

Abstract

How can we learn motor tasks efficiently without expert knowledge?

- experience-based autonomous learning from scratch
- no task-specific prior assumptions
- artificial learning often requires many (millions) trials, humans don't
- use key features from human/animal learning to make artificial learning more efficient
- probabilistic model for predictions is of central importance

1 Key Ingredients and Algorithm

- some important features of human experience-based learning:
 - **generalization** and predictions using a forward model
 - representation and incorporation of **uncertainty** into the decision-making process
 - describe features by a **probabilistic model** of the world
- properties of a probabilistic model
 - extract more useful information from data
 - represent and quantify uncertainty
 - simultaneous consideration of all “plausible” transitions

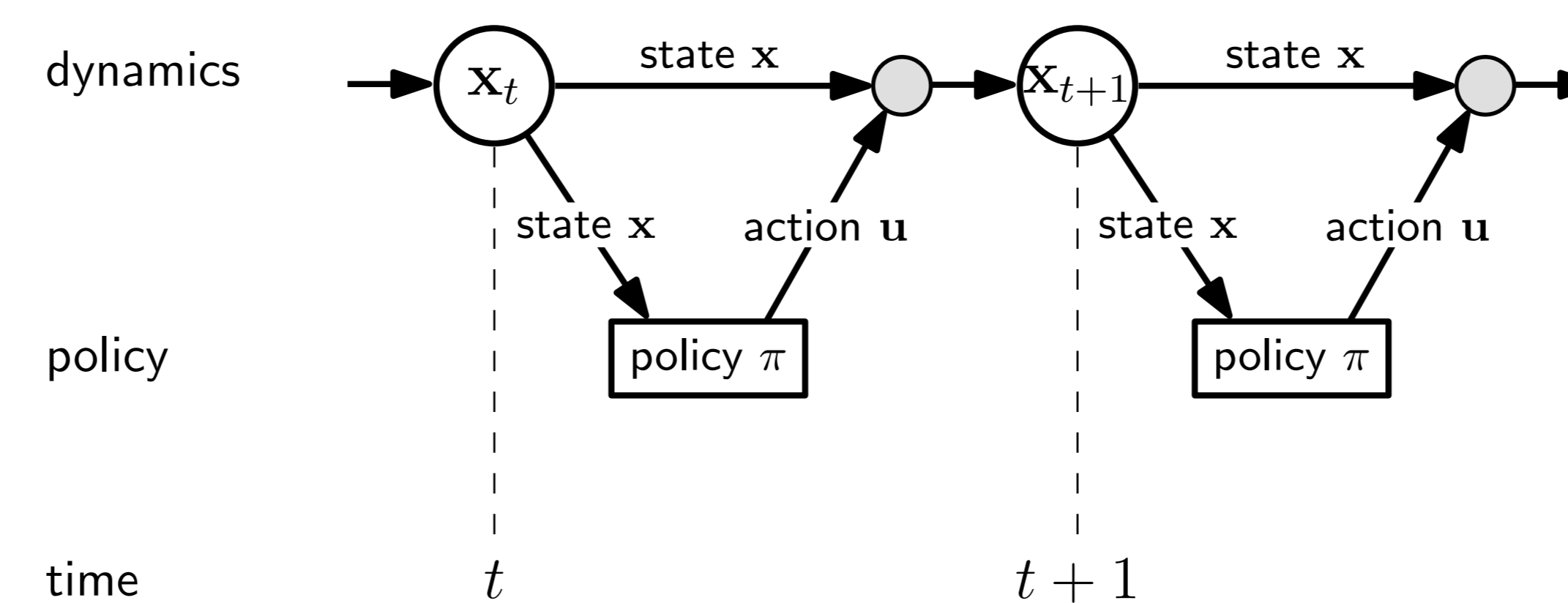
Algorithm 1 Learning algorithm

```

init: try some policy (real world)           ▷ interaction
loop
  record collected experience
  update probabilistic dynamics model         ▷ internal
  loop                                       ▷ policy optimization
    simulate model with current policy       ▷ internal
    compute corresponding expected long-term cost ▷ internal
    improve policy                           ▷ internal
  end loop
  try new policy (real world)               ▷ interaction
end loop
    
```

2 Some Details

- use **Gaussian processes** (GPs) to learn short-term transition dynamics
 - adaptive, non-parametric, probabilistic
 - tractable Bayesian inference (no sampling required)
- **cascade** short-term predictions to obtain long-term predictions [2]
 - **crucial:** keep track of uncertainty evolution
- explicitly consider **distributions** over states and actions during internal simulation



- **analytic** expression for expected long-term cost V^π along path τ

$$V^\pi(\mathbf{x}_0) = \mathbb{E}_\tau \left[\sum_{t=0}^T \ell(\mathbf{x}_t) | \pi \right] = \sum_{t=0}^T \mathbb{E}_{\mathbf{x}_t} [\ell(\mathbf{x}_t)]$$

- **saturating immediate cost function** ℓ
 - does not penalize unimportant details of the state distribution
 - indirectly controls exploration/exploitation even for a greedy policy

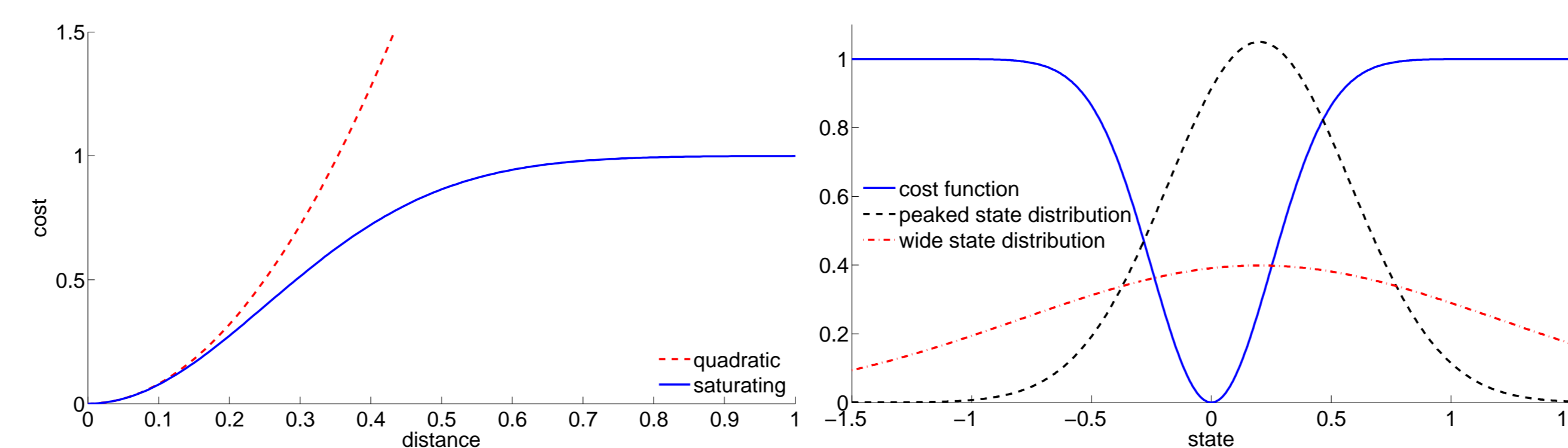


Figure 1: Saturating immediate cost ℓ . Left panel: comparison to quadratic cost. Right panel: allowance for “natural” exploration/exploitation due to probabilistic modeling.

3 Results

- learn to solve tasks from scratch according to Algorithm 1
- probabilistic GP dynamics model based on experience
- cost function only penalizes distance from the target

3.1 Pendubot (Double Pendulum)

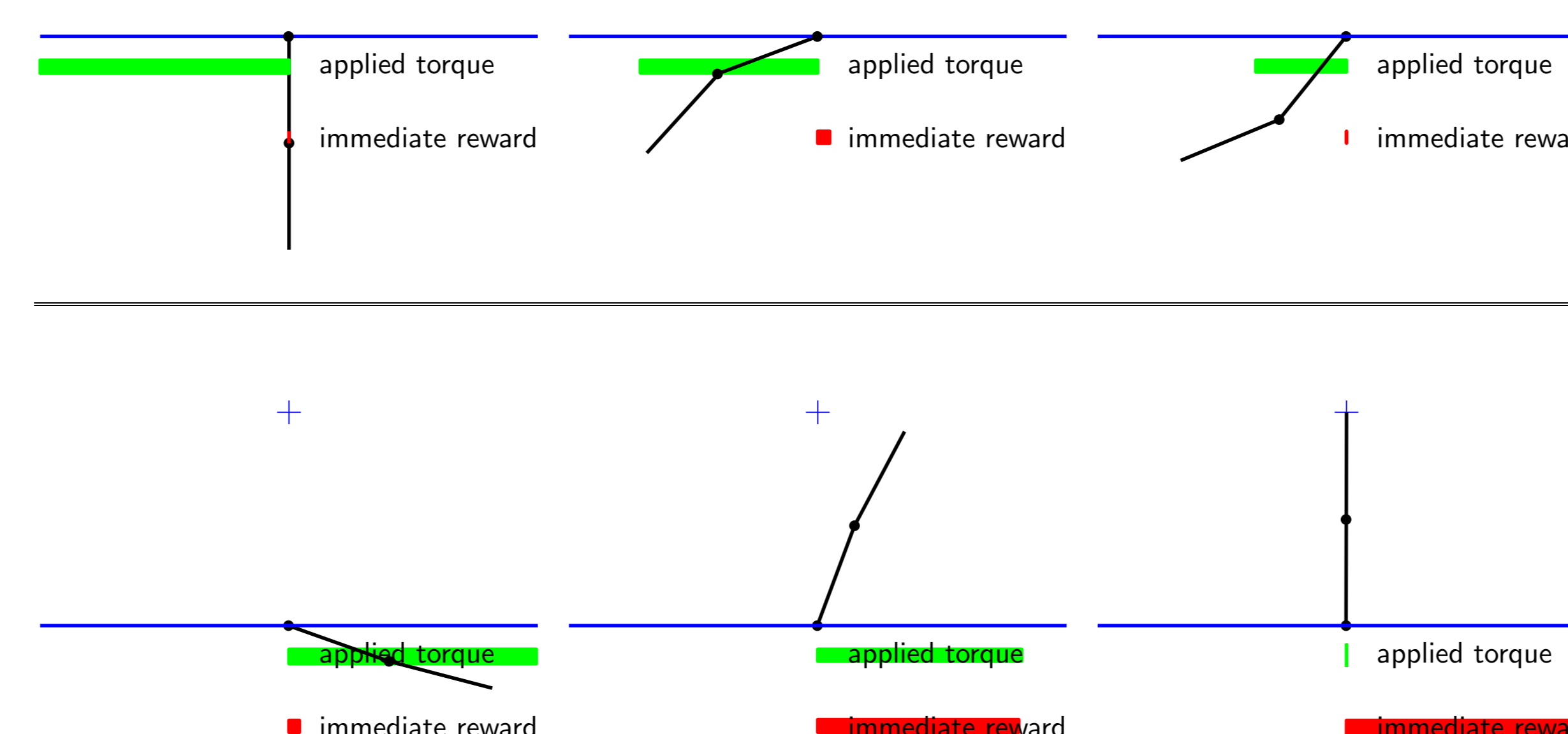


Figure 2: Swing-up of the Pendubot (only first joint is actuated) using experience of about 2 minutes.

3.2 Inverted Pendulum

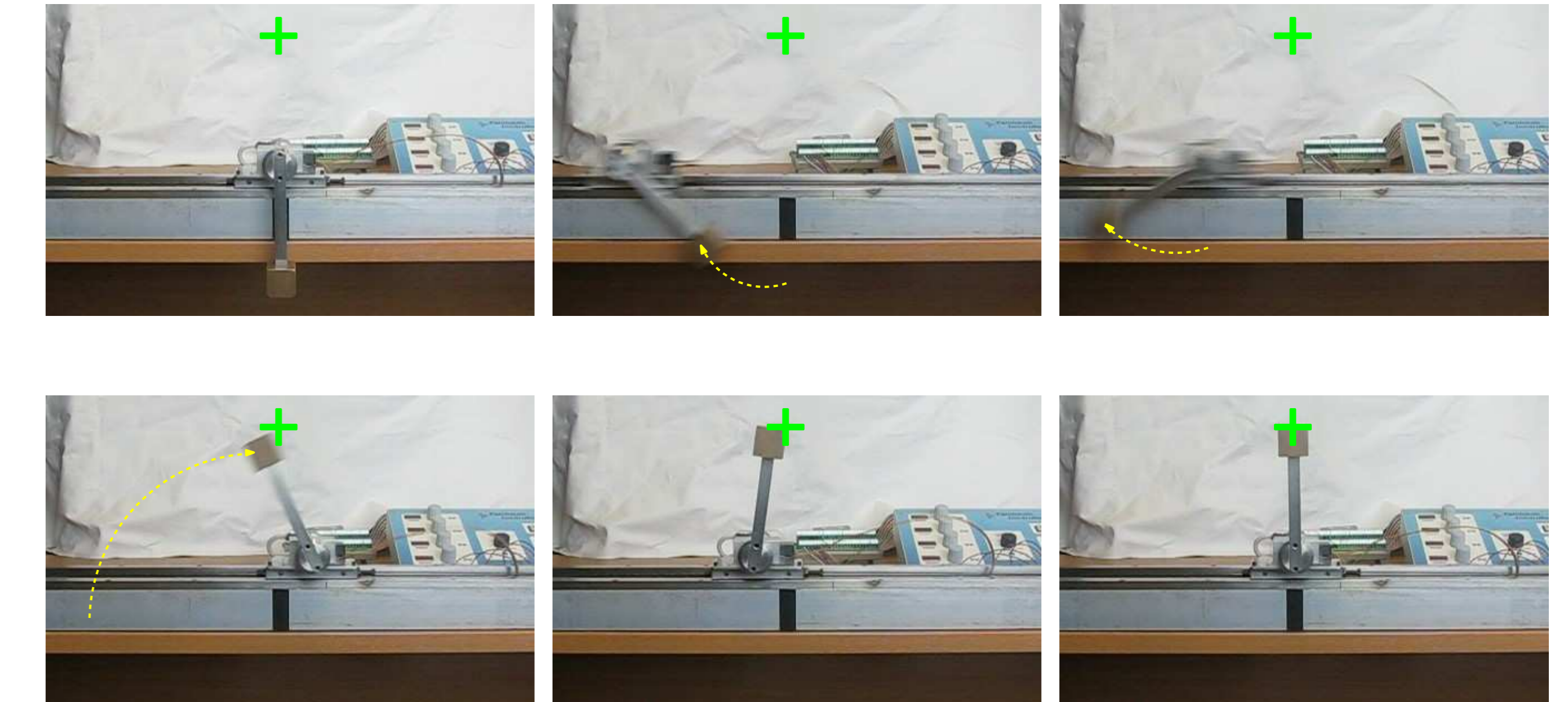


Figure 3: Snapshots of a typical trajectory after having learned the task. Learning the swing up plus balancing required 17.5 seconds experience.

4 Discussion

- algorithm **learns** important details of the tasks (e.g., low velocities around the target state)
- predicted uncertainty along a good trajectory declines to zero
- probabilistic model is crucial: deterministic model does not work
- unprecedented speed of learning (number of trials)
- hardware experiment demonstrates success and applicability

5 Wrap-up

- probabilistic models capture two key ingredients of human learning: generalization and explicit uncertainty modeling
- efficient learning from scratch without expert knowledge
- coherent Bayesian averaging over unknowns is crucial
- no success with deterministic models

References

- [1] C. E. Rasmussen and M. P. Deisenroth. *Probabilistic Inference for Fast Learning in Control*. Chapter in *Recent Advances in Reinforcement Learning*, vol. 5323 of *Lecture Notes in Computer Science*, pp. 229–242. Springer-Verlag, November 2008.
- [2] J. Quiñero-Candela, A. Girard, J. Larsen, and C. E. Rasmussen. Propagation of Uncertainty in Bayesian Kernel Models—Application to Multiple-Step Ahead Forecasting. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2003)*, pp. 701–704, April 2003.