Useful Models for Robot Learning

Marc Deisenroth

Department of Computer Science
University College London

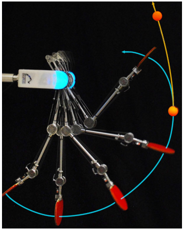
Steindór Sæmundsson

Department of Computing
Imperial College London

NeurIPS Workshop on Robot Learning

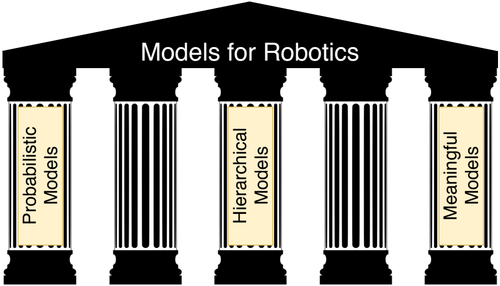
December 14, 2019

Challenges in Robot Learning

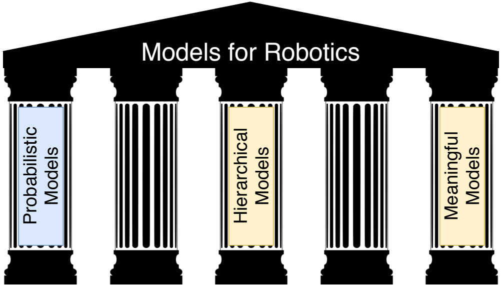


- Automatic adaption in robotics ⏩ Learning
- Practical constraint: data efficiency
- Models are useful for data-efficient learning in robotics

3 Models for Data-Efficient Robot Learning



1. **Probabilistic models**
   - Fast reinforcement learning
2. **Hierarchical models**
   - Infer task similarities within a meta-learning framework
3. **Physically meaningful models**
   - Encode real-world constraints into learning



Models for Robotics
Probabilistic Models
Hierarchical Models
Meaningful Models

Reinforcement Learning and Optimal Control

$$\boldsymbol{x}_{t+1} = f(\boldsymbol{x}_t, \boldsymbol{u}_t) + \boldsymbol{w}\,, \qquad \boldsymbol{u}_t = \pi(\boldsymbol{x}_t, \boldsymbol{\theta})$$

State

Control

Policy

Policy parameters

Transition function

Reinforcement Learning and Optimal Control

$$\boldsymbol{x}_{t+1} = f(\boldsymbol{x}_t, \boldsymbol{u}_t) + \boldsymbol{w}\,, \qquad \boldsymbol{u}_t = \pi(\boldsymbol{x}_t, \boldsymbol{\theta})$$

State — Transition function — Control — Policy — Policy parameters

Objective (Controller Learning)

Find policy parameters $\boldsymbol{\theta}^*$ that minimize the expected long-term cost

$$J(\boldsymbol{\theta}) = \sum_{t=1}^{T} \mathbb{E}[c(\boldsymbol{x}_t)|\boldsymbol{\theta}]\,, \qquad p(\boldsymbol{x}_0) = \mathcal{N}\big(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0\big)\,.$$

Instantaneous cost $c(\boldsymbol{x}_t)$, e.g., $\|\boldsymbol{x}_t - \boldsymbol{x}_{\text{target}}\|^2$

- Typical objective in optimal control and reinforcement learning (Bertsekas, 2005; Sutton & Barto, 1998)

Fast Reinforcement Learning

Objective

Minimize expected long-term cost $J(\boldsymbol{\theta}) = \sum_t \mathbb{E}[c(\boldsymbol{x}_t)|\boldsymbol{\theta}]$

Deisenroth et al. (IEEE-TPAMI, 2015): *Gaussian Processes for Data-Efficient Learning in Robotics and Control*

Objective

Minimize expected long-term cost $J(\boldsymbol{\theta}) = \sum_t \mathbb{E}[c(\boldsymbol{x}_t)|\boldsymbol{\theta}]$

PILCO Framework: High-Level Steps

1. Probabilistic model for transition function
   - ▶▶ System identification

Deisenroth et al. (IEEE-TPAMI, 2015): *Gaussian Processes for Data-Efficient Learning in Robotics and Control*

Objective

Minimize expected long-term cost $J(\boldsymbol{\theta}) = \sum_t \mathbb{E}[c(\boldsymbol{x}_t)|\boldsymbol{\theta}]$

PILCO Framework: High-Level Steps

1. Probabilistic model for transition function
   - System identification
2. Compute long-term state evolution $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \ldots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$

Deisenroth et al. (IEEE-TPAMI, 2015): *Gaussian Processes for Data-Efficient Learning in Robotics and Control*

Fast Reinforcement Learning

Objective

Minimize expected long-term cost $J(\boldsymbol{\theta}) = \sum_t \mathbb{E}[c(\boldsymbol{x}_t)|\boldsymbol{\theta}]$

PILCO Framework: High-Level Steps

1. Probabilistic model for transition function
   ▶▶ System identification
2. Compute long-term state evolution $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \ldots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$
3. Policy improvement

Deisenroth et al. (IEEE-TPAMI, 2015): *Gaussian Processes for Data-Efficient Learning in Robotics and Control*

Fast Reinforcement Learning

Objective

Minimize expected long-term cost $J(\boldsymbol{\theta}) = \sum_t \mathbb{E}[c(\boldsymbol{x}_t)|\boldsymbol{\theta}]$

PILCO Framework: High-Level Steps

1. Probabilistic model for transition function
   ▶▶ System identification
2. Compute long-term state evolution $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \ldots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$
3. Policy improvement
4. Apply controller

Deisenroth et al. (IEEE-TPAMI, 2015): *Gaussian Processes for Data-Efficient Learning in Robotics and Control*

Objective

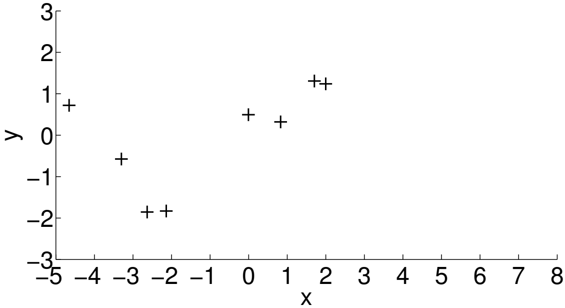
Minimize expected long-term cost $J(\boldsymbol{\theta}) = \sum_t \mathbb{E}[c(\boldsymbol{x}_t)|\boldsymbol{\theta}]$

PILCO Framework: High-Level Steps

1. **Probabilistic model for transition function f**
   ▸▸ **System identification**
2. Compute long-term predictions $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \ldots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$
3. Policy improvement
4. Apply controller

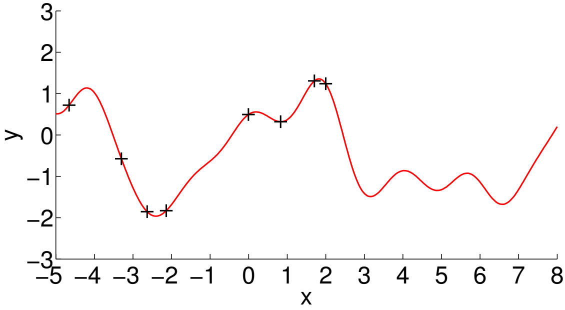
Deisenroth et al. (IEEE-TPAMI, 2015): *Gaussian Processes for Data-Efficient Learning in Robotics and Control*

Model learning problem: Find a function $f : x \mapsto f(x) = y$



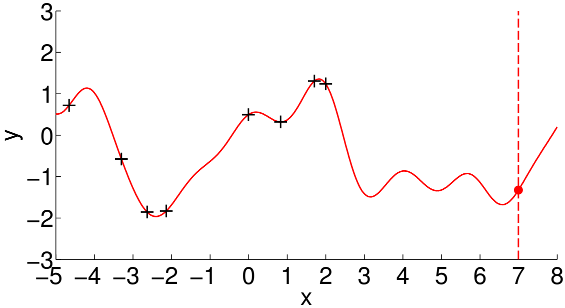
Observed function values

Model learning problem: Find a function $f : x \mapsto f(x) = y$



Plausible model

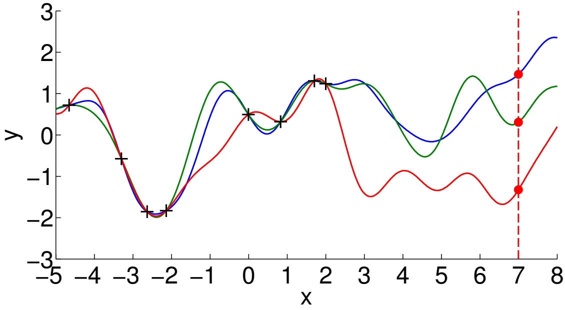
Model learning problem: Find a function $f : x \mapsto f(x) = y$



Plausible model

Predictions? Decision Making?

Model learning problem: Find a function $f : x \mapsto f(x) = y$

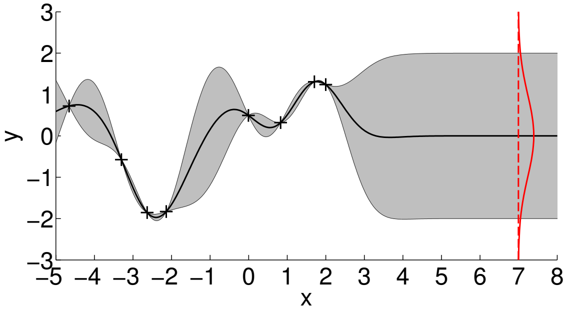


More plausible models

Predictions? Decision Making? Model Errors!

Model Learning (System Identification)

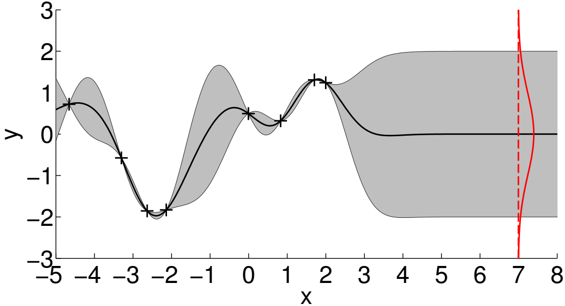
Model learning problem: Find a function $f : x \mapsto f(x) = y$



Distribution over plausible functions

Model Learning (System Identification)

Model learning problem: Find a function $f : x \mapsto f(x) = y$



Distribution over plausible functions

- Express uncertainty about the underlying function to be robust to model errors
- Gaussian process for model learning (Rasmussen & Williams, 2006)

Objective

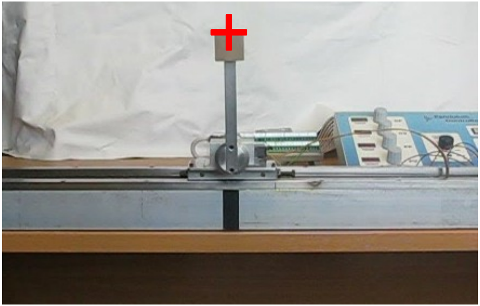
Minimize expected long-term cost $J(\boldsymbol{\theta}) = \sum_t \mathbb{E}[c(\boldsymbol{x}_t)|\boldsymbol{\theta}]$

PILCO Framework: High-Level Steps

1. Probabilistic model for transition function f
   ▸▸ System identification
2. Compute long-term predictions $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \ldots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$
3. Policy optimization via gradient descent
4. Apply controller

Deisenroth et al. (IEEE-TPAMI, 2015): *Gaussian Processes for Data-Efficient Learning in Robotics and Control*

Standard Benchmark: Cart-Pole Swing-up

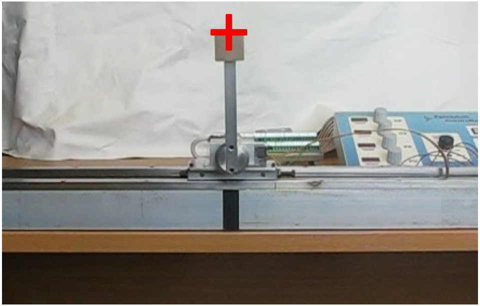


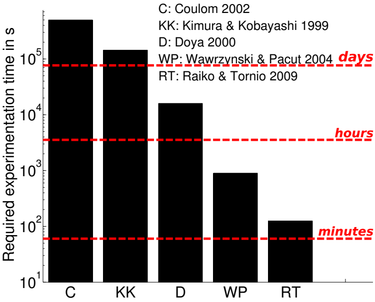
- Swing up and balance a freely swinging pendulum on a cart
- No knowledge about nonlinear dynamics ⏩ Learn from scratch
- Cost function $c(\boldsymbol{x}) = 1 - \exp(-\|\boldsymbol{x} - \boldsymbol{x}_{\text{target}}\|^2)$

- Code: https://github.com/ICL-SML/pilco-matlab

Deisenroth & Rasmussen (ICML, 2011): *PILCO: A Model-based and Data-efficient Approach to Policy Search*

Standard Benchmark: Cart-Pole Swing-up



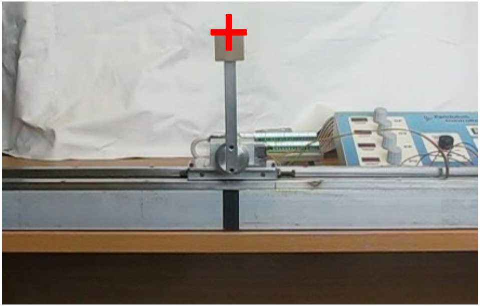


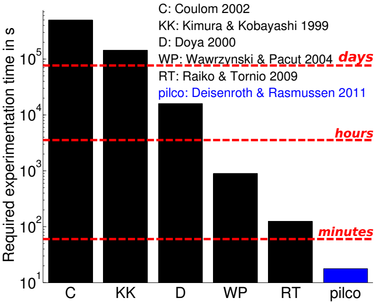
- Swing up and balance a freely swinging pendulum on a cart
- No knowledge about nonlinear dynamics ⏩ Learn from scratch
- Cost function $c(\boldsymbol{x}) = 1 - \exp(-\|\boldsymbol{x} - \boldsymbol{x}_{\text{target}}\|^2)$

- Code: https://github.com/ICL-SML/pilco-matlab

Deisenroth & Rasmussen (ICML, 2011): *PILCO: A Model-based and Data-efficient Approach to Policy Search*

Standard Benchmark: Cart-Pole Swing-up

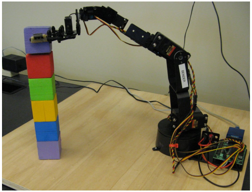




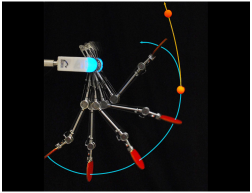
- Swing up and balance a freely swinging pendulum on a cart
- No knowledge about nonlinear dynamics ⏩ Learn from scratch
- Cost function $c(\boldsymbol{x}) = 1 - \exp(-\|\boldsymbol{x} - \boldsymbol{x}_{\text{target}}\|^2)$
- **Unprecedented learning speed** compared to state-of-the-art
- Code: `https://github.com/ICL-SML/pilco-matlab`

Deisenroth & Rasmussen (ICML, 2011): *PILCO: A Model-based and Data-efficient Approach to Policy Search*

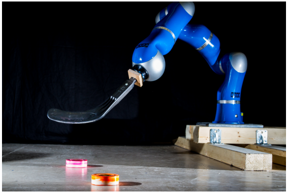
Wide Applicability



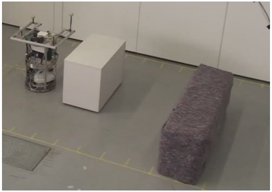
with D Fox



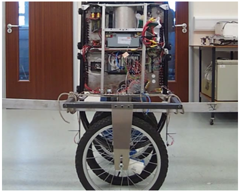
with P Englert, A Paraschos, J Peters



with A Kupcsik, J Peters, G Neumann



B Bischoff (Bosch), ESANN 2013



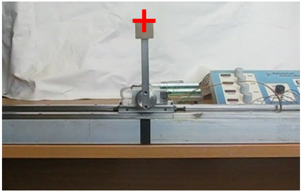
A McHutchon (U Cambridge)

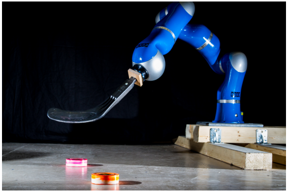
B Bischoff (Bosch), ECML 2013

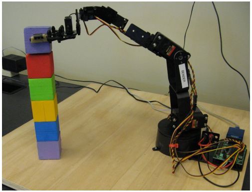
▶▶ Application to a wide range of robotic systems

Deisenroth et al. (RSS, 2011): *Learning to Control a Low-Cost Manipulator using Data-efficient Reinforcement Learning*
Englert et al. (ICRA, 2013): *Model-based Imitation Learning by Probabilistic Trajectory Matching*
Deisenroth et al. (ICRA, 2014): *Multi-Task Policy Search for Robotics*
Kupcsik et al. (AIJ, 2017): *Model-based Contextual Policy Search for Data-Efficient Generalization of Robot Skills*

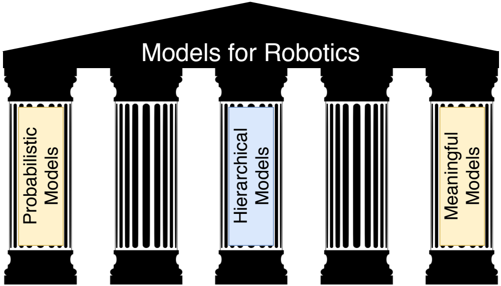
Summary (1)







- In robotics, data-efficient learning is critical
- Probabilistic, model-based RL approach
  - Reduce model bias
  - Unprecedented learning speed
  - Wide applicability



Steindór Sæmundsson

Katja Hofmann

Meta Learning (Schmidhuber 1987)

Generalize knowledge from known tasks to new (related) tasks

Meta Learning

Meta Learning (Schmidhuber 1987)

Generalize knowledge from known tasks to new (related) tasks

- Different robot configurations (link lengths, weights, ...)
- Re-use experience gathered so far generalize learning to new dynamics that are similar
  ▶▶ Accelerated learning

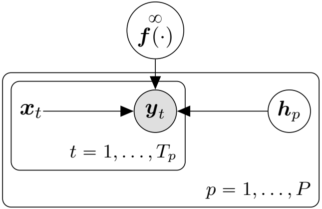
Approach

- Separate global and task-specific properties
- Shared global parameters describe general dynamics
- Describe task-specific (local) properties with latent variable

Approach

- Separate global and task-specific properties
- Shared global parameters describe general dynamics
- Describe task-specific (local) properties with latent variable
- Online variational inference of local properties

Meta Model Learning with Latent Variables

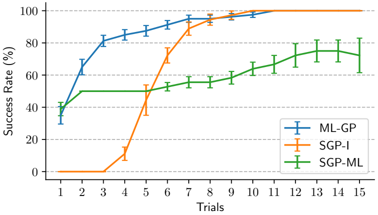


$$\boldsymbol{y}_t = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{h}_p)$$

- GP captures global properties of the dynamics
- Latent variable $\boldsymbol{h}_p$ encodes local properties
  - ▶▶ Variational inference to find a posterior on latent task

Sæmundsson et al. (UAI, 2018): *Meta Reinforcement Learning with Latent Variable Gaussian Processes*

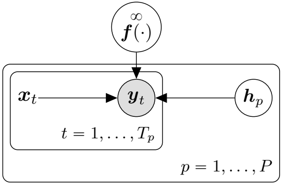
Meta-RL (Cart Pole): Few-Shot Generalization



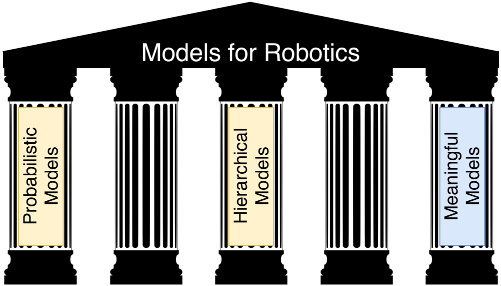
- Train on 6 tasks with different configurations (length/mass)
- Few-shot generalization on 4 unseen configurations
- Success: solve all 10 (6 training + 4 test) tasks
- Meta learning: blue
- Independent (GP-MPC): orange
- Aggregated experience model (no latents): green

▸▸ **Meta RL generalizes well to unseen tasks**

Sæmundsson et al. (UAI, 2018): *Meta Reinforcement Learning with Latent Variable Gaussian Processes*



- Generalize knowledge from known situations to unseen ones
  ▶▶ **Few-shot learning**
- Latent variable can be used to infer task similarities
- Significant speed-up in model learning and model-based RL



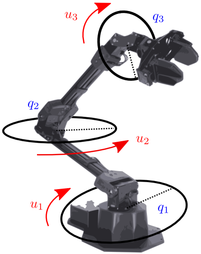
Steindór Sæmundsson

Alexander Terenin

Katja Hofmann

Physically Meaningful Models

Motivation: Data-efficiency and interpretability

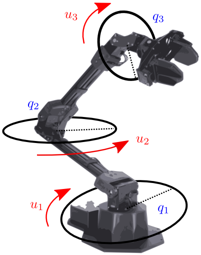


Equations of motion

$$u = \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) - \frac{\partial L}{\partial q}$$

Physically Meaningful Models

Motivation: Data-efficiency and interpretability



Equations of motion

$$u = \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) - \frac{\partial L}{\partial q}$$

Physical Structure:

- Conservation laws
- Position/velocity and mass/force
- Configuration constraints

Lagrangian Mechanics

- **Lagrangian**: Encodes "type" of physics, symmetries.

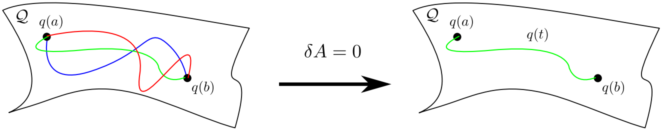
$$L(q(t), \dot{q}(t))$$

- **Lagrangian**: Encodes "type" of physics, symmetries.

$$L(q(t), \dot{q}(t))$$

- **Hamilton's Principle**:

$$A = \int_a^b L(q(t), \dot{q}(t))dt$$

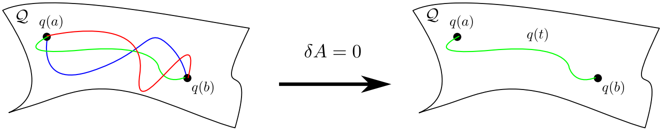


- **Lagrangian**: Encodes "type" of physics, symmetries.

$$L(q(t), \dot{q}(t))$$

- **Hamilton's Principle**:

$$A = \int_a^b L(q(t), \dot{q}(t))dt$$



First idea:

- Learn L instead of dynamics directly
- Encode physical properties in the form of L (e.g., Lutter et al., 2019; Greydanus et al., 2019)

Euler-Lagrange Equations (Equations of motion):

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) - \frac{\partial L}{\partial q} = 0$$

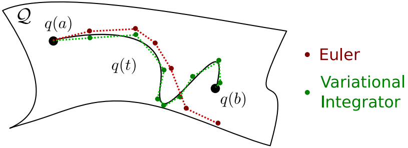
Variational Integrators

Euler-Lagrange Equations (Equations of motion):

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) - \frac{\partial L}{\partial q} = 0$$

Variational Integrators:

- Symplectic
- Momentum preserving
- Bounded energy behavior

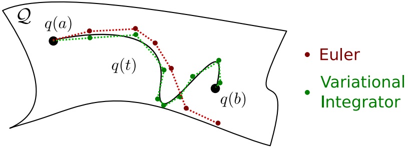


Euler-Lagrange Equations (Equations of motion):

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) - \frac{\partial L}{\partial q} = 0$$

Variational Integrators:

- Symplectic
- Momentum preserving
- Bounded energy behavior



Second idea: Discretize in a way that preserves the physics

Recipe for Variational Integrator Network

1. Write down parameterized Lagrangian:

$$L_\theta(q(t), \dot{q}(t))$$

Sæmundsson et al. (arXiv:1910.09349): *Variational Integrator Networks for Physically Meaningful Embeddings*

Recipe for Variational Integrator Network

1. Write down parameterized Lagrangian:

$$L_\theta(q(t), \dot{q}(t))$$

2. Derive **explicit** variational integrator:

$$q_{t+1} = f_\theta(q_t, q_{t-1})$$

Sæmundsson et al. (arXiv:1910.09349): *Variational Integrator Networks for Physically Meaningful Embeddings*

Recipe for Variational Integrator Network

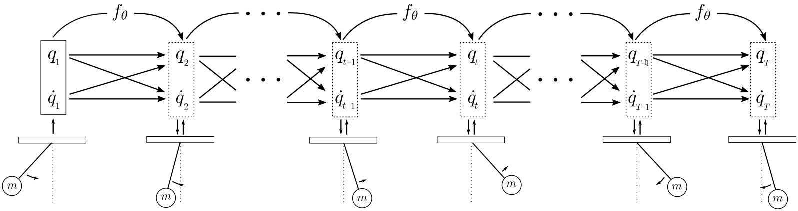
1. Write down parameterized Lagrangian:

$$L_\theta(q(t), \dot{q}(t))$$

2. Derive **explicit** variational integrator:

$$q_{t+1} = f_\theta(q_t, q_{t-1})$$

3. f_θ defines the network architecture



Sæmundsson et al. (arXiv:1910.09349): *Variational Integrator Networks for Physically Meaningful Embeddings*

Recipe for Variational Integrator Network

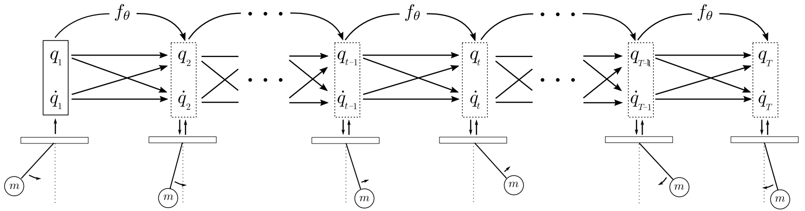
1. Write down parameterized Lagrangian:

$$L_\theta(q(t), \dot{q}(t))$$

2. Derive **explicit** variational integrator:

$$q_{t+1} = f_\theta(q_t, q_{t-1})$$

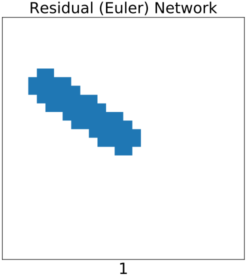
3. f_θ defines the network architecture



▶▶ Define dynamics on $\mathbb{R}^D$ or on manifolds (e.g., SO(2))

Sæmundsson et al. (arXiv:1910.09349): *Variational Integrator Networks for Physically Meaningful Embeddings*

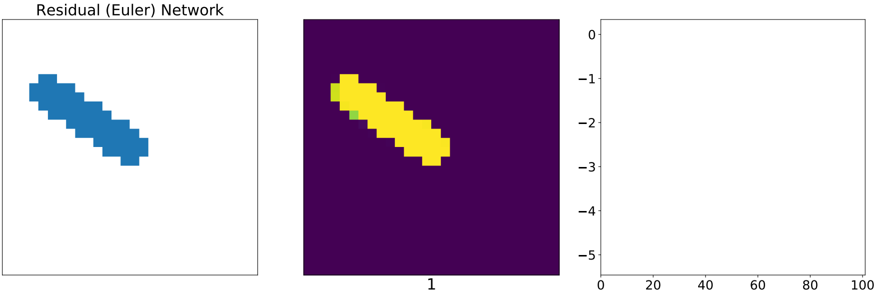
Learning from Pixel Data: Forecasting



- Observations: 28×28 pixel images of pendulum
- Training data: 40 images

Sæmundsson et al. (arXiv:1910.09349): *Variational Integrator Networks for Physically Meaningful Embeddings*

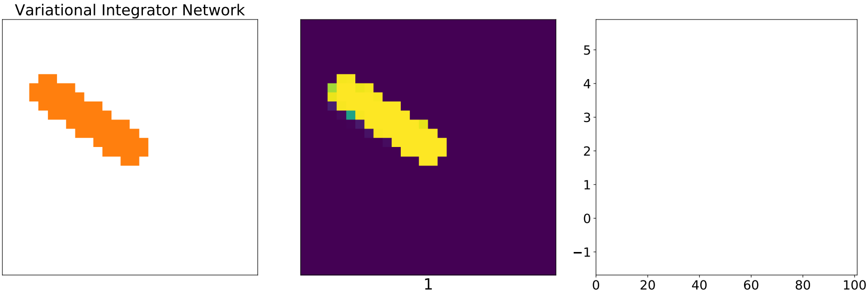
Learning from Pixel Data: Forecasting



- Observations: 28×28 pixel images of pendulum
- Training data: 40 images
- Residual-VAE: Forecasting is not meaningful

Sæmundsson et al. (arXiv:1910.09349): *Variational Integrator Networks for Physically Meaningful Embeddings*

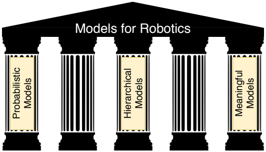
Learning from Pixel Data: Forecasting

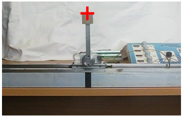


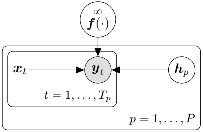
- Observations: 28×28 pixel images of pendulum
- Training data: 40 images
- Residual-VAE: Forecasting is not meaningful
- VIN-VAE: Physically meaningful long-term forecasts in latent and observation space

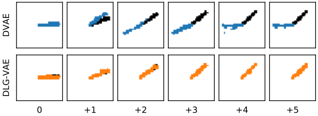
Sæmundsson et al. (arXiv:1910.09349): *Variational Integrator Networks for Physically Meaningful Embeddings*

Wrap-up



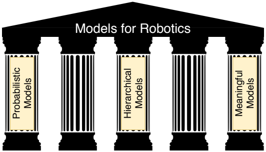


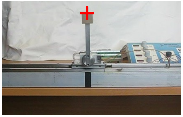


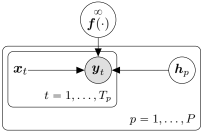


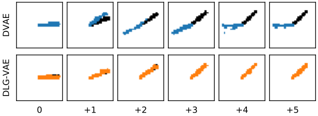
- **Data efficiency** is a practical challenge for autonomous robots
- Three useful models for data-efficient learning in robotics
  1. Probabilistic models for fast reinforcement learning
  2. Hierarchical models for learning task similarities within a meta-learning framework
  3. Physically meaningful models to encode real-world constraints into learning

Wrap-up









- **Data efficiency** is a practical challenge for autonomous robots
- Three useful models for data-efficient learning in robotics
  1. Probabilistic models for fast reinforcement learning
  2. Hierarchical models for learning task similarities within a meta-learning framework
  3. Physically meaningful models to encode real-world constraints into learning

Thank you for your attention

References I

[1] D. P. Bertsekas. *Dynamic Programming and Optimal Control*, volume 1 of *Optimization and Computation Series*. Athena Scientific, Belmont, MA, USA, 3rd edition, 2005.

[2] D. P. Bertsekas. *Dynamic Programming and Optimal Control*, volume 2 of *Optimization and Computation Series*. Athena Scientific, Belmont, MA, USA, 3rd edition, 2007.

[3] B. Bischoff, D. Nguyen-Tuong, T. Koller, H. Markert, and A. Knoll. Learning Throttle Valve Control Using Policy Search. In *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases*, 2013.

[4] M. P. Deisenroth, P. Englert, J. Peters, and D. Fox. Multi-Task Policy Search for Robotics. In *Proceedings of the IEEE International Conference on Robotics and Automation*, 2014.

[5] M. P. Deisenroth, D. Fox, and C. E. Rasmussen. Gaussian Processes for Data-Efficient Learning in Robotics and Control. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2):408–423, 2015.

[6] M. P. Deisenroth and C. E. Rasmussen. PILCO: A Model-Based and Data-Efficient Approach to Policy Search. In *Proceedings of the International Conference on Machine Learning*, 2011.

[7] M. P. Deisenroth, C. E. Rasmussen, and D. Fox. Learning to Control a Low-Cost Manipulator using Data-Efficient Reinforcement Learning. In *Proceedings of Robotics: Science and Systems*, Los Angeles, CA, USA, June 2011.

[8] P. Englert, A. Paraschos, J. Peters, and M. P. Deisenroth. Model-based Imitation Learning by Probabilistic Trajectory Matching. In *Proceedings of the IEEE International Conference on Robotics and Automation*, 2013.

[9] P. Englert, A. Paraschos, J. Peters, and M. P. Deisenroth. Probabilistic Model-based Imitation Learning. *Adaptive Behavior*, 21:388–403, 2013.

[10] S. Greydanus, M. Dzamba, and J. Yosinski. Hamiltonian Neural Networks. In *Advances in Neural Information Processing Systems*, 2019.

[11] A. Kupcsik, M. P. Deisenroth, J. Peters, L. A. Poha, P. Vadakkepata, and G. Neumann. Model-based Contextual Policy Search for Data-Efficient Generalization of Robot Skills. *Artificial Intelligence*, 2017.

[12] M. Lutter, C. Ritter, and J. Peters. Deep Lagrangian Networks: Using Physics as Model Prior for Deep Learning. In *Proceedings of the International Conference on Learning Representations*, 2019.

References II

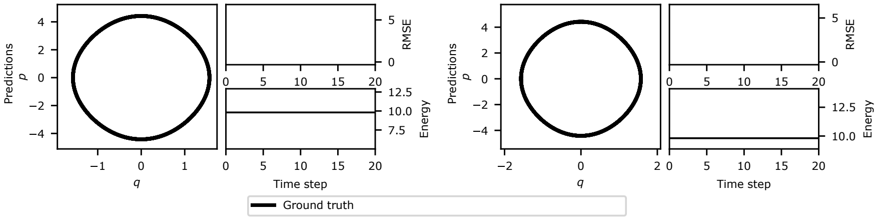
[13] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. The MIT Press, Cambridge, MA, USA, 2006.

[14] S. Sæmundsson, K. Hofmann, and M. P. Deisenroth. Meta Reinforcement Learning with Latent Variable Gaussian Processes. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, 2018.

[15] J. Schmidhuber. Evolutionary Principles in Self-Referential Learning. Master's thesis, 1987.

[16] S. Sæmundsson, A. Terenin, K. Hofmann, and M. P. Deisenroth. Variational Integrator Networks for Physically Meaningful Embeddings. In *arXiv:1910.09349*, 2019.

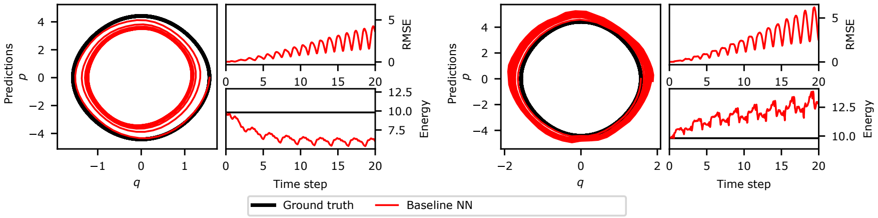
Learning from Noisy Data: Pendulum



Pendulum System. **Left:** 150 observations; **Right:** 750 observations.

Sæmundsson et al. (arXiv:1910.09349): *Variational Integrator Networks for Physically Meaningful Embeddings*

Learning from Noisy Data: Pendulum

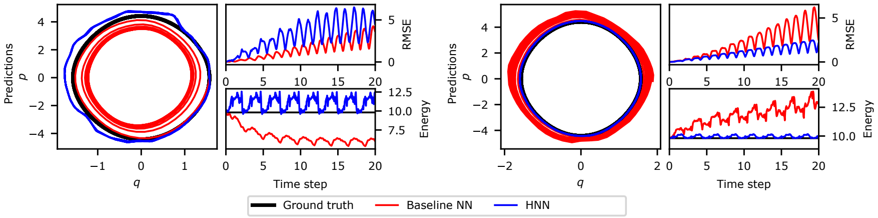


Pendulum System. **Left:** 150 observations; **Right:** 750 observations.

- Baseline neural network: Dissipates/adds energy for low and moderate data

Sæmundsson et al. (arXiv:1910.09349): *Variational Integrator Networks for Physically Meaningful Embeddings*

Learning from Noisy Data: Pendulum

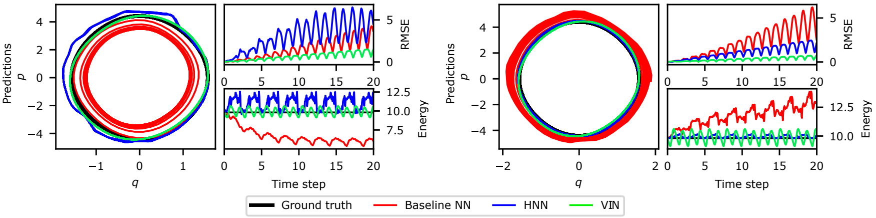


Pendulum System. **Left:** 150 observations; **Right:** 750 observations.

- Baseline neural network: Dissipates/adds energy for low and moderate data
- Hamiltonian neural network (Greydanus et al., 2019): Overfits in low-data regime

Sæmundsson et al. (arXiv:1910.09349): *Variational Integrator Networks for Physically Meaningful Embeddings*

Learning from Noisy Data: Pendulum



Pendulum System. **Left:** 150 observations; **Right:** 750 observations.

- Baseline neural network: Dissipates/adds energy for low and moderate data
- Hamiltonian neural network (Greydanus et al., 2019): Overfits in low-data regime
- Variational integrator network: Conserves energy and generalizes better in both regimes

Sæmundsson et al. (arXiv:1910.09349): *Variational Integrator Networks for Physically Meaningful Embeddings*