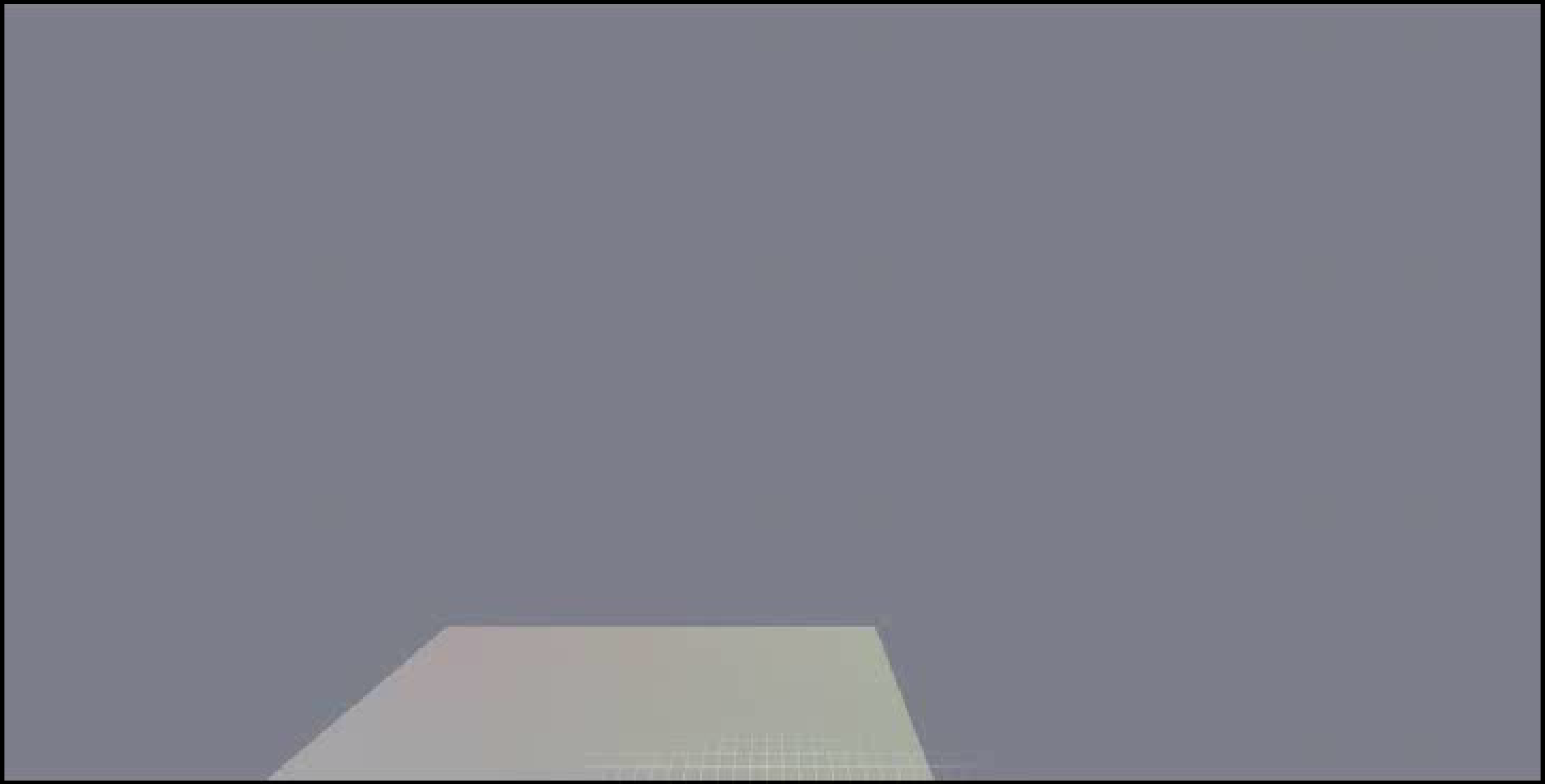


- Dynamic Time Warping (Needleman&Wunsch 1970, Sakoe&Chiba, 1978)
- Extended Kalman filter / smoother

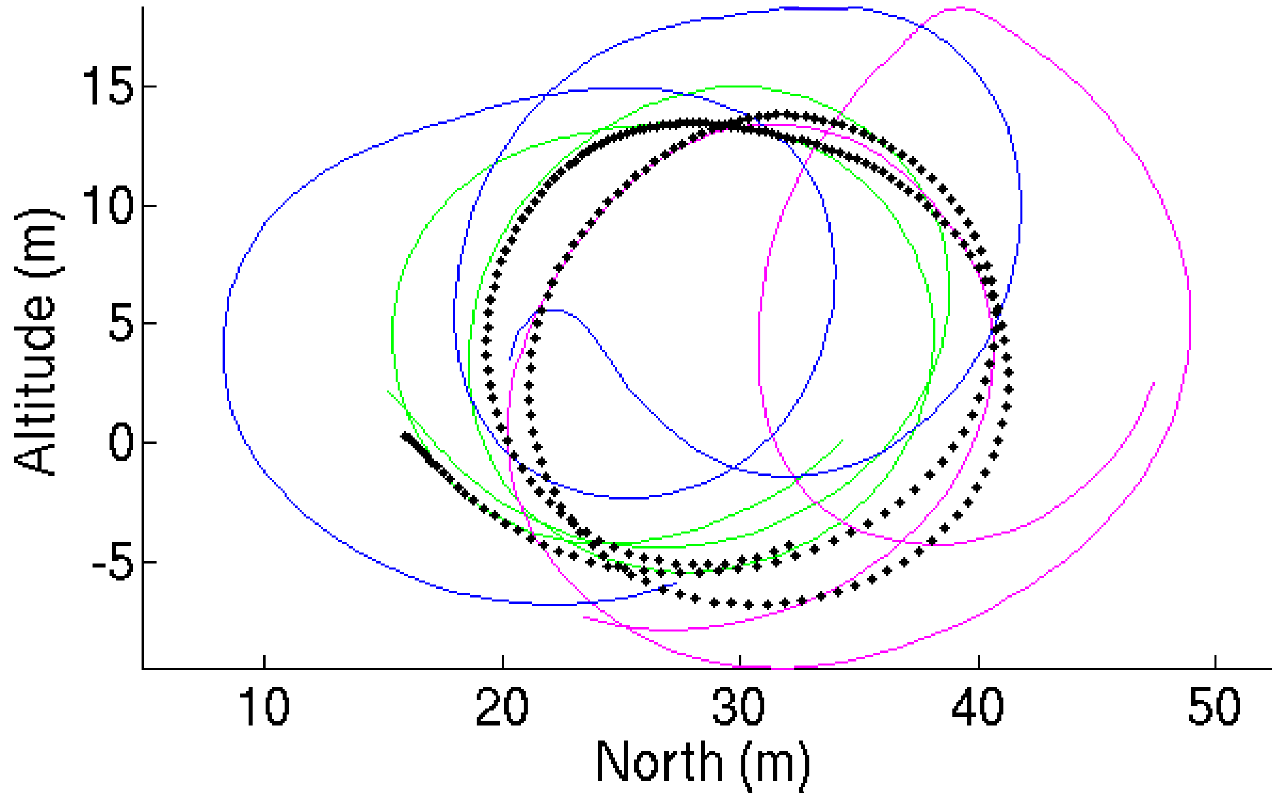
Abbeel, Coates, Ng, IJRR 2010

Results: Time-aligned demonstrations

- White helicopter is inferred “intended” trajectory.



Results: Loops



- Even without prior knowledge, the inferred trajectory is much closer to an ideal loop.

Abbeel, Coates, Ng, IJRR 2010

Learning to perform dynamic maneuvers: outline

- Learning a target trajectory
- **Learning a dynamics model**
- Autonomous flight results
  - Aerobatics
  - Auto-Rotation Landings

Baseline dynamics model

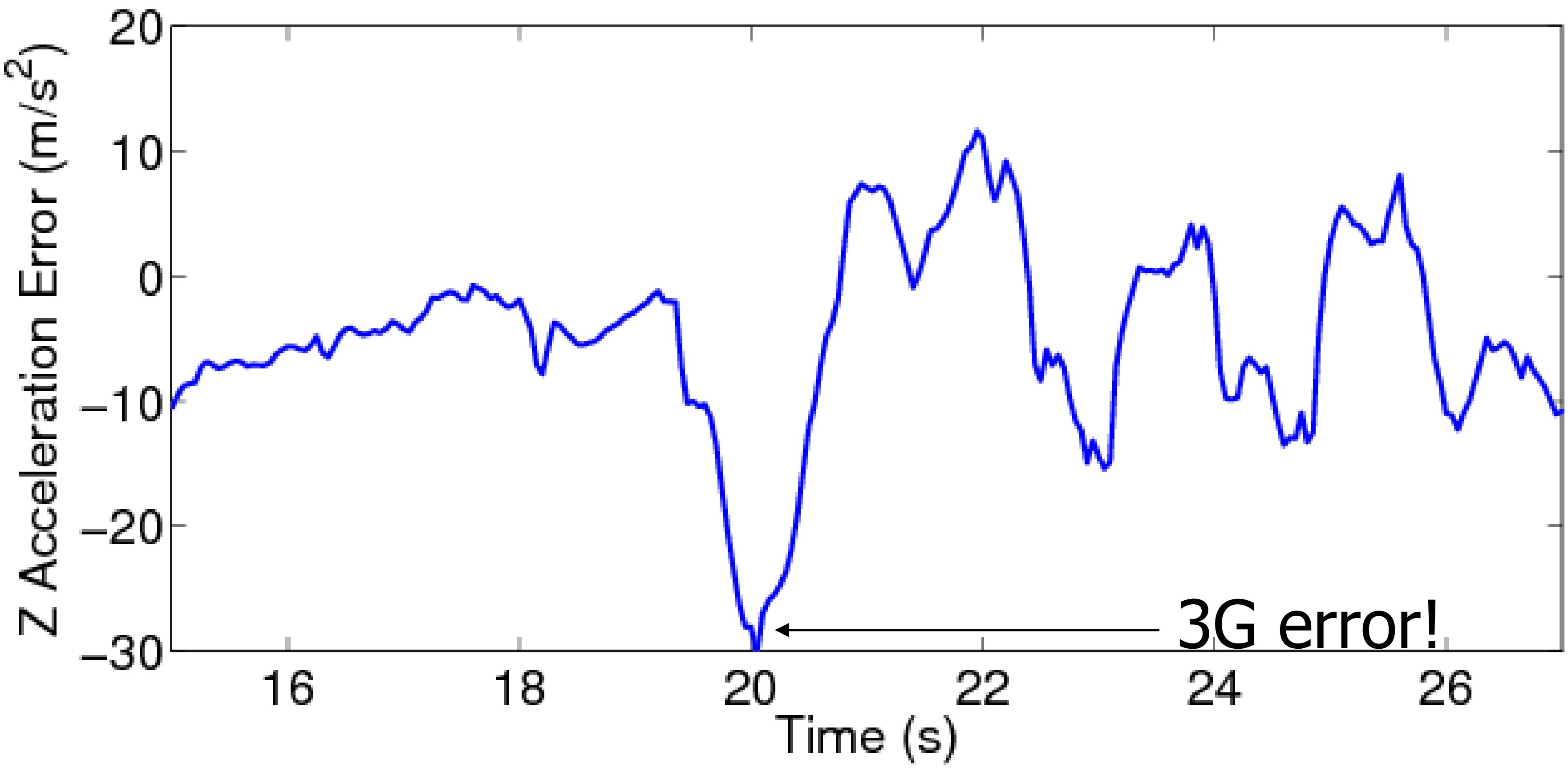
Collect sweeps data to estimate model parameters

$$
\begin{aligned}
\dot{u} &= v * r - w * q + g_u + C_u' * [u], \\
\dot{v} &= w * p - u * r + g_v + C_v' * [v], \\
\dot{w} &= u * q - v * p + g_w + C_w' * [1; \quad w; \quad i_{\text{col}}], \\
\dot{p} &= C_p' * [1; \quad p; \quad i_{\text{lat}}], \\
\dot{q} &= C_q' * [1; \quad q; \quad i_{\text{lon}}], \\
\dot{r} &= C_r' * [1; \quad r; \quad i_{\text{rud}}],
\end{aligned}
$$

(u, v, w)	velocity in helicopter's reference frame
(p, q, r)	angular rates in helicopter's reference frame
(g_u, g_v, g_w)	components of gravity in helicopter's reference frame
$C.$	parameters—estimated from data using least squares

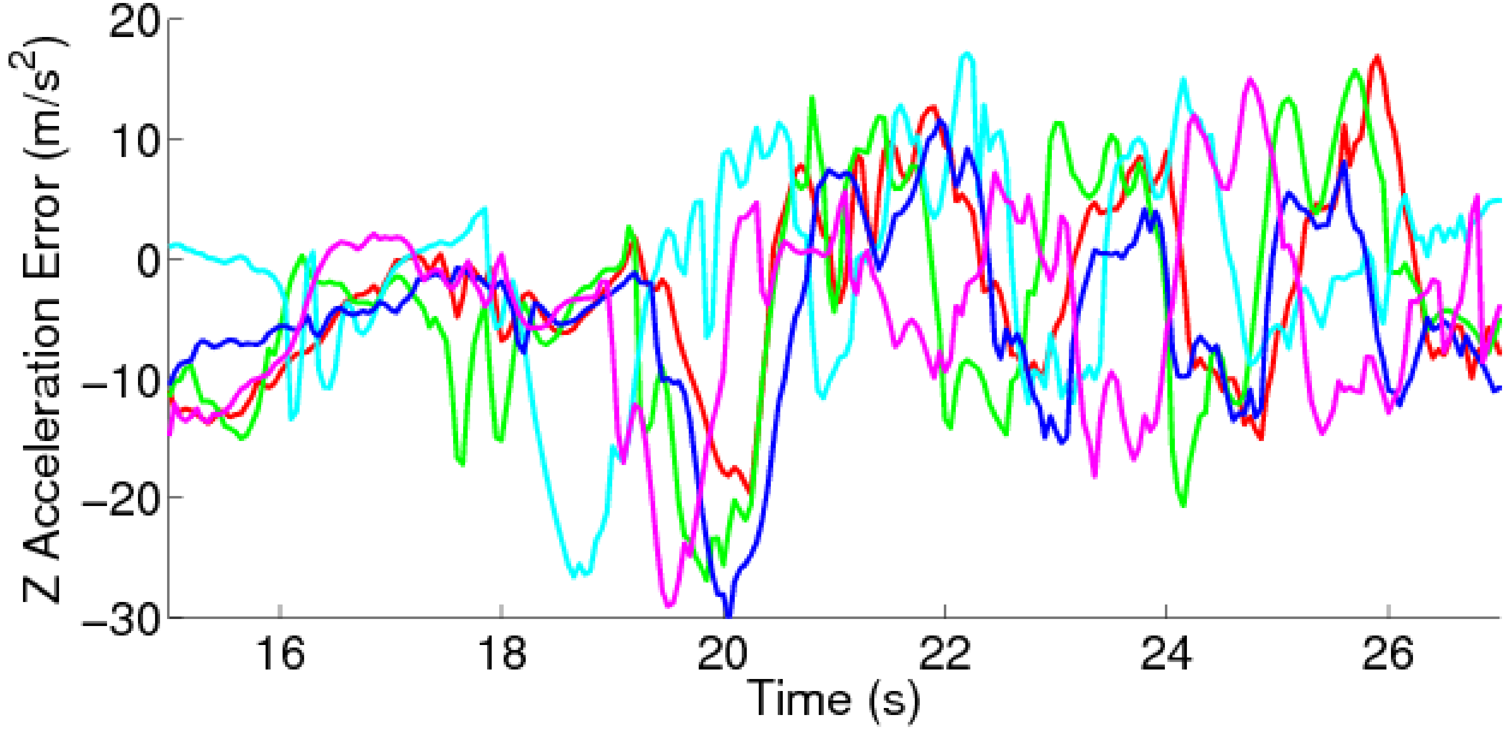
Abbeel, Ganapathi, Ng, NIPS 2006

Empirical evaluation of standard modeling approach



Abbeel, Coates, Ng, IJRR 2010

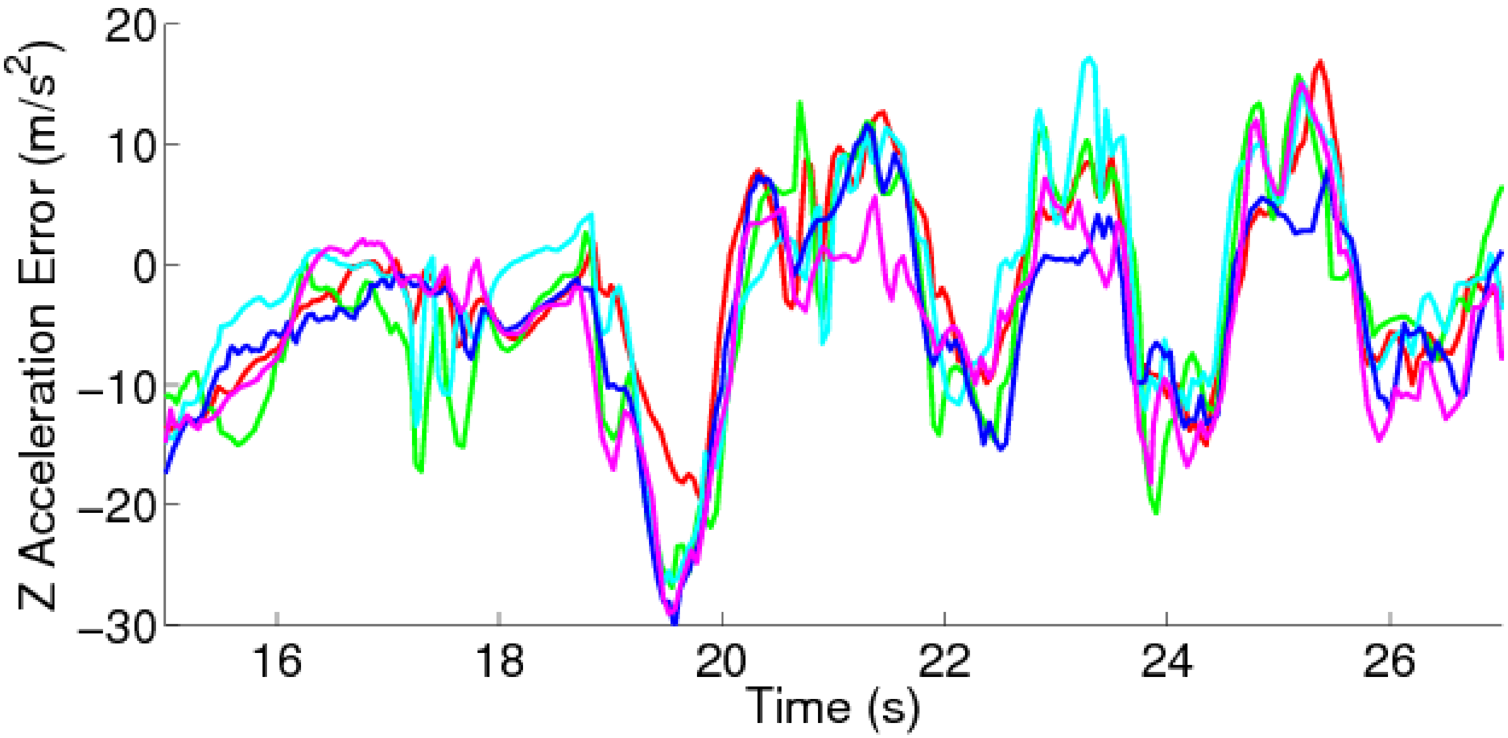
Key observation



- Errors observed in the "baseline" model are clearly consistent after aligning demonstrations.

Abbeel, Coates, Ng, IJRR 2010

Key observation



- Errors observed in the "baseline" model are clearly consistent after aligning demonstrations.

Abbeel, Coates, Ng, IJRR 2010

Key observation

- If we fly the same trajectory repeatedly, errors are consistent over time once we align the data.
  - There are many unmodeled variables that we can’t expect our model to capture accurately.
    - Air (!), actuator delays, etc.
  - If we fly the same trajectory repeatedly, the hidden variables tend to be the same each time.

~ muscle memory for human pilots

Abbeel, Coates, Ng, IJRR 2010

Trajectory-specific local models

- Learn locally-weighted model from aligned demonstration data
  - Since data is aligned in time, we can weight by *time* to exploit repeatability of unmodeled variables.
  - For model at time t: $W(t') = \exp(- (t - t')^2 / \sigma^2)$
  - Obtain a model for each time t into the maneuver by running weighted regression for each time t

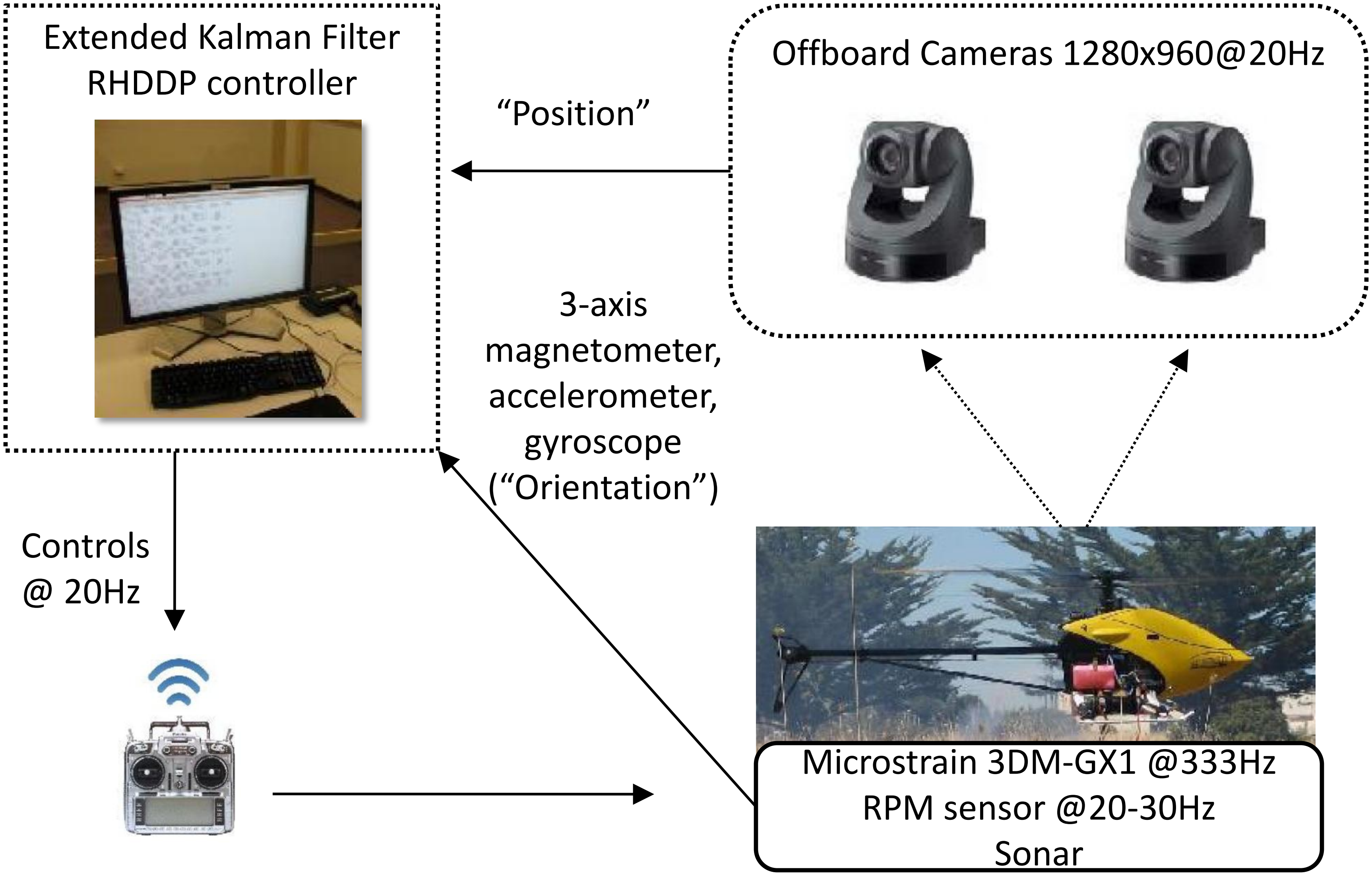
Abbeel, Coates, Ng, IJRR 2010

Learning to perform dynamic maneuvers: outline

- Learning a target trajectory
- Learning a dynamics model
- **Autonomous flight results**
  - Aerobatics
  - Auto-Rotation Landings

Abbeel, Coates, Ng, IJRR 2010

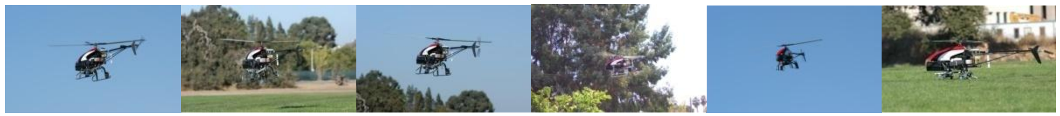
Experimental Setup



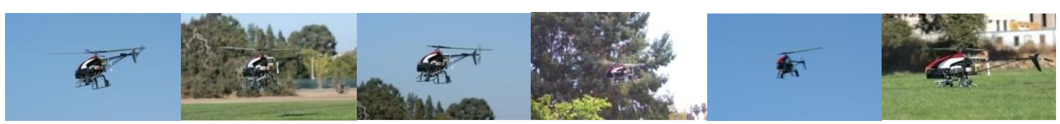
Abbeel, Coates, Quigley, Ng, NIPS 2007

Experimental procedure

1. Collect sweeps to build a baseline dynamics model
2. Our expert pilot demonstrates the airshow several times.



3. Learn a target trajectory.
4. Learn a dynamics model.
5. Find the optimal control policy for learned target and dynamics model.
6. Autonomously fly the airshow



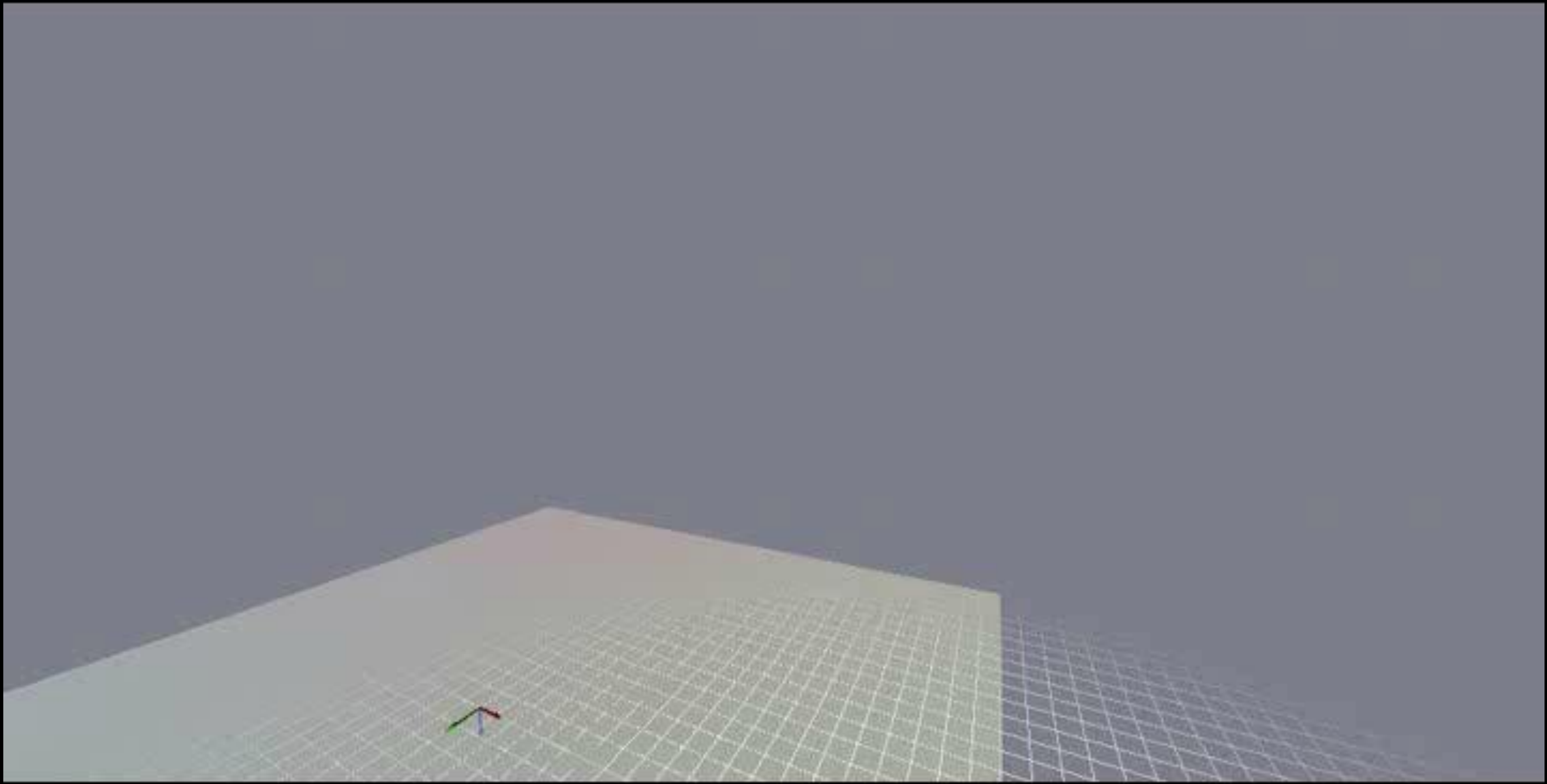
7. Learn an improved dynamics model. Go back to step 4.

→ **Learn to fly new maneuvers in < 1hour.**

Abbeel, Coates, Ng, IJRR 2010

Results: Autonomous airshow

Results: Flight accuracy

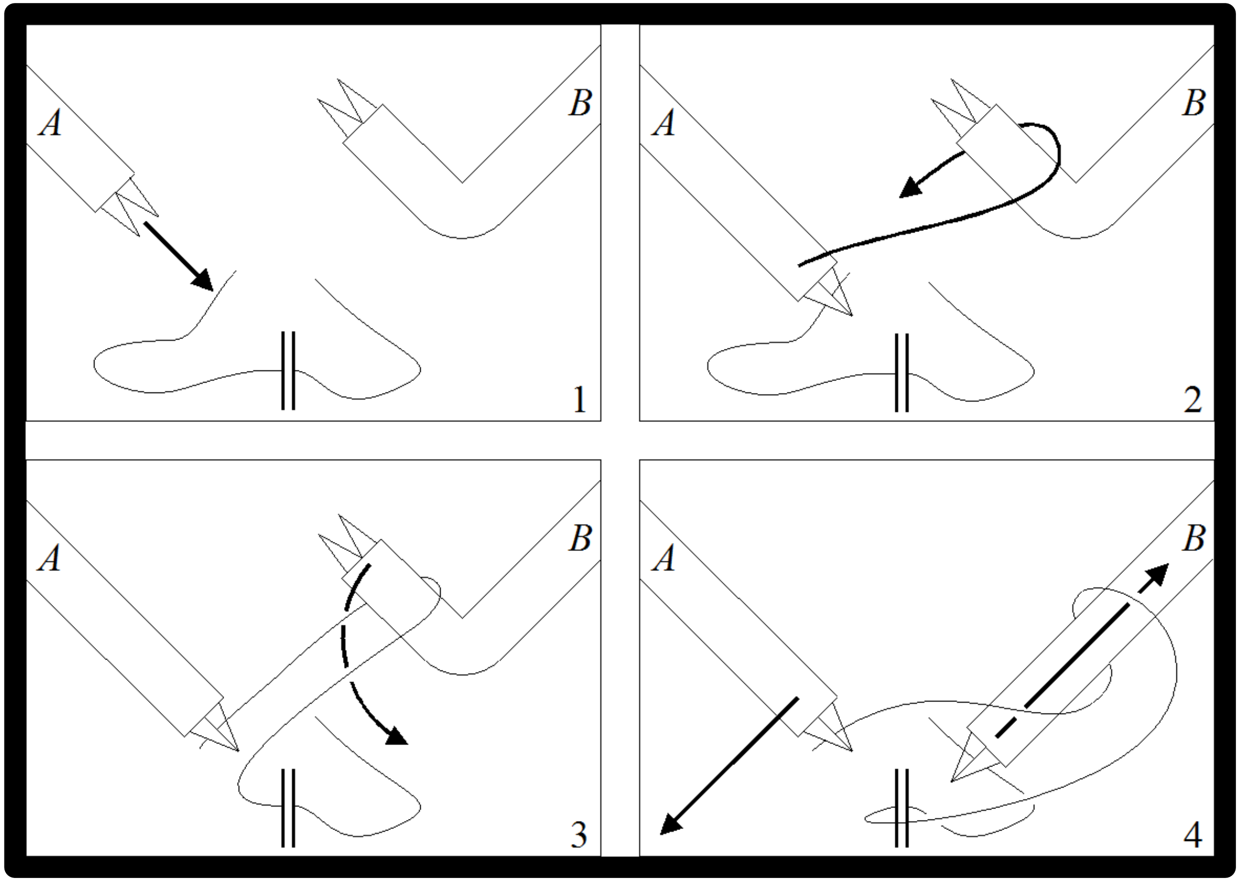


Thus far

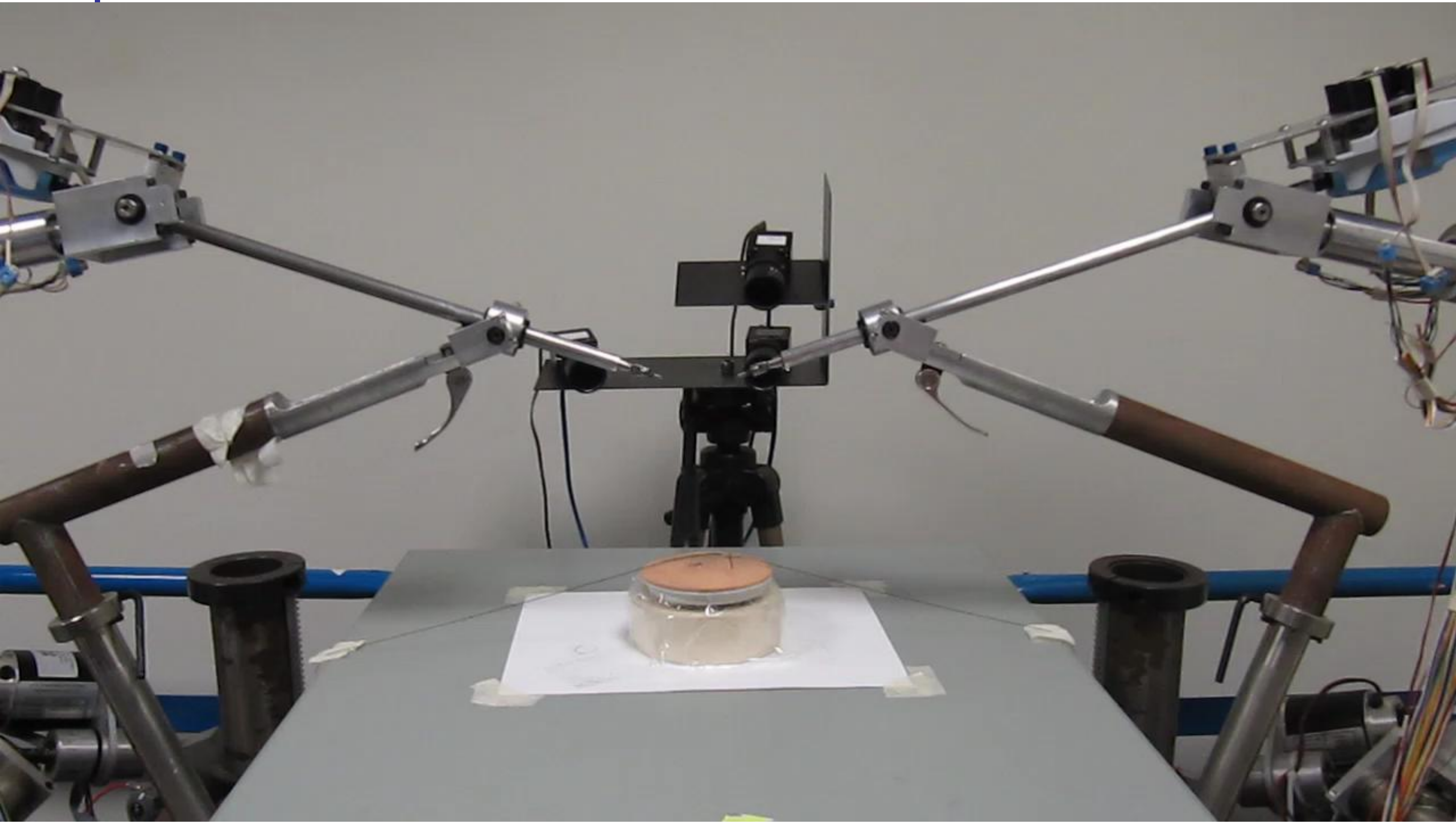
- Apprenticeship learning
  - Learn to perform task from expert demonstrations
  - Enabled by far most advanced helicopter aerobatics

- How about:

Surgical knot tie



Surgical knot tie



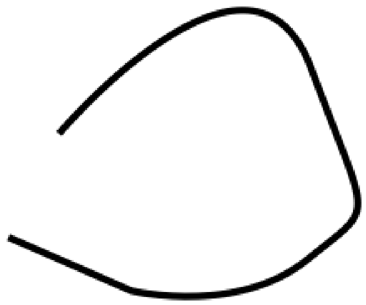
van den Berg, Miller, Duckworth, Humphrey, Wan, Fu, Goldberg, Abbeel, ICRA 2010

Surgical knot tie

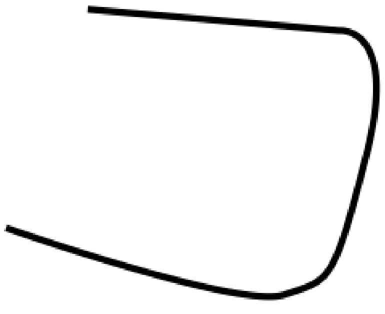
- Open loop
- If careful about initial conditions
  - 50% success rate

Generalizing Trajectories

- The problem
  - Human demonstrated knot-tie in this rope

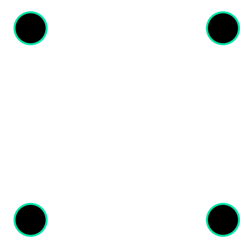


  - Robot has to tie a knot in this rope

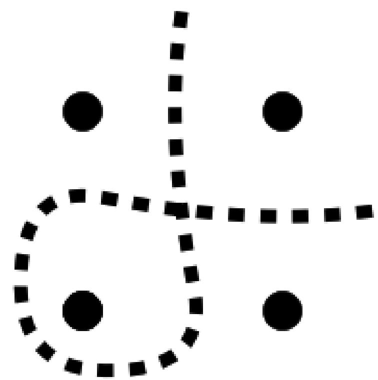


Cartoon Problem Setting

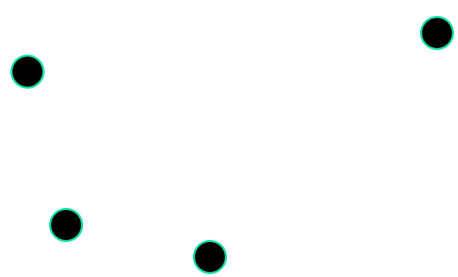
Train situation:



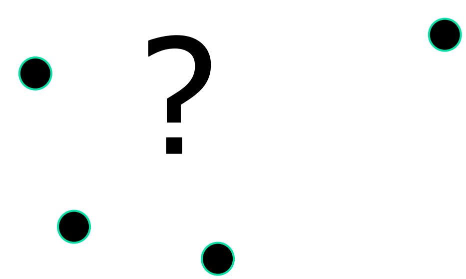
Verb demonstration: --- trajectory



Test situation:



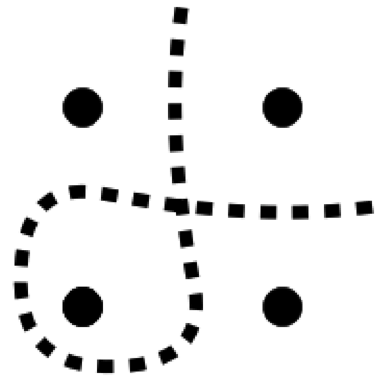
How to perform verb here?

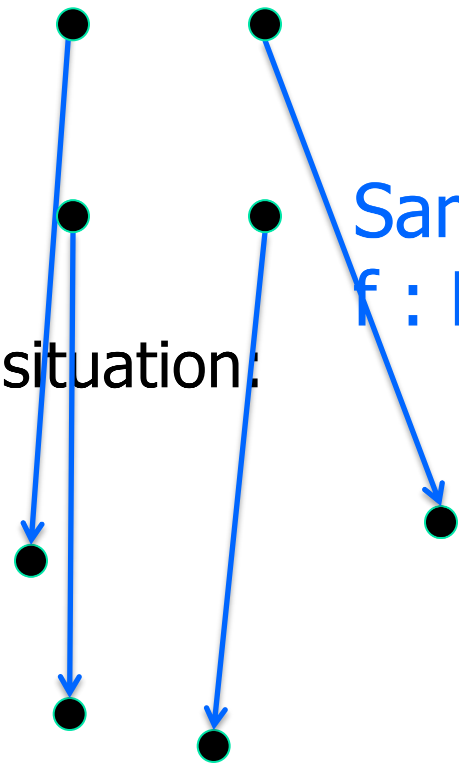


Cartoon Problem Setting

Train situation:

Verb demonstration: --- trajectory

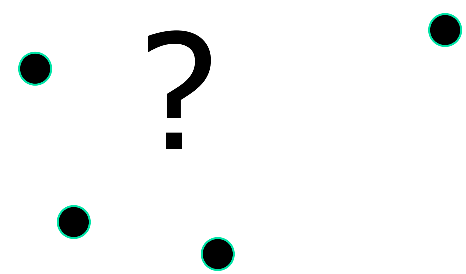




Samples of $f : R^2 \rightarrow R^2$

Test situation:

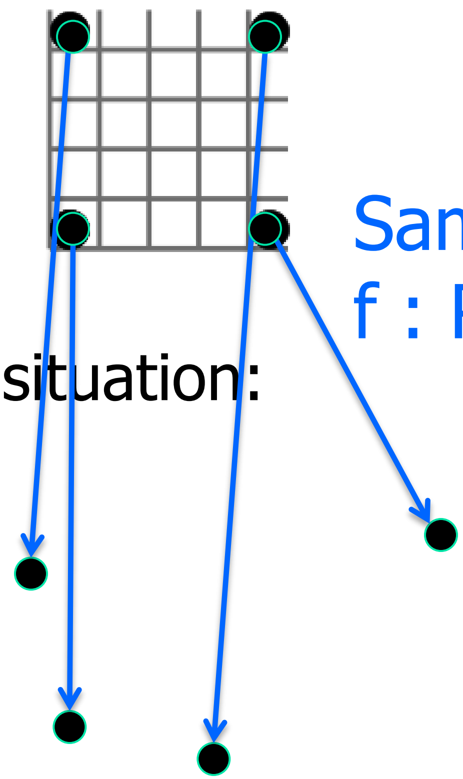
How to perform verb here?

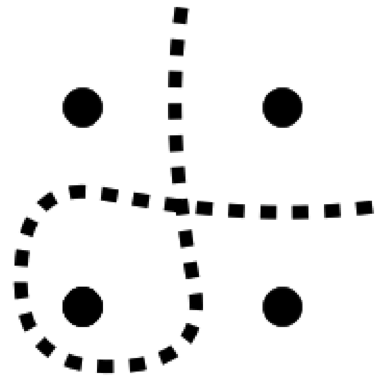


Cartoon Problem Setting

Train situation:

Verb demonstration: --- trajectory

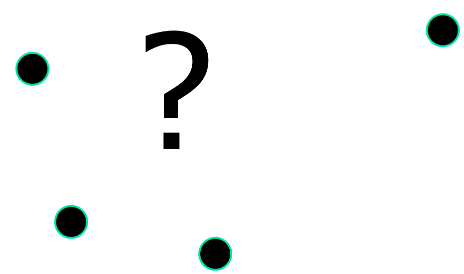




Samples of
$f : R^2 \rightarrow R^2$

Test situation:

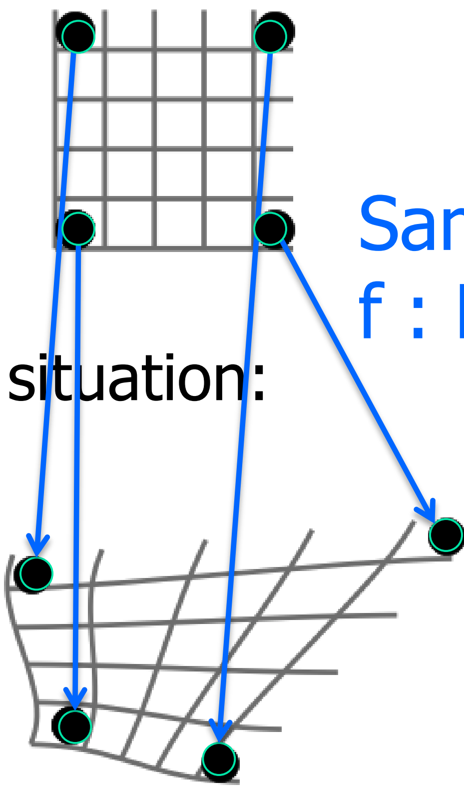
How to perform verb here?

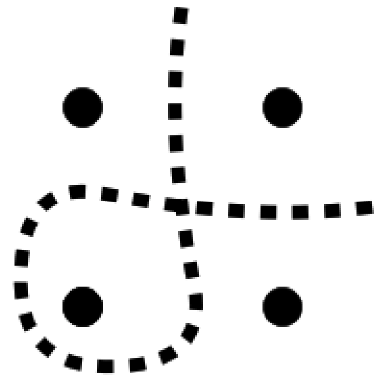


Cartoon Problem Setting

Train situation:

Verb demonstration: --- trajectory

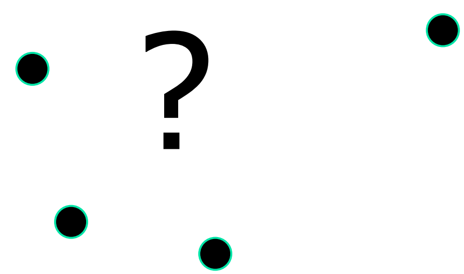




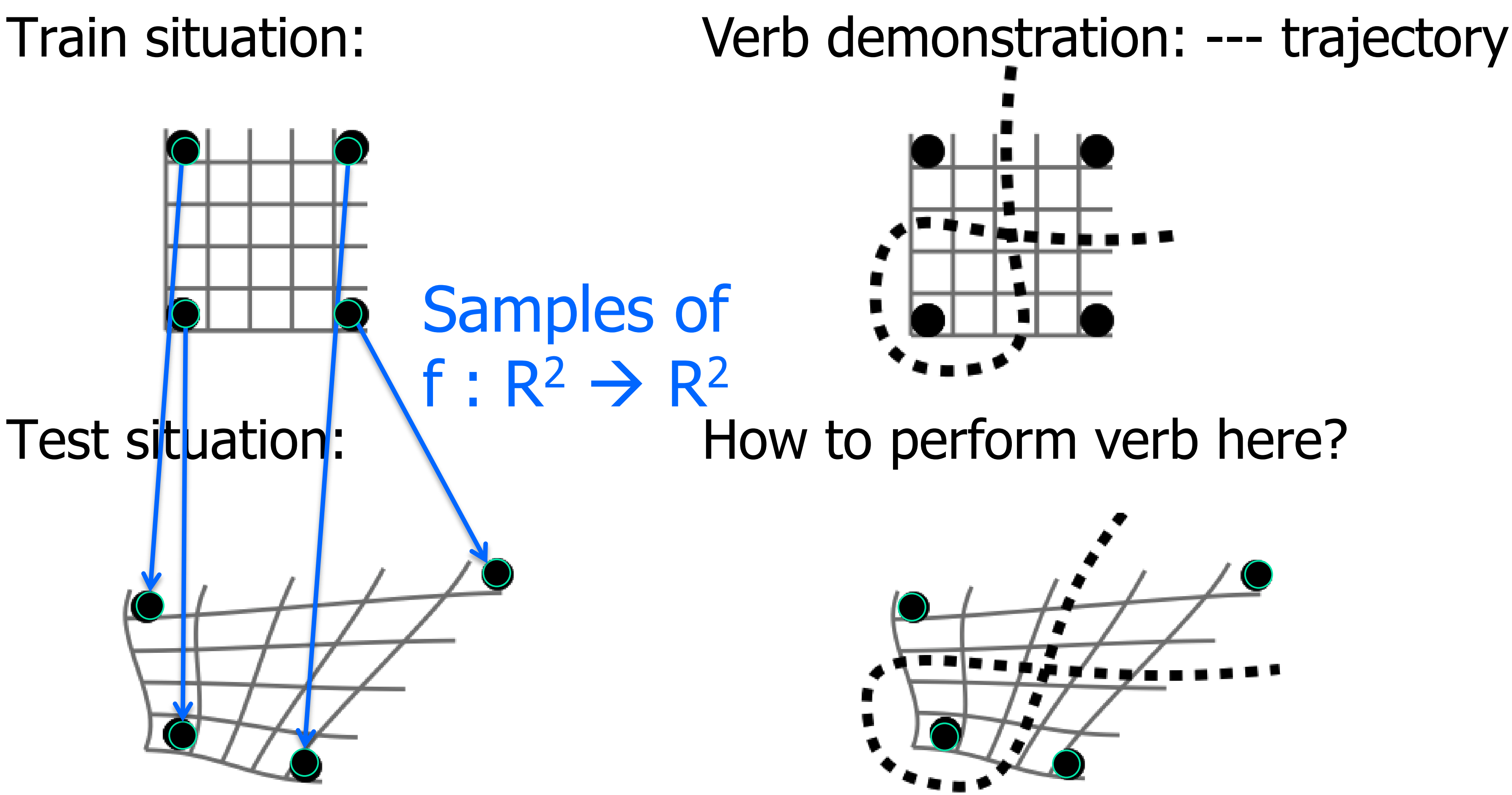
Samples of
$f : R^2 \rightarrow R^2$

Test situation:

How to perform verb here?



Cartoon Problem Setting



Learning f : R³ → R³ from samples

$$\min_{f \in \{\mathbb{R}^3 \to \mathbb{R}^3\}} \int_{x \in \mathbb{R}^3} \|D^2 f\|^2_{\mathrm{Frob}}(x)\,dx$$

$$\text{s.t.} \quad f(x^{(i)}_{\mathrm{train}}) = x^{(i)}_{\mathrm{test}} \quad \forall i \in 1, \ldots, m$$

- Observations
  - Translations, rotations and scaling are FREE
  - Can be solved efficiently manipulating matrices of size of number of examples

Learning f : R³ → R³ from samples

$$\min_{f \in \{\mathbb{R}^3 \to \mathbb{R}^3\}} \quad \int_{x \in \mathbb{R}^3} \|D^2 f\|^2_{\text{Frob}}(x)dx$$

$$\text{s.t.} \quad f(x^{(i)}_{\text{train}}) = x^{(i)}_{\text{test}} \quad \forall i \in 1, \ldots, m$$

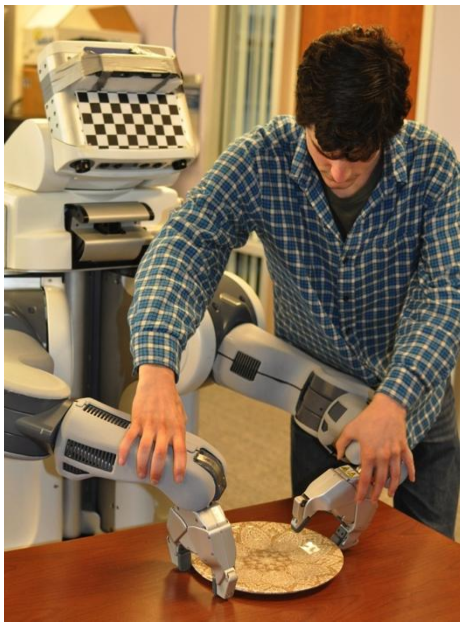
- Solution has form:

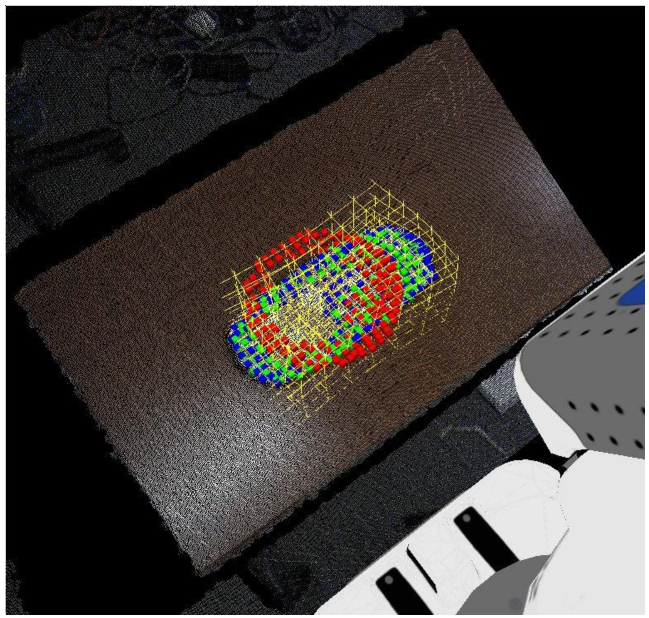
$$f(x) = \sum_{i=1}^{m} a_i K(x_i, x) + b^\top x + c,$$

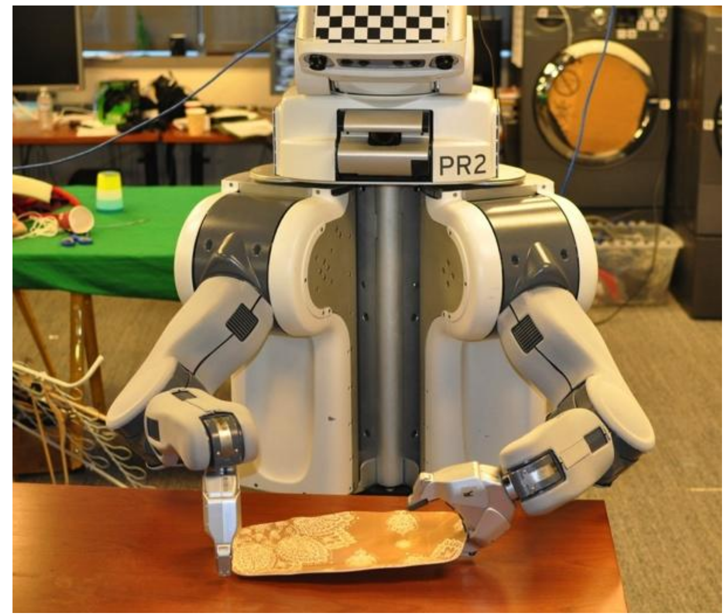
$$K(x,y) = \begin{cases} c_0 r^{4-d} \ln r, & d = 2 \text{ or } d = 4 \\ c_1 r^{4-d}, & \text{otherwise} \end{cases} \qquad \text{with} \quad r = \|x - y\|_2.$$

Wahba, Spline models for observational data. Philadelphia: Society for Industrial and Applied Mathematics. 1990.
Evgeniou, Pontil, Poggio, Regularization Networks and Support Vector Machines. Advances in Computational Mathematics. 2000
Hastie, Tibshirani, Friedman, Elements of Statistical Learning, Chapter 5. 2008.

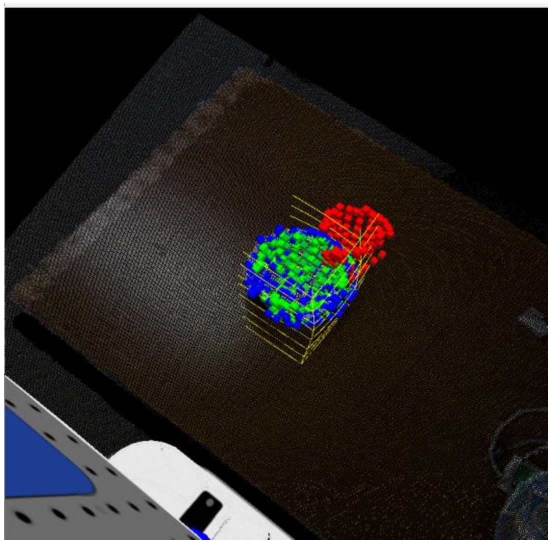
Experiments: Plate Pick-Up

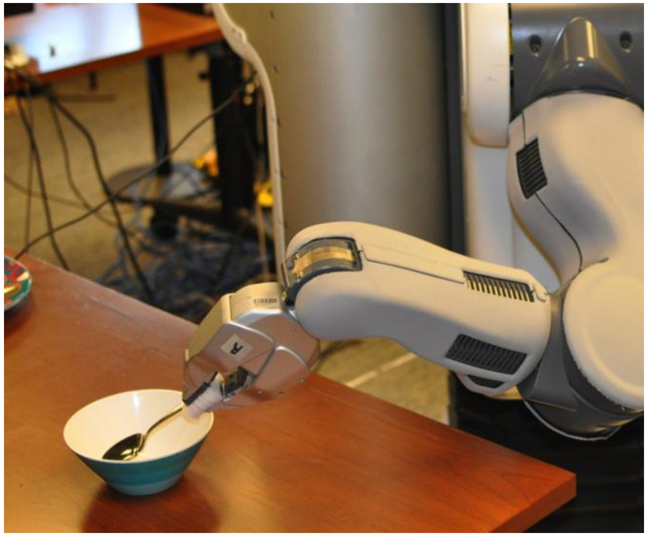




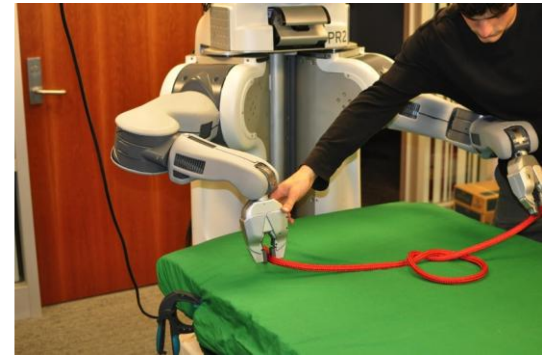


Experiments: Scooping

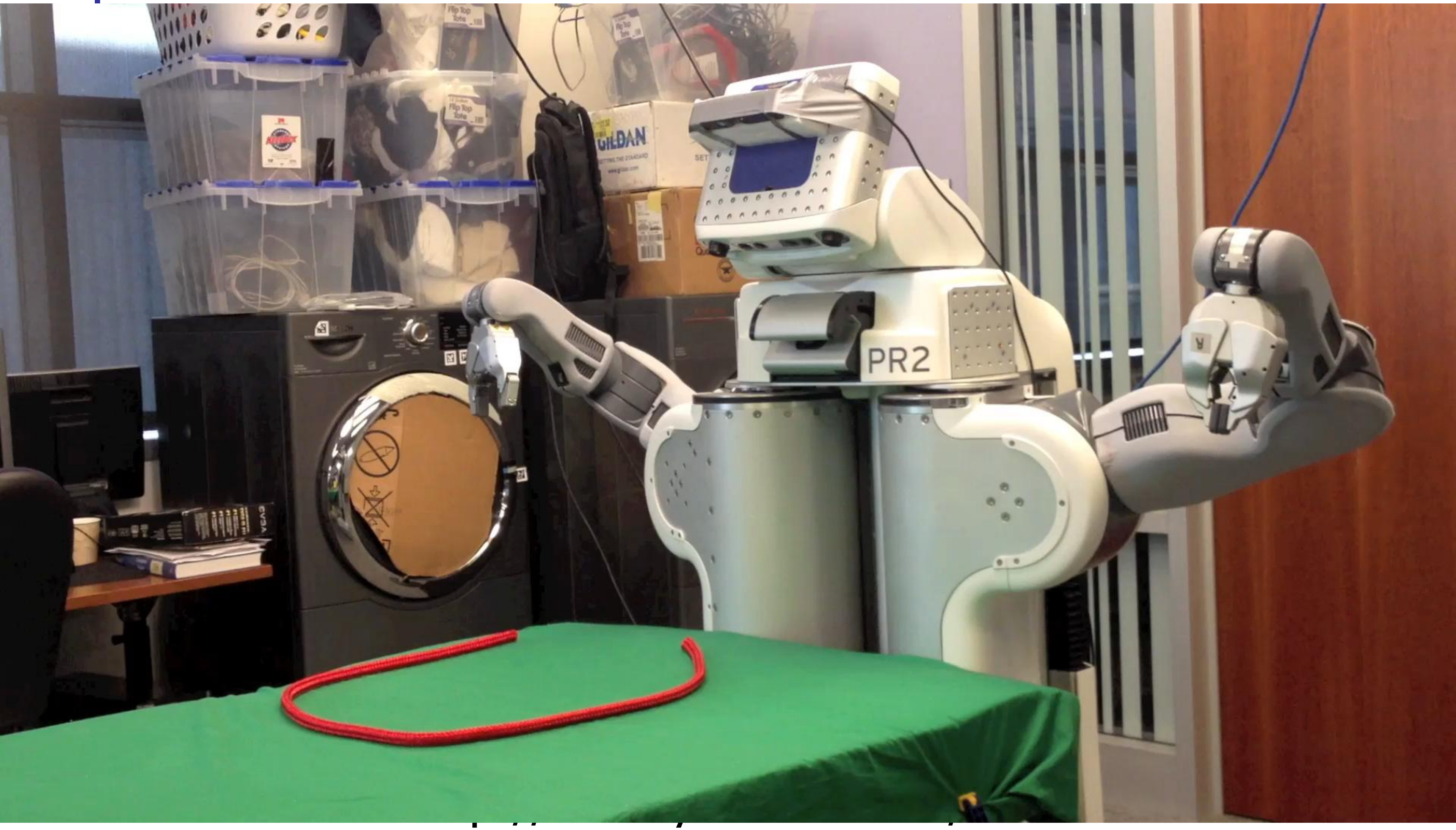




Experiment: Knot-Tie



Autonomous tying of a knot for a previously unseen situation

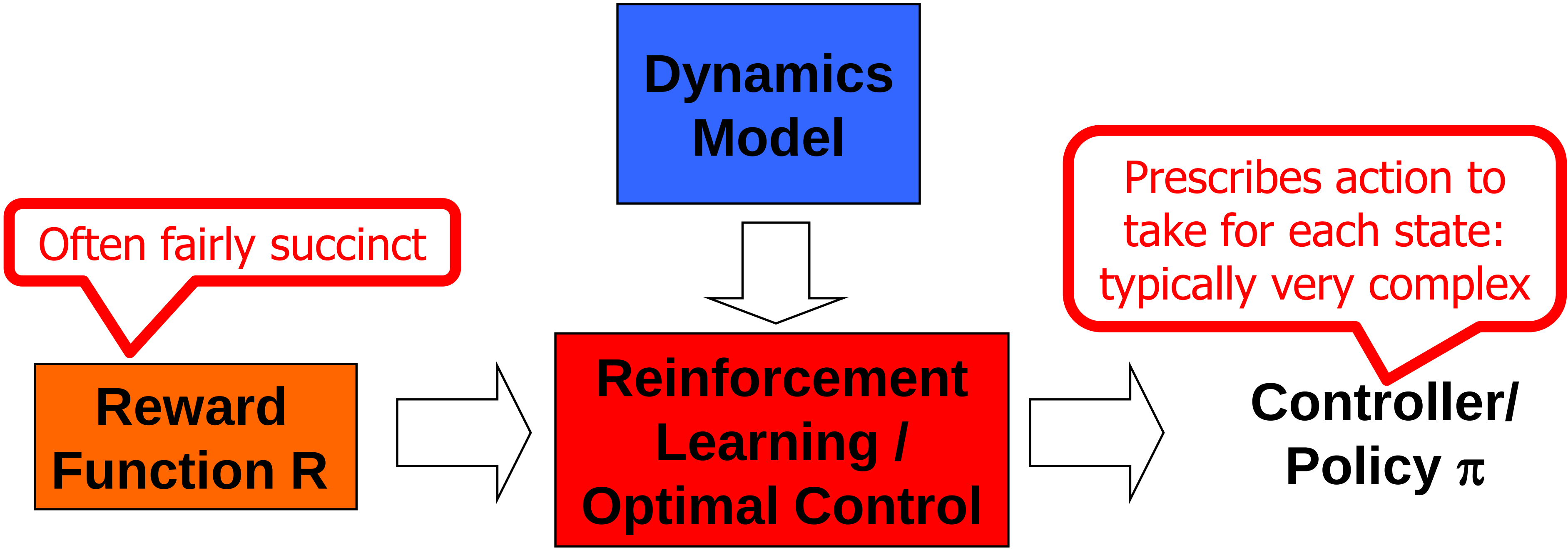


Schulman & Abbeel, TBD 2012

Outline

- Apprenticeship learning
  - Learning to perform dynamic maneuvers
    - Helicopter
  - Learning to perform tasks that require generalization
    - Rope manipulation
- Brief highlights of three other projects on machine learning for robotics
  - Inverse optimal control
    - Quadruped locomotion
  - Safe exploration
    - Martian surface exploration
  - Perception
    - Clothes manipulation

Problem Structure



$$\arg\max_{\pi} \mathrm{E}[\sum_t R(s_t)|\pi]$$

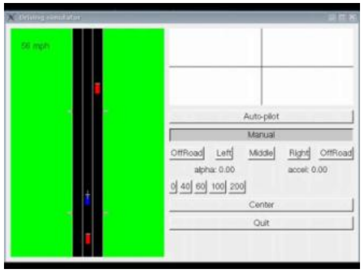
E.g., $R^* = w_1^* 1\{''\,in\ right\ lane''\} + w_2^* 1\{''\,safe\ distance''\}$

Inverse RL History

- 1964, Kalman posed the inverse optimal control problem and solved it in the 1D input case
- 1994, Boyd+al.: a linear matrix inequality (LMI) characterization for the general linear quadratic setting
- 2000, Ng and Russell: first MDP formulation, reward function ambiguity pointed out and a few solutions suggested
- 2004, Abbeel and Ng: inverse RL for apprenticeship learning---reward feature matching

Inverse RL Examples

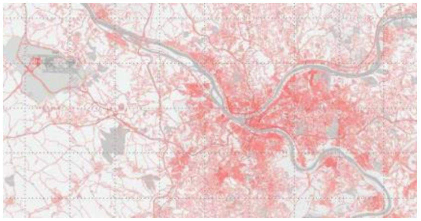
- Simulated highway driving
  - Abbeel and Ng, ICML 2004,
  - Syed and Schapire, NIPS 2007



- Aerial imagery based navigation
  - Ratliff, Bagnell and Zinkevich, ICML 2006

- Parking lot navigation
  - Abbeel, Dolgov, Ng and Thrun, IROS 2008

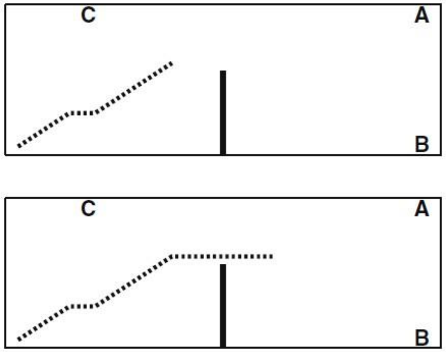
- Urban navigation
  - Ziebart, Maas, Bagnell and Dey, AAAI 2008



Inverse RL Examples (ctd)

- Human path planning
  - Mombaur, Truong and Laumond, AURO 2009

- Human goal inference
  - Baker, Saxe and Tenenbaum, Cognition 2009



- Quadruped locomotion
  - Ratliff, Bradley, Bagnell and Chestnutt, NIPS 2007
  - Kolter, Abbeel and Ng, NIPS 2008

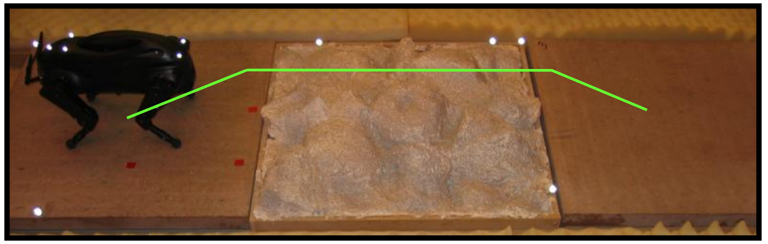
Quadruped

- Reward function trades off 25 features.

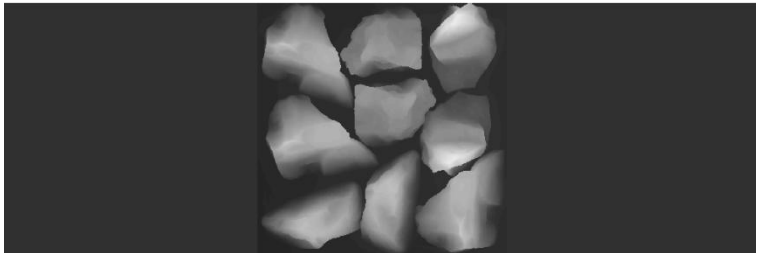
Hierarchical max margin [Kolter, Abbeel & Ng, 2008]

Experimental setup

- Demonstrate path across the "training terrain"



- Run our apprenticeship learning algorithm to find the reward function
- Receive "testing terrain"---height map.



- Find the optimal policy with respect to the *learned reward function* for crossing the testing terrain.

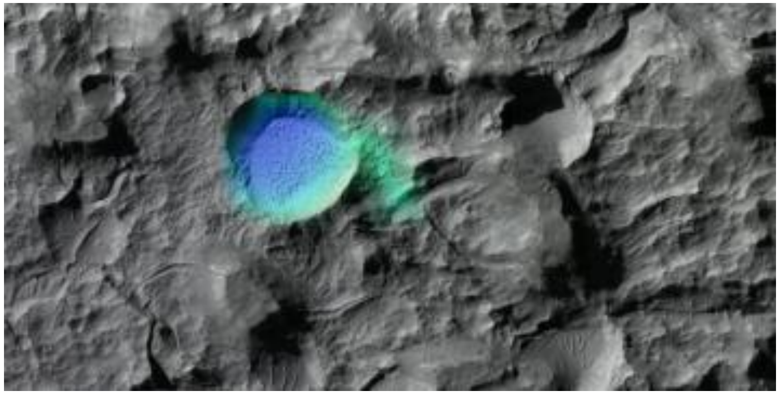
Hierarchical max margin [Kolter, Abbeel & Ng, 2008]

Without learning

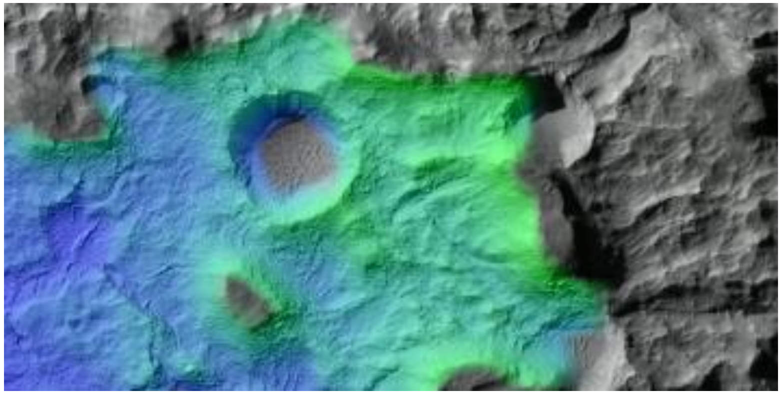
With learned reward function

Safe exploration

- Existing exploration methods performance:



- Safe exploration:



Moldovan & Abbeel, ICML 2012

Safe exploration --- towards:

Exploration levels in this video (still) human guided

Safe exploration – Key idea

$$\max_{\pi_o,\pi_r} \quad \mathrm{E}^{\gamma p}_{s_0,\pi_o} \sum_t \left(r_{S_t,A_t} + \xi^{\beta}_{S_t,A_t} \right)$$

find policy that optimizes expected exploration bonus

$$\text{s.t.} \quad \mathrm{E}_{\beta} \mathrm{E}^{P}_{s_0,\pi_o} \left[\mathrm{E}^{P}_{S_T,\pi_r} [B_{s_0}] \right] \geq \delta.$$

while if recalled at any given time, possible to return "home" with high probability

- Constraint is NP-hard to work with per distribution over all possible worlds
- We have derived an efficient (conservative) approximation

Perception and clothes manipulation

Maitin-Shepard, Cusumano-Towner, Lei & Abbeel, ICRA 2010; Cusumano-Towner, Singh, Miller & Abbeel, ICRA 2011; Miller, van den Berg, Fritz, Darrell, Goldberg & Abbeel, IJRR 2011; Wang, Miller, Fritz, Darrell & Abbeel, ICRA 2011

Conclusion

- Apprenticeship learning
  - Learning to perform dynamic maneuvers
    - Helicopter
  - Learning to perform tasks that require generalization
    - Rope manipulation
- Brief highlights of three other projects on machine learning for robotics
  - Inverse optimal control
    - Quadruped locomotion
  - Safe exploration
    - Martian surface exploration
  - Perception
    - Clothes manipulation

Thank you.