Learning to Select Expert Demonstrations for Deformable Object Manipulation

Dylan Hadfield-Menell, Alex Lee, Sandy Huang, Eric Tzeng, Pieter Abbeel

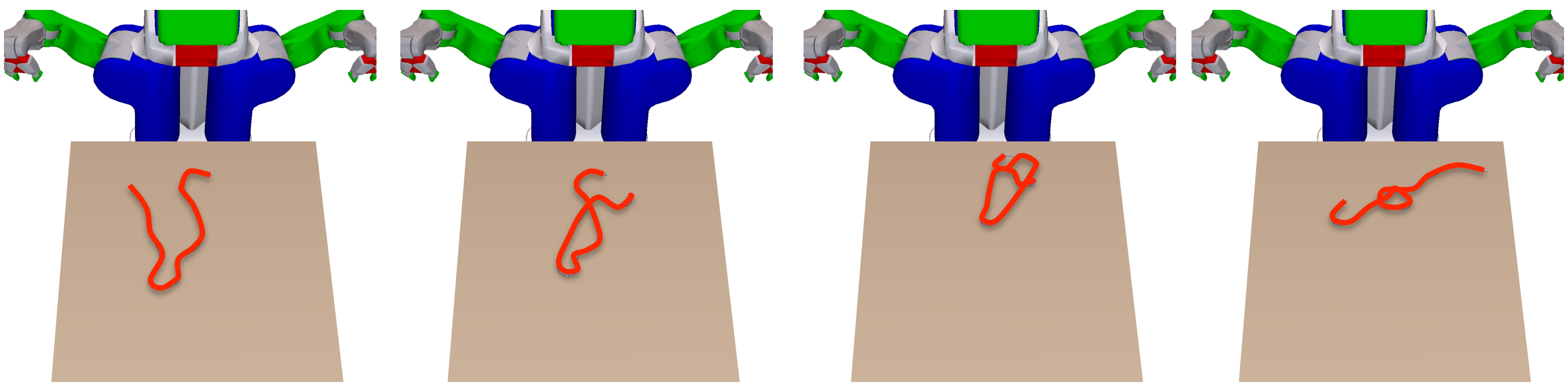
Workshop on Information-Based Grasp and Manipulation Planning
July 13, 2014
RSS 2014

Vision

- We’d like robots to be able to do lots of things
- Need deformable object manipulation
- Ease of programming

Deformable Object Manipulation

- High-Dimensional, Continuous State and Action Spaces
- Long Time Horizons
- Complex Dynamics
- Example: Knot-Tying with the PR2



$\mathcal{S} \subset \mathbb{R}^{230}$ $\mathcal{A} \subset \mathbb{R}^{14}$ $H \approx 100$

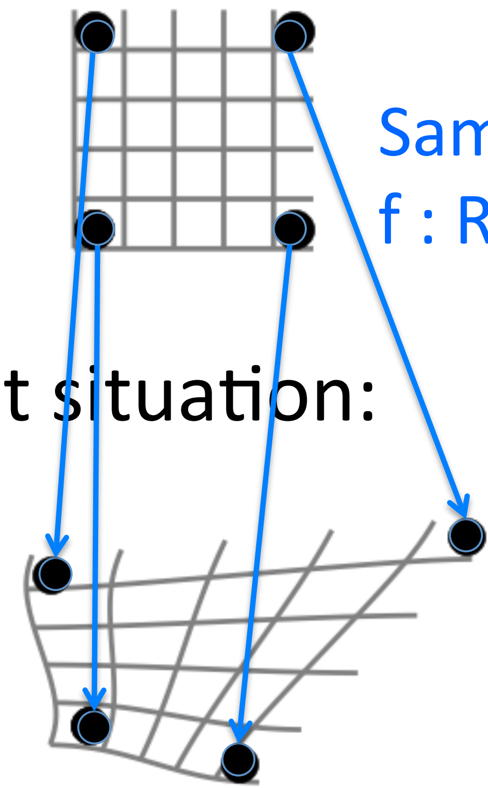
Trajectory Transfer

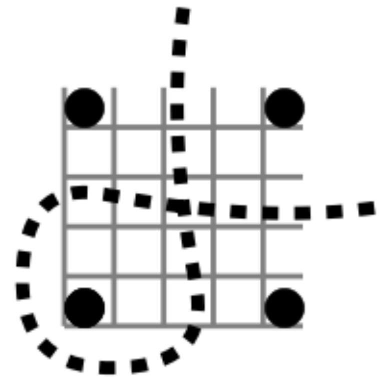
- Planning for deformable object manipulation is a serious challenge
  - Substantial improvements in existing methods before tractability
- Solution: Don't plan!
  - modify demonstration trajectories to fit the current situation

Trajectory Transfer: Cartoon Problem Setting

Train situation:

Trajectory demonstration

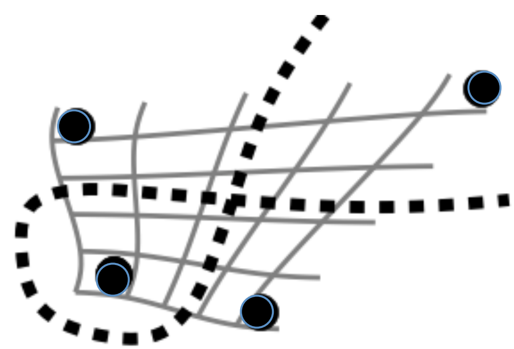




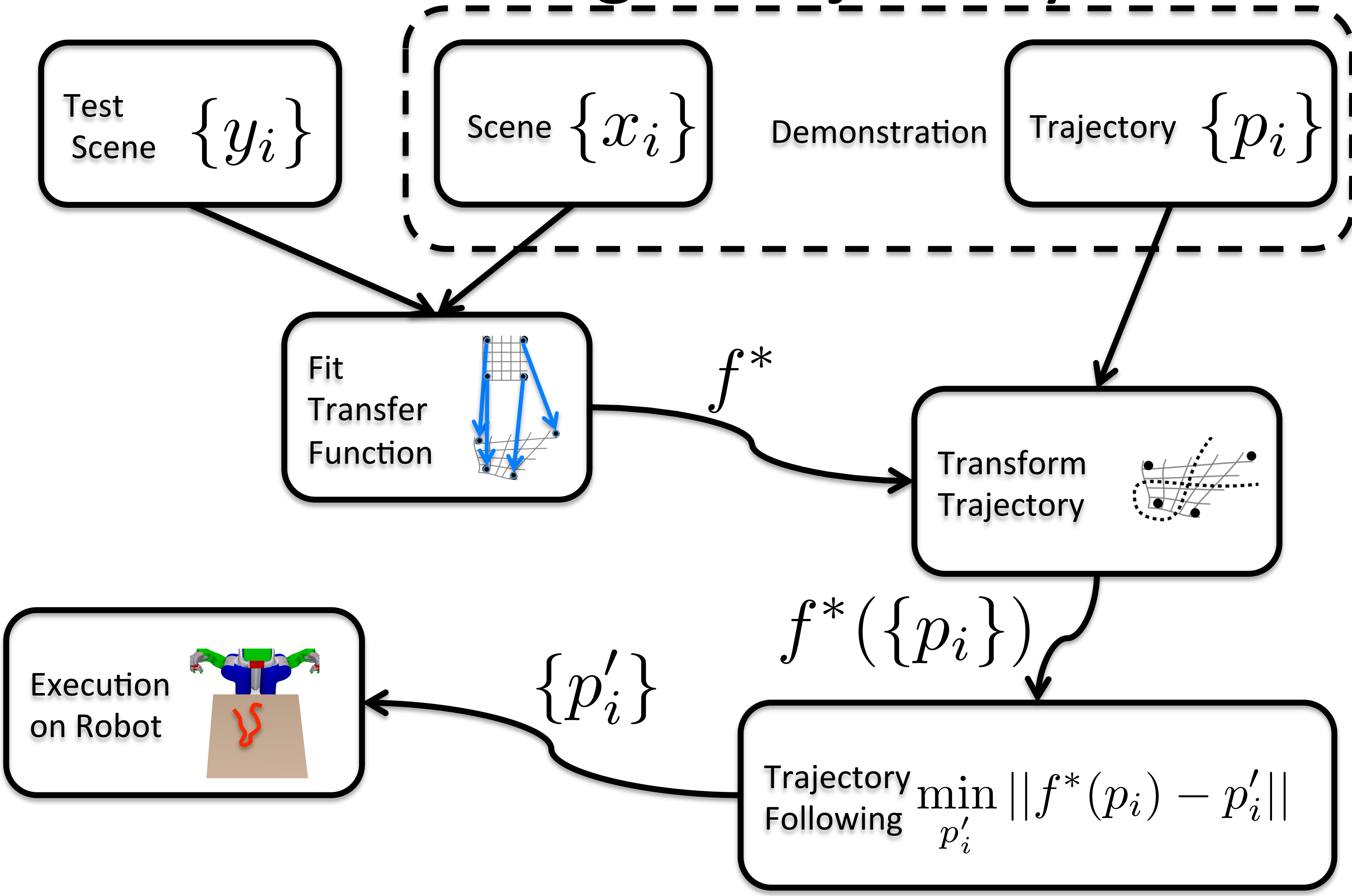
Samples of
f : $R^3 \rightarrow R^3$

Test situation:

What trajectory here?

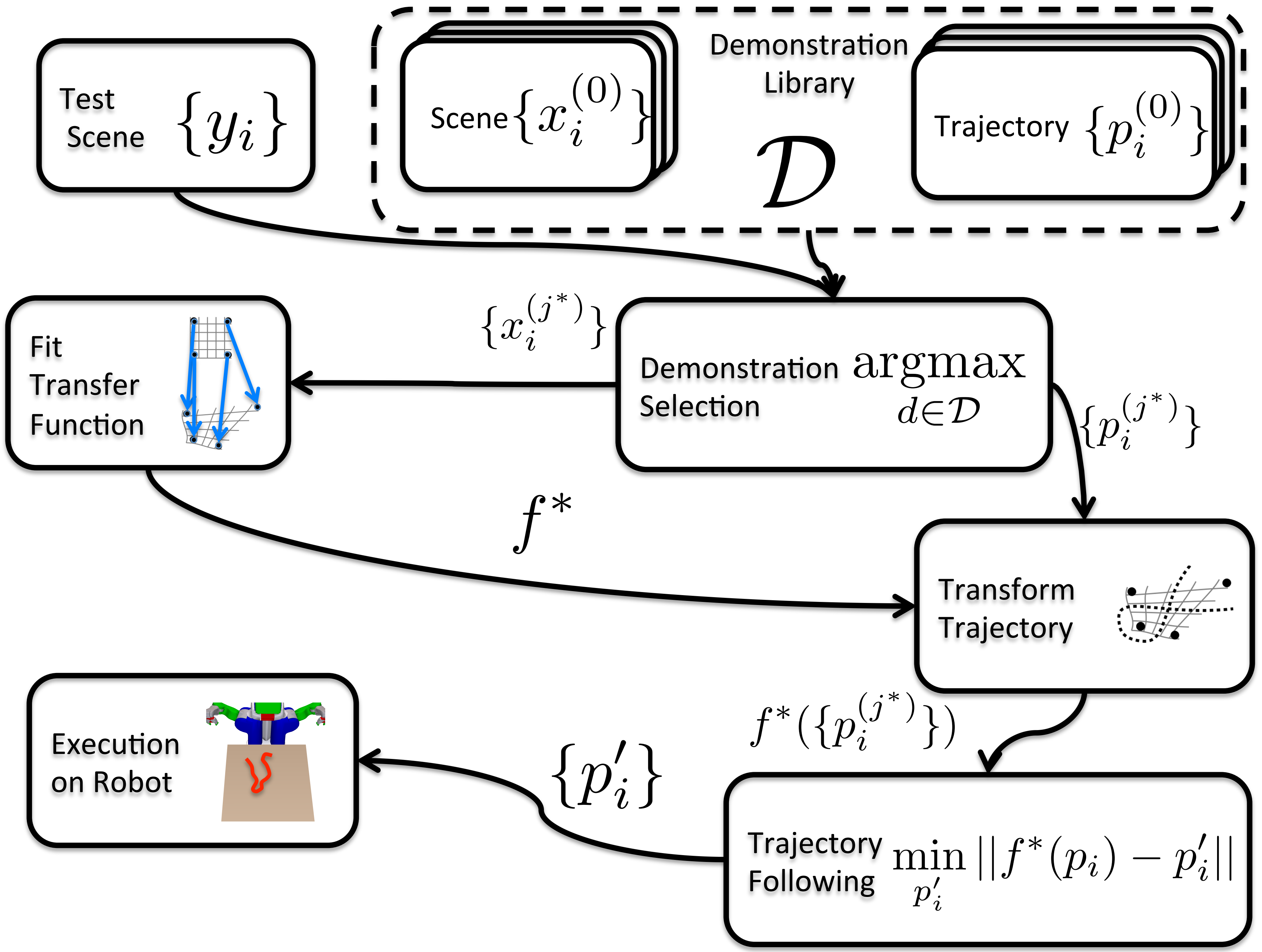


Transferring a Trajectory



Example Trajectory Transfer

- J. Schulman, J. Ho, C. Lee, P. Abbeel. 'Generalization of robotic manipulation through the use of non-rigid registration.' ISRR 2013.
- J. Schulman, A. Gupta, S. Venkatesan, M. Taylor-Frederick, P. Abbeel. 'A case study of trajectory transfer through non-rigid registration for a simplified suturing scenario.' IROS 2013.
- A. Lee, S. Huang, D. Hadfield-Menell, E. Tzeng, P. Abbeel. 'Unifying scene registration and trajectory optimization for learning from demonstrations with application to manipulation of deformable objects.' IROS 2014

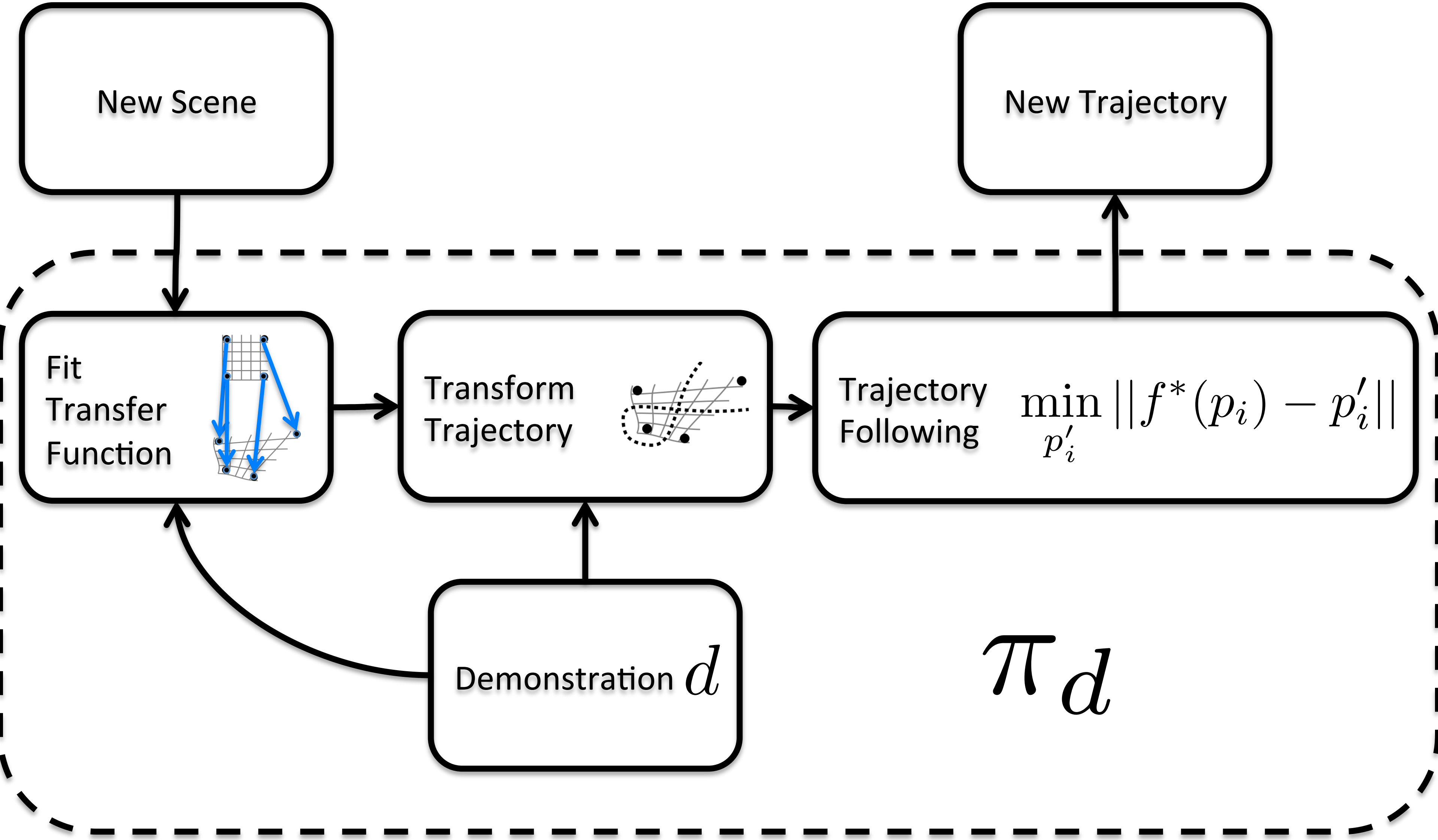


Test Scene $\{y_i\}$
Scene $\{x_i^{(0)}\}$
Demonstration Library
$\mathcal{D}$
Trajectory $\{p_i^{(0)}\}$
Fit Transfer Function
$\{x_i^{(j^*)}\}$
Demonstration Selection $\arg\max_{d \in \mathcal{D}}$
$\{p_i^{(j^*)}\}$
f^*
Transform Trajectory
$f^*(\{p_i^{(j^*)}\})$
Execution on Robot
$\{p'_i\}$
Trajectory Following $\min_{p'_i} ||f^*(p_i) - p'_i||$

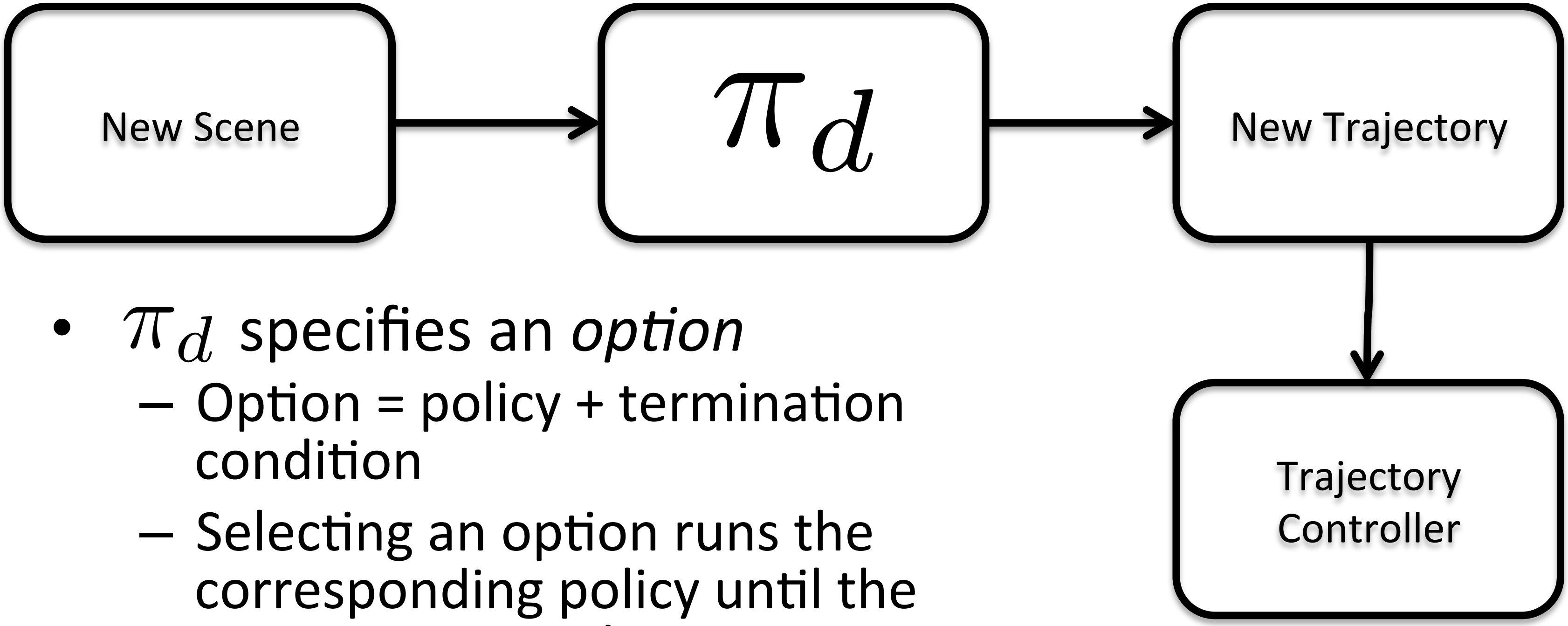
How do we select the ‘best’ Demonstration?

- Different demonstrations may have very different results under transfer
  - Selecting the wrong one may move to a state where we don’t have good demonstrations!
- [Schulman et al. ISRR 2013]
  - Select nearest neighbor with respect to rigidity of the transformation
- How to improve on this?
  - Need a framework for demonstration selection!

Demo + Transfer Method → Policy



Demo + Transfer Method ➔ Policy



- π_d specifies an *option*
  - Option = policy + termination condition
  - Selecting an option runs the corresponding policy until the termination condition

$$M \;+\; \mathcal{D} \;\Rightarrow\; M_{\mathcal{D}}$$

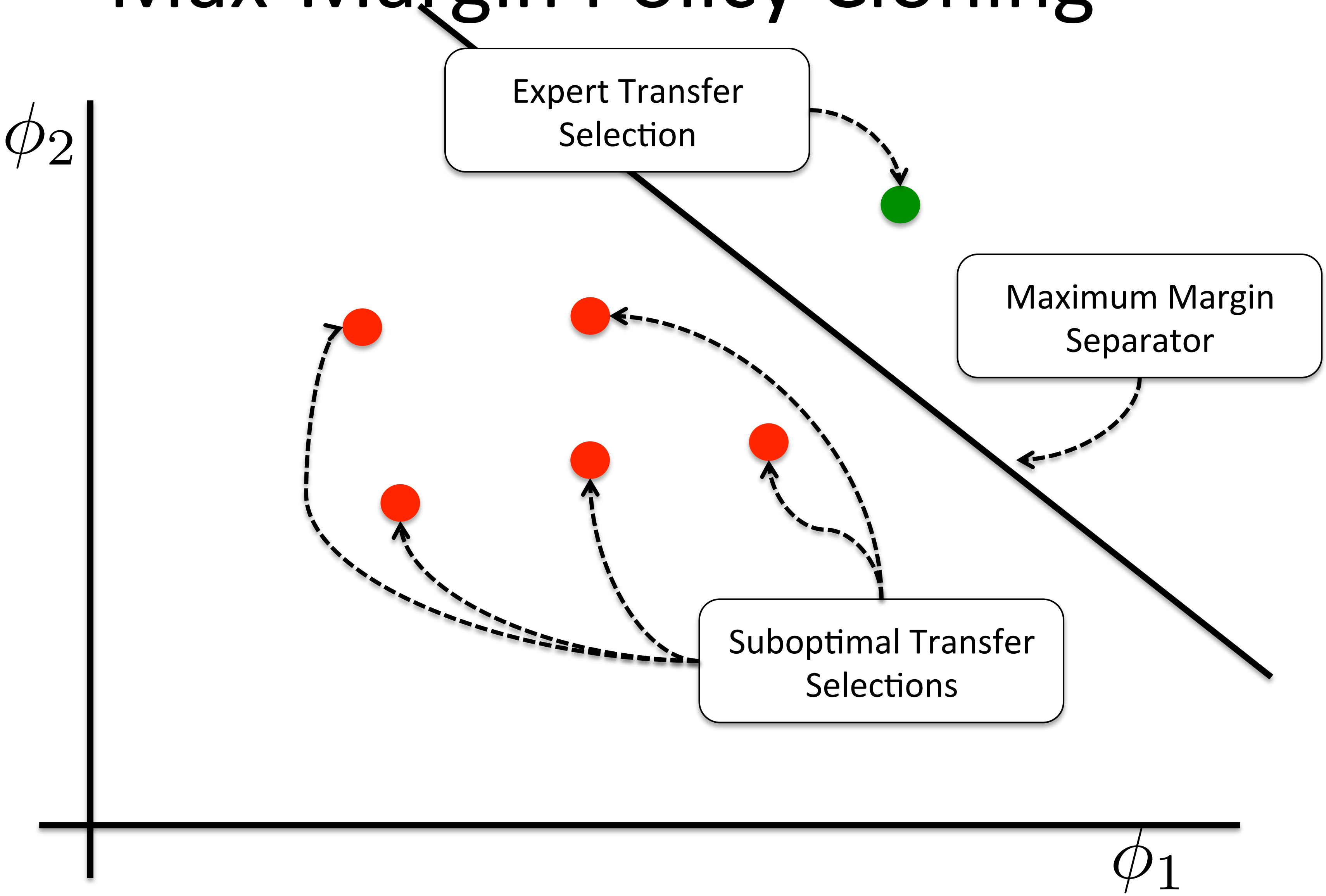
Original (intractable) MDP   Demonstration Library   Options MDP

	M	vs $M_{\mathcal{D}}$
$\lvert\mathcal{A}\rvert$	$\mathbb{R}^{14}$	$\lvert\mathcal{D}\rvert \approx 150$
H	≈ 100	≈ 4
$\lvert\mathcal{S}\rvert$	$\mathbb{R}^{230}$	$\mathbb{R}^{230}$

Takeaways

- Heuristic Method from ISRR paper is a policy for $M_{\mathcal{D}}$
- Learning policies is something we know how to do
- Can we apply that here?
  - State space is still a challenge
- Solution: use expert knowledge again
  - This time about *which* demonstrations to transfer

Max-Margin Policy Cloning



Max-Margin Policy Cloning

Maximize the Margin

Prefer Expert Selections

$$\min_w \boxed{w^\top w}$$

$$\text{s.t.} \boxed{w^\top \phi(s, d_{exp}) \geq w^\top \phi(s, d') + 1; \forall s}$$

Details

- Expert Selections gathered by watching multiple transfers from same state and selecting `best'
- Structured margin to capture similarity between demonstrations
- Slack variables to cope with sub-optimality in choices

Max-Margin Q-function Estimation

- Policy Cloning is good, but has some drawbacks
  - Ranking function has no natural interpretation
  - No direct notion of progress
  - No comparisons between states
- We have a bunch of other information
  - Cost function for MDP, Bellman constraints on value function…etc
- Solution: modify Max-Margin Policy Cloning to learn an approximate Q-function

Max-Margin Q-function Estimation

Maximize the Margin

Minimize Bellman Error

$$\min_{w,\xi_i} \boxed{w^\top w} + \boxed{\sum |\xi_i|}$$

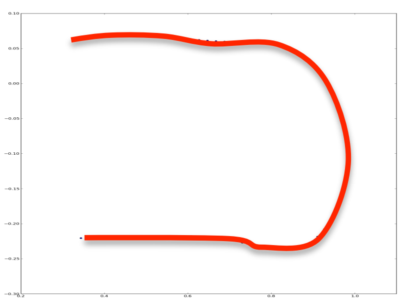
$$\text{s.t.} \quad \boxed{w^\top \phi(s_i, d_{exp}(s_i) = w^\top \phi(s_{i+1}, d_{exp}(s_{i+1}) - \gamma C + \xi_i}$$

$$\boxed{w^\top \phi(s, d_{exp}(s)) \geq w^\top \phi(s, d') + 1; \forall s}$$

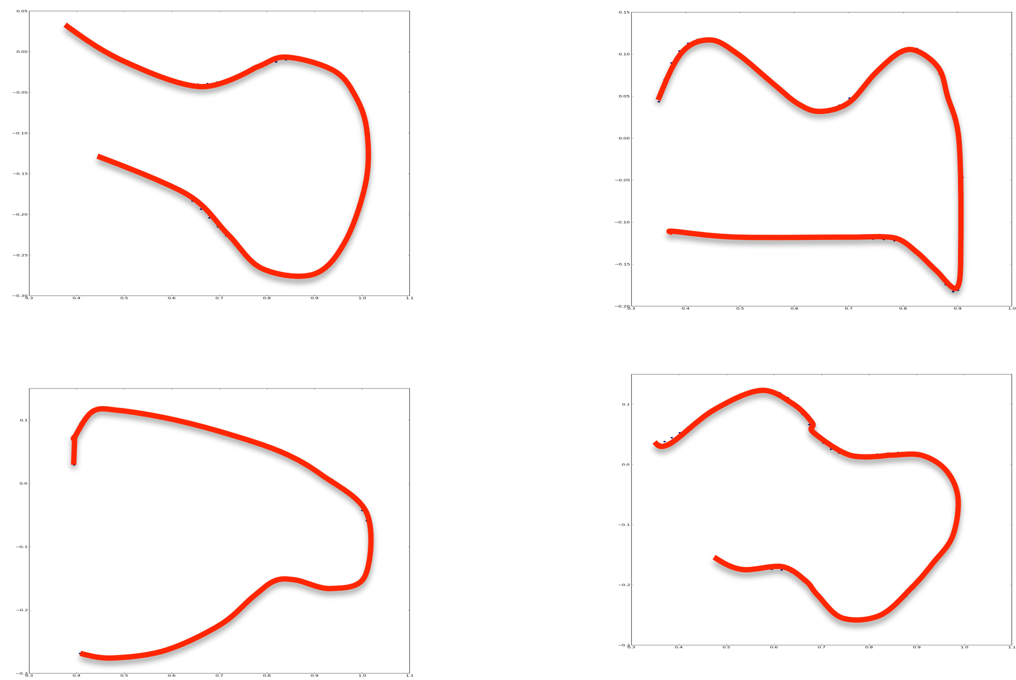
Prefer Expert Selections

Evaluation on Overhand Knot-Tying

- Distribution over initial states
  - Initial states from demonstrations with 10cm perturbations at 7 random locations along rope
- Compare success rate for tying overhand knot on 500 perturbed instances

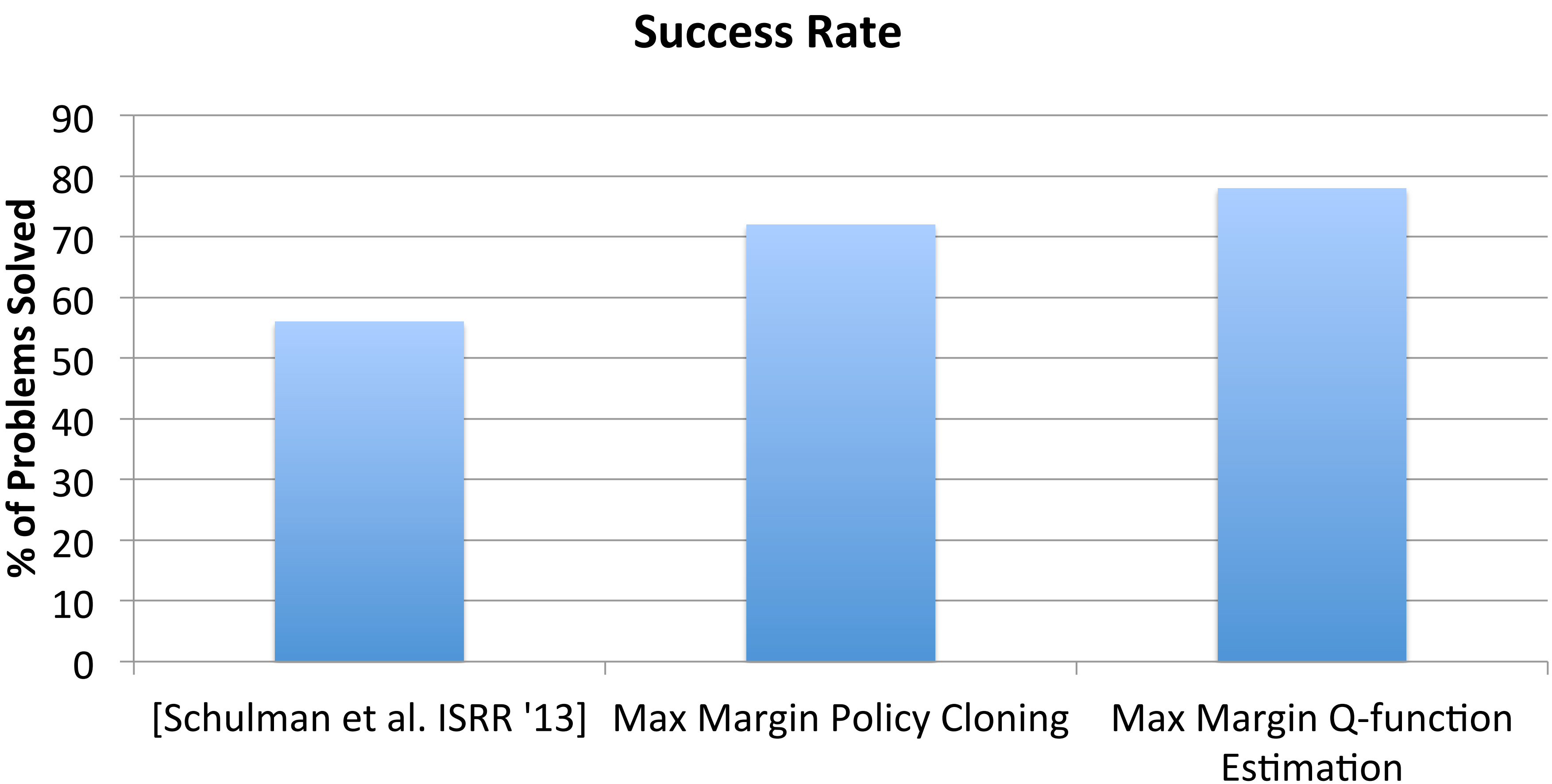


Example Initial State



Samples from Perturbed Distribution

Evaluation on Overhand Knot-Tying

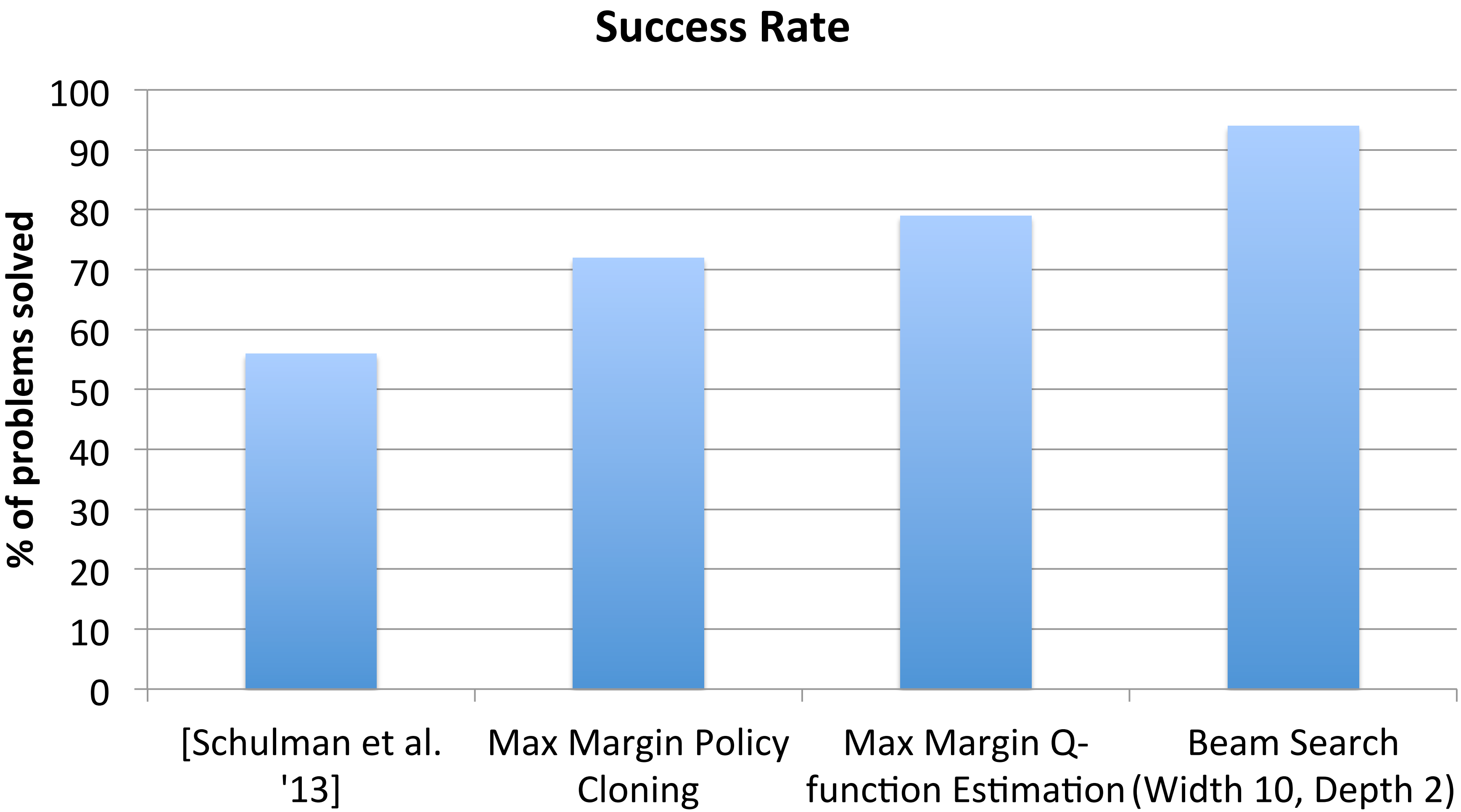


56% 72% 79%

Search

- We have an estimate of the Q-function
- If we have access to a simulator, we can do a local expansion of the state space graph
- Select the action that maximizes the Q-function at the search horizon
- Large Branching Factor → Beam Search

Evaluation on Overhand Knot-Tying



56% 72% 79% 94%

Next Steps

- More difficult tasks
  - More complex knots → longer time horizon
- Other robots
  - Humanoid robot demonstration from motion capture
  - More complicated end effectors
- Transferring more than trajectories?
  - Linear Feedback controllers? Arbitrary policies?