Data-Efficient Robot Learning

Marc Deisenroth
Centre for Artificial Intelligence
Department of Computer Science
University College London

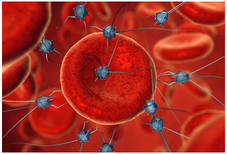
@mpd37
m.deisenroth@ucl.ac.uk
https://deisenroth.cc

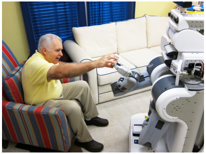
Creative Machine Learning

The AI Talks Series (AI Core Online Event)

April 26, 2020

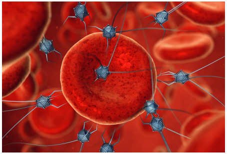
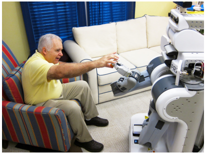
Autonomous Robots



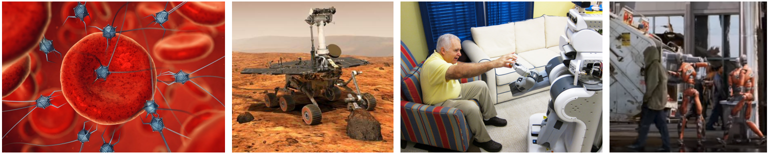



- Vision: Autonomous robots support humans in everyday activities ⏩ Fast learning and automatic adaptation

Autonomous Robots

- Vision: Autonomous robots support humans in everyday activities ⏩ Fast learning and automatic adaptation
- Currently: Data-hungry learning or human guidance



- Vision: Autonomous robots support humans in everyday activities ▶▶ Fast learning and automatic adaptation
- Currently: Data-hungry learning or human guidance

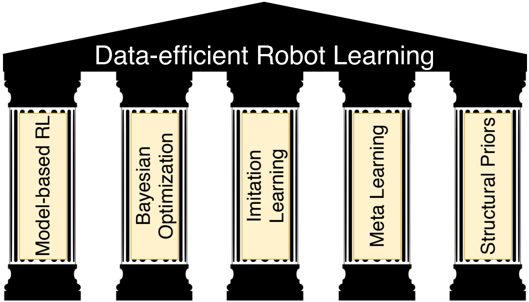
Fully **autonomous learning and decision making with little data** in real-life situations

Central Problem

Data-Efficient Robot Learning

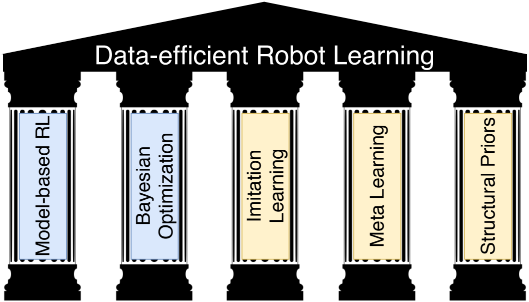
Ability to learn and make decisions in physical domains without requiring large quantities of data

Outline



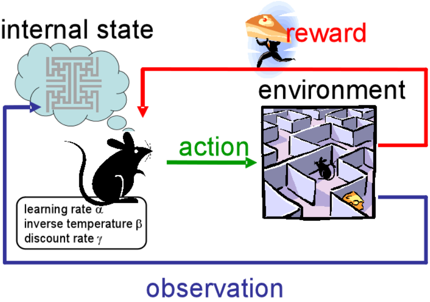
Data-efficient Robot Learning
Model-based RL
Bayesian Optimization
Imitation Learning
Meta Learning
Structural Priors

Outline



Two approaches toward data-efficient robot learning:

1. Model-based reinforcement learning
2. Bayesian optimization



- Learn to solve a task
- Trial-and-error interaction with the environment
- Feedback via reward/cost function

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

State Control Policy Policy parameters

Transition function

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

State Control Policy Policy parameters

Transition function

Objective (Controller Learning)

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

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

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)

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

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 $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \dots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$

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 f
   - System identification
2. Compute long-term predictions $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \ldots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$
3. Policy improvement

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

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*

Objective

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

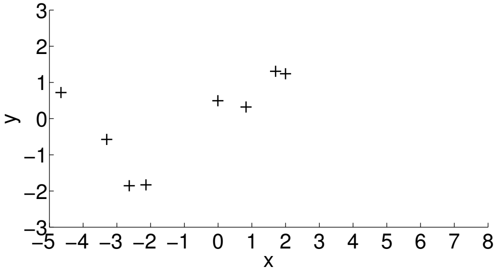
PILCO Framework: High-Level Steps

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 (System Identification)

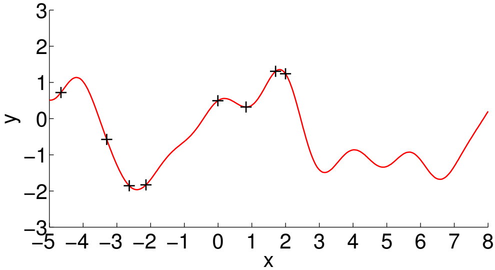
Model learning problem: Find a function $f : x \mapsto f(x) = y$



Observed function values

Model Learning (System Identification)

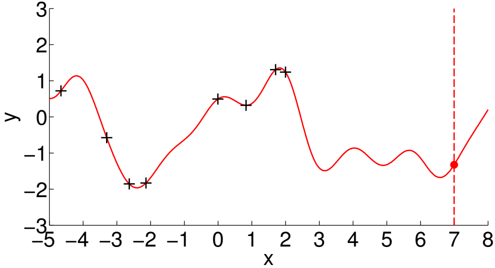
Model learning problem: Find a function $f : x \mapsto f(x) = y$



Plausible model

Model Learning (System Identification)

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

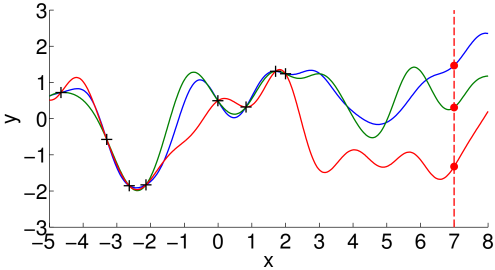


Plausible model

Predictions? Decision Making?

Model Learning (System Identification)

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

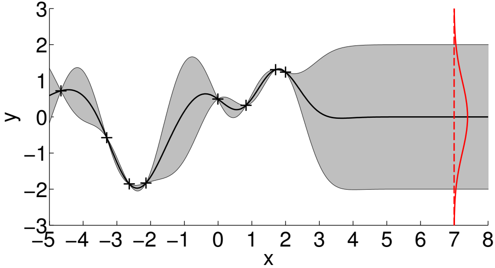


More plausible models

Predictions? Decision Making? Model Errors!

Model Learning (System Identification)

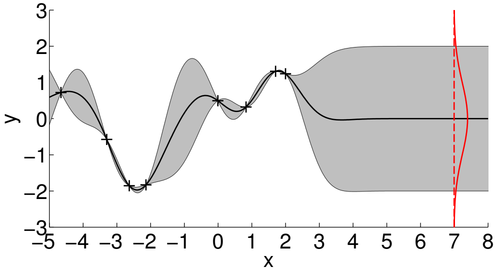
Model learning problem: Find a function $f : x \mapsto f(x) = y$



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)

- Flexible regression model
- Probability distribution over functions
- Fully specified by
  - Mean function m (average function)
  - Covariance function k (assumptions on structure)

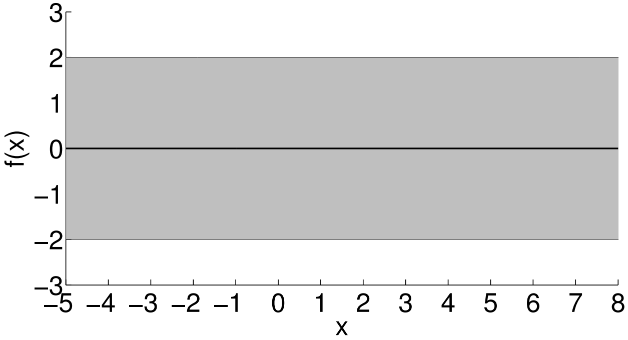
$$k(\boldsymbol{x}_p, \boldsymbol{x}_q) = \mathrm{Cov}[f(\boldsymbol{x}_p), f(\boldsymbol{x}_q)]$$

- Flexible regression model
- Probability distribution over functions
- Fully specified by
  - Mean function m (average function)
  - Covariance function k (assumptions on structure)

$$k(\boldsymbol{x}_p, \boldsymbol{x}_q) = \mathrm{Cov}[f(\boldsymbol{x}_p), f(\boldsymbol{x}_q)]$$

- Predictive distribution at test input $\boldsymbol{x}_*$ is Gaussian (Bayes' theorem):

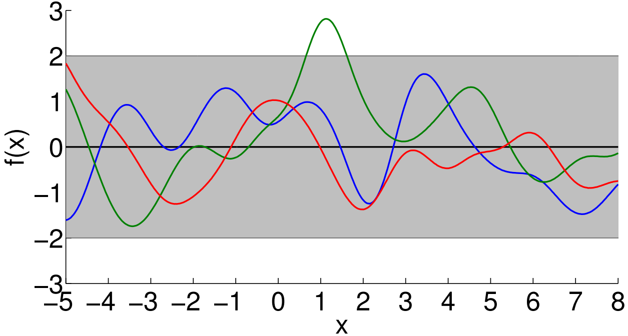
$$p(f(\boldsymbol{x}_*)|\underbrace{\boldsymbol{x}_*}_{\text{Test input}}, \underbrace{\boldsymbol{X}, \boldsymbol{y}}_{\text{Training data}}) = \mathcal{N}\big(f(\boldsymbol{x}_*)\,|\,m(\boldsymbol{x}_*),\, \sigma^2(\boldsymbol{x}_*)\big)$$



Prior belief about the function

Predictive (marginal) mean and variance:

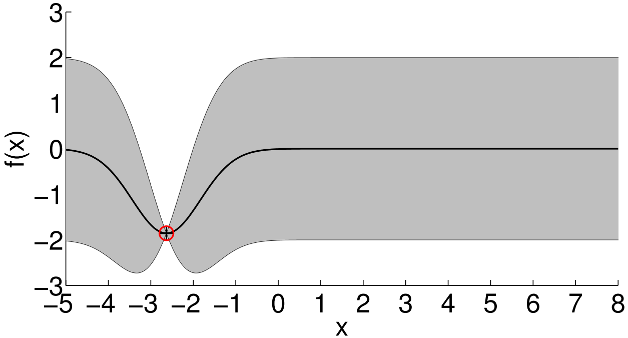
$$
\begin{aligned}
\mathbb{E}[f(\boldsymbol{x}_*)|\boldsymbol{x}_*, \varnothing] &= m(\boldsymbol{x}_*) = 0 \\
\mathbb{V}[f(\boldsymbol{x}_*)|\boldsymbol{x}_*, \varnothing] &= \sigma^2(\boldsymbol{x}_*) = k(\boldsymbol{x}_*, \boldsymbol{x}_*)
\end{aligned}
$$



Prior belief about the function

Predictive (marginal) mean and variance:

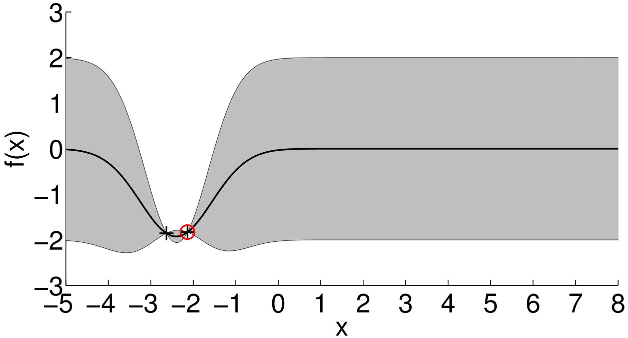
$$\mathbb{E}[f(\boldsymbol{x}_*)|\boldsymbol{x}_*, \varnothing] = m(\boldsymbol{x}_*) = 0$$
$$\mathbb{V}[f(\boldsymbol{x}_*)|\boldsymbol{x}_*, \varnothing] = \sigma^2(\boldsymbol{x}_*) = k(\boldsymbol{x}_*, \boldsymbol{x}_*)$$



Posterior belief about the function

Predictive (marginal) mean and variance:

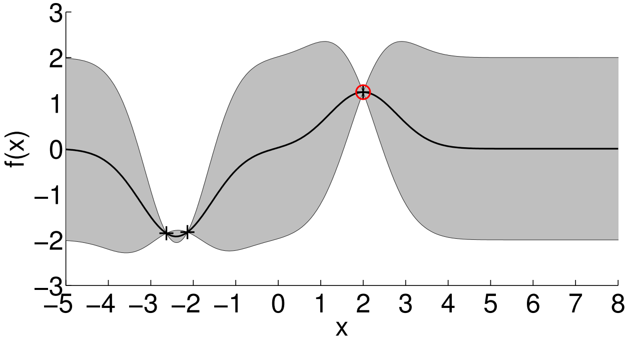
$$\mathbb{E}[f(\boldsymbol{x}_*)|\boldsymbol{x}_*, \boldsymbol{X}, \boldsymbol{y}] = m(\boldsymbol{x}_*) = k(\boldsymbol{X}, \boldsymbol{x}_*)^\top k(\boldsymbol{X}, \boldsymbol{X})^{-1}\boldsymbol{y}$$
$$\mathbb{V}[f(\boldsymbol{x}_*)|\boldsymbol{x}_*, \boldsymbol{X}, \boldsymbol{y}] = \sigma^2(\boldsymbol{x}_*) = k(\boldsymbol{x}_*, \boldsymbol{x}_*) - k(\boldsymbol{X}, \boldsymbol{x}_*)^\top k(\boldsymbol{X}, \boldsymbol{X})^{-1} k(\boldsymbol{X}, \boldsymbol{x}_*)$$



Posterior belief about the function

Predictive (marginal) mean and variance:

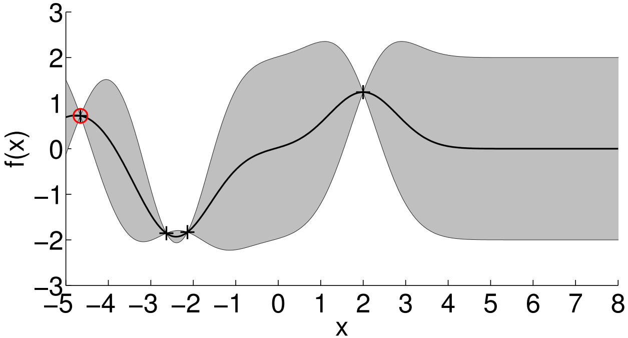
$$\mathbb{E}[f(\boldsymbol{x}_*)|\boldsymbol{x}_*, \boldsymbol{X}, \boldsymbol{y}] = m(\boldsymbol{x}_*) = k(\boldsymbol{X}, \boldsymbol{x}_*)^\top k(\boldsymbol{X}, \boldsymbol{X})^{-1}\boldsymbol{y}$$
$$\mathbb{V}[f(\boldsymbol{x}_*)|\boldsymbol{x}_*, \boldsymbol{X}, \boldsymbol{y}] = \sigma^2(\boldsymbol{x}_*) = k(\boldsymbol{x}_*, \boldsymbol{x}_*) - k(\boldsymbol{X}, \boldsymbol{x}_*)^\top k(\boldsymbol{X}, \boldsymbol{X})^{-1} k(\boldsymbol{X}, \boldsymbol{x}_*)$$



Posterior belief about the function

Predictive (marginal) mean and variance:

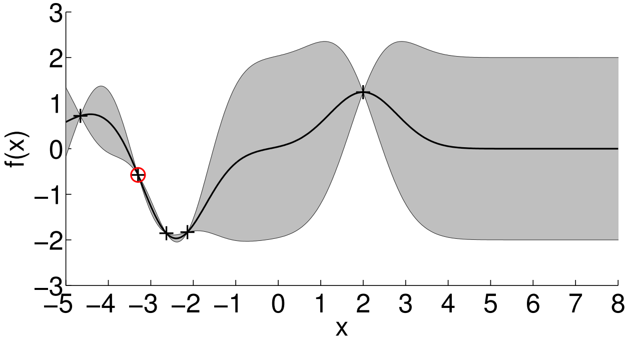
$$\mathbb{E}[f(\boldsymbol{x}_*)|\boldsymbol{x}_*, \boldsymbol{X}, \boldsymbol{y}] = m(\boldsymbol{x}_*) = k(\boldsymbol{X}, \boldsymbol{x}_*)^\top k(\boldsymbol{X}, \boldsymbol{X})^{-1}\boldsymbol{y}$$
$$\mathbb{V}[f(\boldsymbol{x}_*)|\boldsymbol{x}_*, \boldsymbol{X}, \boldsymbol{y}] = \sigma^2(\boldsymbol{x}_*) = k(\boldsymbol{x}_*, \boldsymbol{x}_*) - k(\boldsymbol{X}, \boldsymbol{x}_*)^\top k(\boldsymbol{X}, \boldsymbol{X})^{-1}k(\boldsymbol{X}, \boldsymbol{x}_*)$$



Posterior belief about the function

Predictive (marginal) mean and variance:

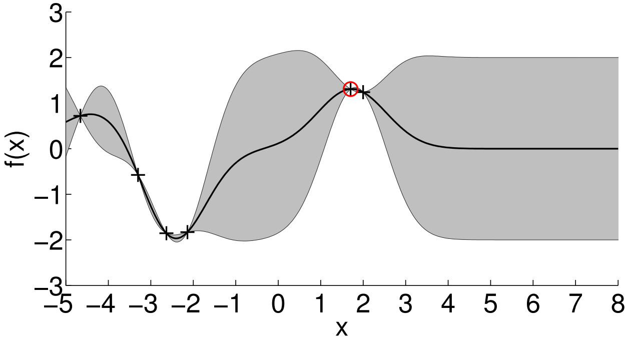
$$\mathbb{E}[f(\boldsymbol{x}_*)|\boldsymbol{x}_*, \boldsymbol{X}, \boldsymbol{y}] = m(\boldsymbol{x}_*) = k(\boldsymbol{X}, \boldsymbol{x}_*)^\top k(\boldsymbol{X}, \boldsymbol{X})^{-1}\boldsymbol{y}$$
$$\mathbb{V}[f(\boldsymbol{x}_*)|\boldsymbol{x}_*, \boldsymbol{X}, \boldsymbol{y}] = \sigma^2(\boldsymbol{x}_*) = k(\boldsymbol{x}_*, \boldsymbol{x}_*) - k(\boldsymbol{X}, \boldsymbol{x}_*)^\top k(\boldsymbol{X}, \boldsymbol{X})^{-1}k(\boldsymbol{X}, \boldsymbol{x}_*)$$



Posterior belief about the function

Predictive (marginal) mean and variance:

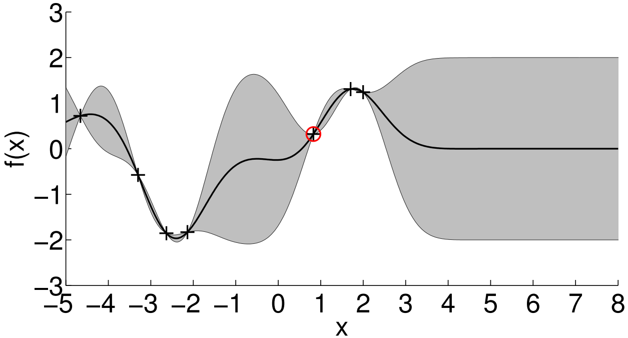
$$\mathbb{E}[f(\boldsymbol{x}_*)|\boldsymbol{x}_*, \boldsymbol{X}, \boldsymbol{y}] = m(\boldsymbol{x}_*) = k(\boldsymbol{X}, \boldsymbol{x}_*)^\top k(\boldsymbol{X}, \boldsymbol{X})^{-1}\boldsymbol{y}$$
$$\mathbb{V}[f(\boldsymbol{x}_*)|\boldsymbol{x}_*, \boldsymbol{X}, \boldsymbol{y}] = \sigma^2(\boldsymbol{x}_*) = k(\boldsymbol{x}_*, \boldsymbol{x}_*) - k(\boldsymbol{X}, \boldsymbol{x}_*)^\top k(\boldsymbol{X}, \boldsymbol{X})^{-1} k(\boldsymbol{X}, \boldsymbol{x}_*)$$



Posterior belief about the function

Predictive (marginal) mean and variance:

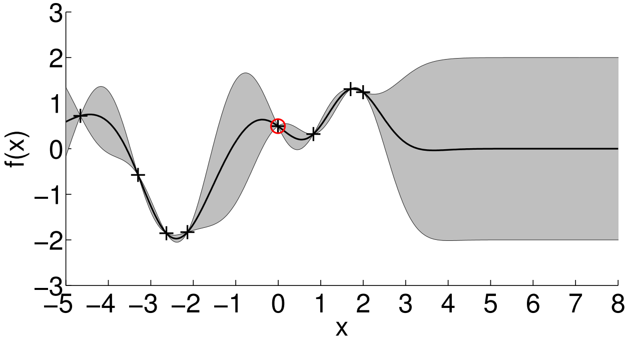
$$\mathbb{E}[f(\boldsymbol{x}_*)|\boldsymbol{x}_*, \boldsymbol{X}, \boldsymbol{y}] = m(\boldsymbol{x}_*) = k(\boldsymbol{X}, \boldsymbol{x}_*)^\top k(\boldsymbol{X}, \boldsymbol{X})^{-1}\boldsymbol{y}$$
$$\mathbb{V}[f(\boldsymbol{x}_*)|\boldsymbol{x}_*, \boldsymbol{X}, \boldsymbol{y}] = \sigma^2(\boldsymbol{x}_*) = k(\boldsymbol{x}_*, \boldsymbol{x}_*) - k(\boldsymbol{X}, \boldsymbol{x}_*)^\top k(\boldsymbol{X}, \boldsymbol{X})^{-1} k(\boldsymbol{X}, \boldsymbol{x}_*)$$



Posterior belief about the function

Predictive (marginal) mean and variance:

$$\mathbb{E}[f(\boldsymbol{x}_*)|\boldsymbol{x}_*, \boldsymbol{X}, \boldsymbol{y}] = m(\boldsymbol{x}_*) = k(\boldsymbol{X}, \boldsymbol{x}_*)^\top k(\boldsymbol{X}, \boldsymbol{X})^{-1}\boldsymbol{y}$$
$$\mathbb{V}[f(\boldsymbol{x}_*)|\boldsymbol{x}_*, \boldsymbol{X}, \boldsymbol{y}] = \sigma^2(\boldsymbol{x}_*) = k(\boldsymbol{x}_*, \boldsymbol{x}_*) - k(\boldsymbol{X}, \boldsymbol{x}_*)^\top k(\boldsymbol{X}, \boldsymbol{X})^{-1} k(\boldsymbol{X}, \boldsymbol{x}_*)$$

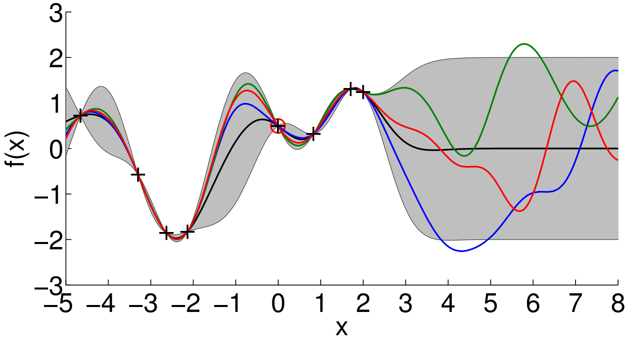


Posterior belief about the function

Predictive (marginal) mean and variance:

$$\mathbb{E}[f(\boldsymbol{x}_*)|\boldsymbol{x}_*, \boldsymbol{X}, \boldsymbol{y}] = m(\boldsymbol{x}_*) = k(\boldsymbol{X}, \boldsymbol{x}_*)^\top k(\boldsymbol{X}, \boldsymbol{X})^{-1}\boldsymbol{y}$$

$$\mathbb{V}[f(\boldsymbol{x}_*)|\boldsymbol{x}_*, \boldsymbol{X}, \boldsymbol{y}] = \sigma^2(\boldsymbol{x}_*) = k(\boldsymbol{x}_*, \boldsymbol{x}_*) - k(\boldsymbol{X}, \boldsymbol{x}_*)^\top k(\boldsymbol{X}, \boldsymbol{X})^{-1} k(\boldsymbol{X}, \boldsymbol{x}_*)$$



Posterior belief about the function

Predictive (marginal) mean and variance:

$$\mathbb{E}[f(\boldsymbol{x}_*)|\boldsymbol{x}_*, \boldsymbol{X}, \boldsymbol{y}] = m(\boldsymbol{x}_*) = k(\boldsymbol{X}, \boldsymbol{x}_*)^\top k(\boldsymbol{X}, \boldsymbol{X})^{-1}\boldsymbol{y}$$
$$\mathbb{V}[f(\boldsymbol{x}_*)|\boldsymbol{x}_*, \boldsymbol{X}, \boldsymbol{y}] = \sigma^2(\boldsymbol{x}_*) = k(\boldsymbol{x}_*, \boldsymbol{x}_*) - k(\boldsymbol{X}, \boldsymbol{x}_*)^\top k(\boldsymbol{X}, \boldsymbol{X})^{-1} k(\boldsymbol{X}, \boldsymbol{x}_*)$$

Objective

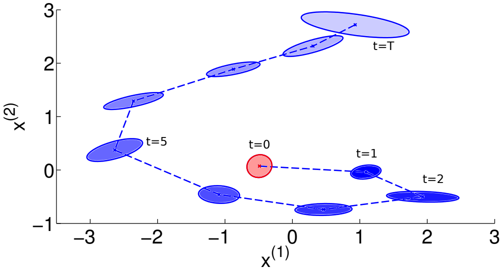
Minimize expected long-term cost $J(\boldsymbol{\theta}) = \sum_t \mathbb{E}[c(\boldsymbol{x}_t)|\boldsymbol{\theta}]$

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*

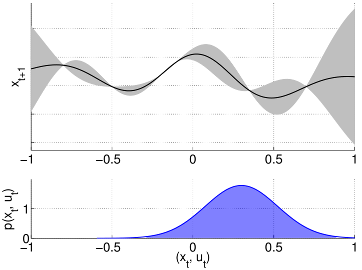
Long-Term Predictions



- Iteratively compute $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \ldots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$

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

Long-Term Predictions

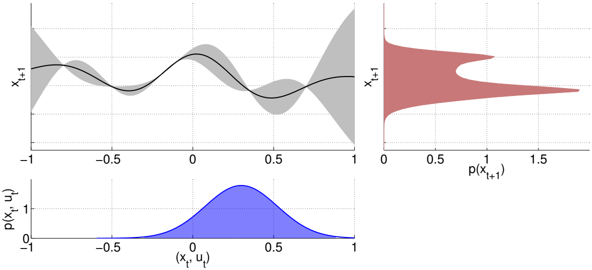


- Iteratively compute $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \ldots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$

$$\underbrace{\boxed{p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, \boldsymbol{u}_t)}}_{\text{GP prediction}} \underbrace{\boxed{p(\boldsymbol{x}_t, \boldsymbol{u}_t|\boldsymbol{\theta})}}_{\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})}$$

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

Long-Term Predictions

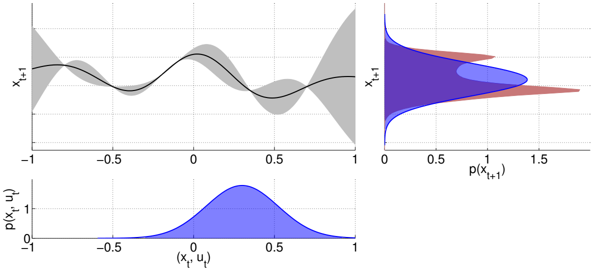


- Iteratively compute $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \ldots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$

$$p(\boldsymbol{x}_{t+1}|\boldsymbol{\theta}) = \iiint \underbrace{p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, \boldsymbol{u}_t)}_{\text{GP prediction}} \underbrace{p(\boldsymbol{x}_t, \boldsymbol{u}_t|\boldsymbol{\theta})}_{\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})} \, df \, d\boldsymbol{x}_t \, d\boldsymbol{u}_t$$

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

Long-Term Predictions



- Iteratively compute $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \dots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$

$$\boxed{p(\boldsymbol{x}_{t+1}|\boldsymbol{\theta})} = \iiint \underbrace{\boxed{p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, \boldsymbol{u}_t)}}_{\text{GP prediction}} \underbrace{\boxed{p(\boldsymbol{x}_t, \boldsymbol{u}_t|\boldsymbol{\theta})}}_{\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})} \, df \, d\boldsymbol{x}_t \, d\boldsymbol{u}_t$$

▶▶ GP moment matching
(Girard et al., 2002; Quiñonero-Candela et al., 2003)

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 $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \dots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$
3. **Policy improvement**
   - Compute expected long-term cost $J(\boldsymbol{\theta})$
   - Find parameters $\boldsymbol{\theta}$ that minimize $J(\boldsymbol{\theta})$
4. Apply controller

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

Policy Improvement

Objective

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

- Know how to predict $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \ldots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$

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

Policy Improvement

Objective

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

- Know how to predict $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \dots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$
- Compute

$$\mathbb{E}[c(\boldsymbol{x}_t)|\boldsymbol{\theta}] = \int c(\boldsymbol{x}_t)\mathcal{N}\big(\boldsymbol{x}_t \,|\, \boldsymbol{\mu}_t,\, \boldsymbol{\Sigma}_t\big)d\boldsymbol{x}_t\,, \quad t = 1, \dots, T\,,$$

and sum them up to obtain $J(\boldsymbol{\theta})$

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

Policy Improvement

Objective

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

- Know how to predict $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \dots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$
- Compute

$$\mathbb{E}[c(\boldsymbol{x}_t)|\boldsymbol{\theta}] = \int c(\boldsymbol{x}_t)\mathcal{N}\big(\boldsymbol{x}_t \,|\, \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t\big)d\boldsymbol{x}_t\,, \quad t = 1, \dots, T\,,$$

  and sum them up to obtain $J(\boldsymbol{\theta})$
- Analytically compute gradient $\mathrm{d}J(\boldsymbol{\theta})/\mathrm{d}\boldsymbol{\theta}$
- Standard gradient-based optimizer (e.g., BFGS) to find $\boldsymbol{\theta}^*$

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

Fast Reinforcement Learning

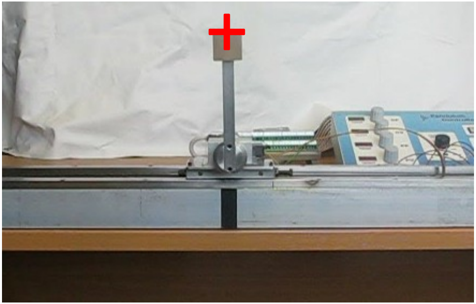
Objective

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 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**

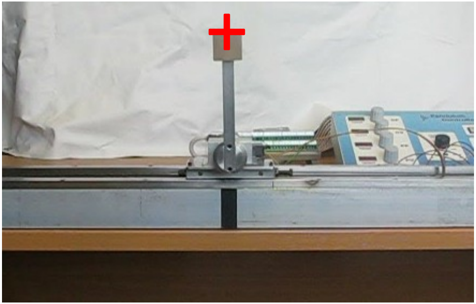
Deisenroth et al. (IEEE-TPAMI, 2015): *Gaussian Processes for Data-Efficient Learning in Robotics and Control*

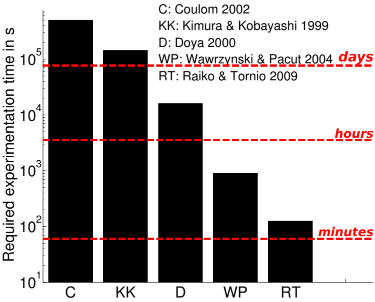


- Swing up and balance a freely swinging pendulum on a cart
- No knowledge about nonlinear dynamics ▶▶ Learn from scratch
- Cost function $c(\boldsymbol{x}) = 1 - \exp(-\|\boldsymbol{x} - \boldsymbol{x}_{\text{target}}\|^2)$
- Code: https://github.com/ICL-SML/pilco-matlab

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

Standard Benchmark: Cart-Pole Swing-up

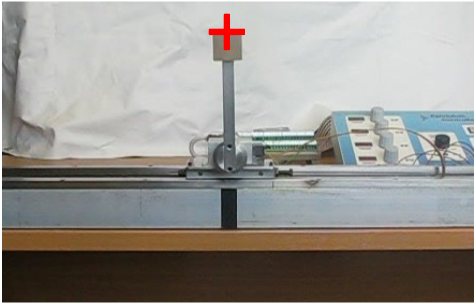


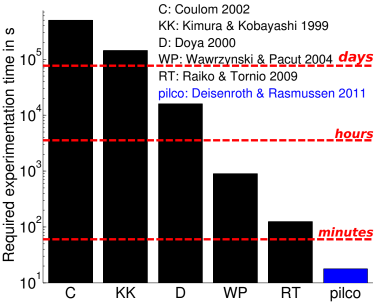


- Swing up and balance a freely swinging pendulum on a cart
- No knowledge about nonlinear dynamics ⏩ Learn from scratch
- Cost function $c(\boldsymbol{x}) = 1 - \exp(-\|\boldsymbol{x} - \boldsymbol{x}_{\text{target}}\|^2)$
- Code: `https://github.com/ICL-SML/pilco-matlab`

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

Standard Benchmark: Cart-Pole Swing-up

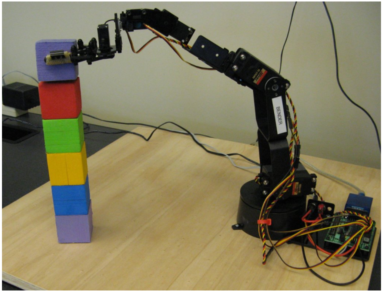




- Swing up and balance a freely swinging pendulum on a cart
- No knowledge about nonlinear dynamics ⏩ Learn from scratch
- Cost function $c(\boldsymbol{x}) = 1 - \exp(-\|\boldsymbol{x} - \boldsymbol{x}_{\text{target}}\|^2)$
- **Unprecedented learning speed** compared to state-of-the-art
- Code: `https://github.com/ICL-SML/pilco-matlab`

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

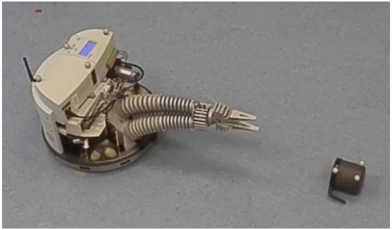
Learning to Control an Off-the-Shelf Robot



- Autonomously learn block-stacking with a low-cost robot
- Kinect camera as only sensor
- Robot very noisy
- Learn forward model and controller from scratch
- Small number of interactions: **Robot wears out quickly**

Deisenroth et al. (RSS, 2011): *Learning to Control a Low-Cost Manipulator using Data-efficient Reinforcement Learning*

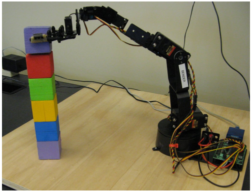
Learning to Pick up Objects



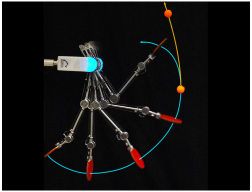
- Robotino XT (compliant behavior)
- Omnidirectional platform with pneumatic arm/trunk
- Motion capture
- 9D states, 9D actions

Bischoff et al. (ICRA, 2014): *Policy Search For Learning Robot Control Using Sparse Data*

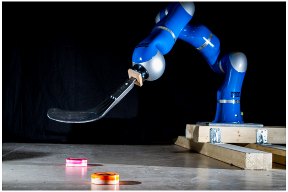
Wide Applicability



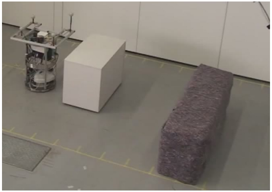
with D Fox



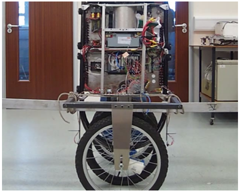
with P Englert, A Paraschos, J Peters



with A Kupcsik, J Peters, G Neumann



B Bischoff (Bosch), ESANN 2013



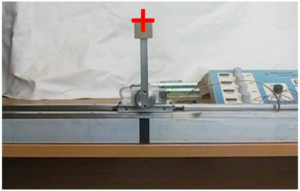
A McHutchon (U Cambridge)

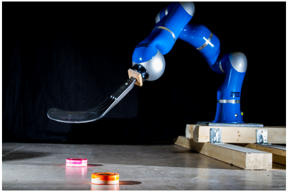
B Bischoff (Bosch), ECML 2013

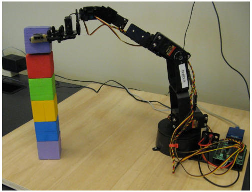
- **Application to a wide range of robotic systems**

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)

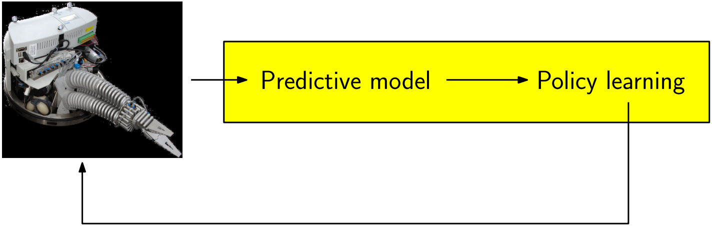






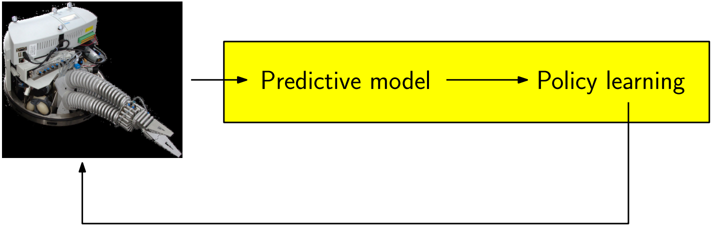
- In robotics, data-efficient learning is critical
- Probabilistic, model-based RL approach
  - Reduce model bias
  - Unprecedented learning speed
  - Wide applicability

Properties



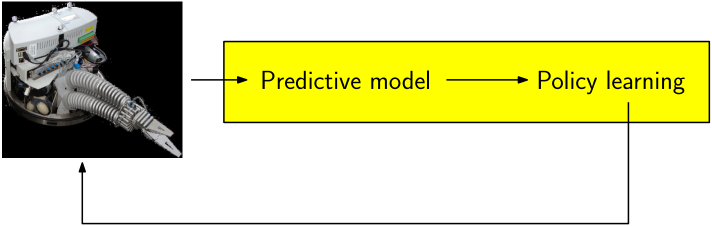
- We can learn flexible (low-level) policies with thousands of parameters

Properties



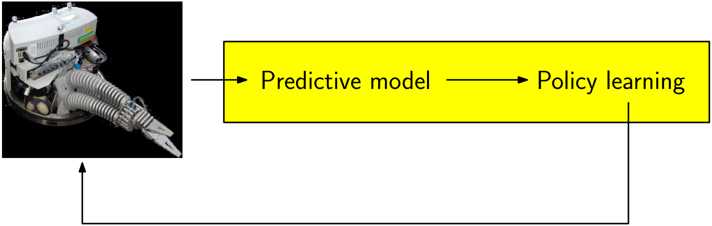
- We can learn flexible (low-level) policies with thousands of parameters
- Critical assumption: Learn a good model of the robot's dynamics ⏩ Smoothness

Properties



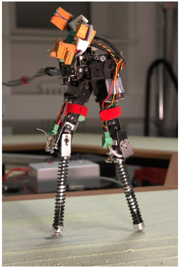
- We can learn flexible (low-level) policies with thousands of parameters
- Critical assumption: Learn a good model of the robot's dynamics ⏩ Smoothness
- Sometimes this assumption is unrealistic

Properties



- We can learn flexible (low-level) policies with thousands of parameters
- Critical assumption: Learn a good model of the robot's dynamics ⏩ Smoothness
- Sometimes this assumption is unrealistic

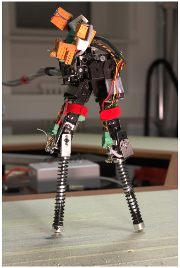
⏩ Alternative approach to data-efficient controller learning



- Learning forward models is not always easy
- E.g., Ground contacts in legged locomotion

Objective

Find parameters $\boldsymbol{\theta}$ of controller $\pi(\boldsymbol{\theta})$



- Learning forward models is not always easy
- E.g., Ground contacts in legged locomotion

Objective

Find parameters $\boldsymbol{\theta}$ of controller $\pi(\boldsymbol{\theta})$

Challenges:

- No forward model
- No analytic cost function, no demonstrations
- Still need to be data efficient (fragile robot)
- Manual parameter search can be tedious

▶▶ Bayesian optimization (e.g., Jones 1998; Osborne et al., 2009)

Bayesian Optimization

Objective (Bayesian Optimization)

Minimize an objective function g, which is very expensive to evaluate

Objective (Bayesian Optimization)

Minimize an objective function g, which is very expensive to evaluate

- Computations are cheap, experiments are expensive

Objective (Bayesian Optimization)

Minimize an objective function g, which is very expensive to evaluate

- Computations are cheap, experiments are expensive

Key Idea:

1. Build a model $\tilde{g}$ of the true objective function g
2. Find $\boldsymbol{\theta}^* \in \arg\min_{\boldsymbol{\theta}} \tilde{g}(\boldsymbol{\theta})$
3. Evaluate true objective g at $\boldsymbol{\theta}^*$
4. Update model $\tilde{g}$

Objective (Bayesian Optimization)

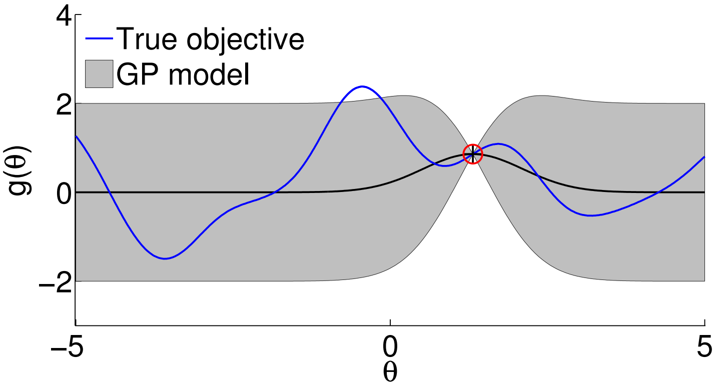
Minimize an objective function g, which is very expensive to evaluate

- Computations are cheap, experiments are expensive

Key Idea:

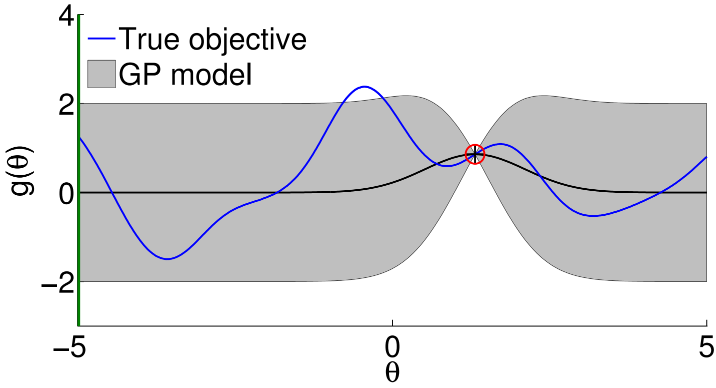
1. Build a model $\tilde{g}$ of the true objective function g
2. Find $\boldsymbol{\theta}^* \in \arg\min_{\boldsymbol{\theta}} \tilde{g}(\boldsymbol{\theta})$
3. Evaluate true objective g at $\boldsymbol{\theta}^*$
4. Update model $\tilde{g}$

- Standard model $\tilde{g}$ is a Gaussian process



- Upper-Confidence-Bound (UCB) criterion to select next point

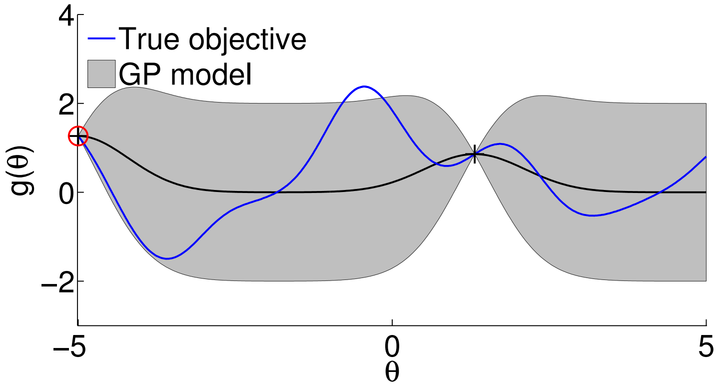
$$\theta^* \in \arg\min_{\theta} \quad \mathbb{E}[\tilde{g}(\theta)] - 2\sqrt{\mathbb{V}[\tilde{g}(\theta)]}$$



- Upper-Confidence-Bound (UCB) criterion to select next point

$$\theta^* \in \arg\min_{\theta} \quad \mathbb{E}[\tilde{g}(\theta)] - 2\sqrt{\mathbb{V}[\tilde{g}(\theta)]}$$

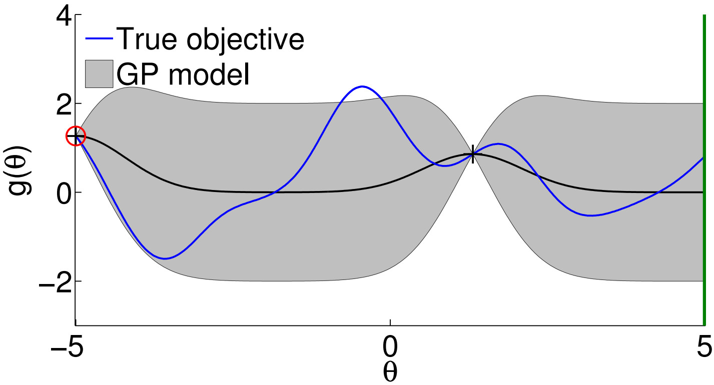
Bayesian Optimization: Illustration



- Upper-Confidence-Bound (UCB) criterion to select next point

$$\theta^* \in \arg\min_{\theta} \quad \mathbb{E}[\tilde{g}(\theta)] - 2\sqrt{\mathbb{V}[\tilde{g}(\theta)]}$$

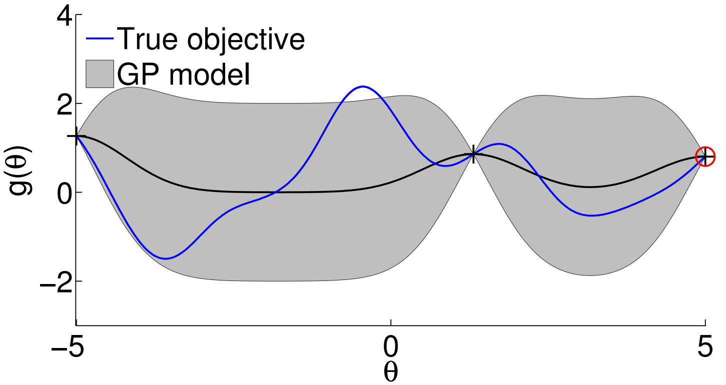
Bayesian Optimization: Illustration



- Upper-Confidence-Bound (UCB) criterion to select next point

$$\theta^* \in \arg\min_{\theta} \quad \mathbb{E}[\tilde{g}(\theta)] - 2\sqrt{\mathbb{V}[\tilde{g}(\theta)]}$$

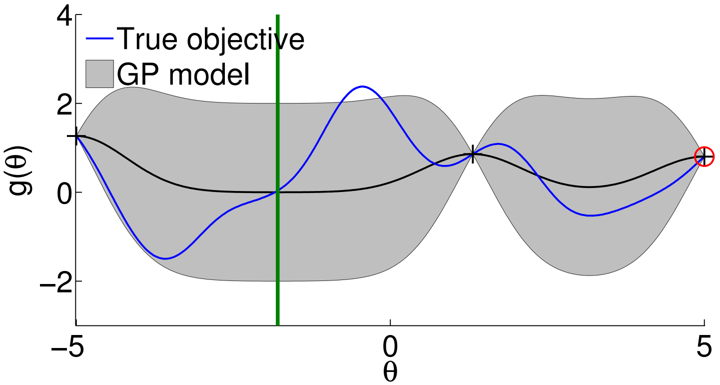
Bayesian Optimization: Illustration



- Upper-Confidence-Bound (UCB) criterion to select next point

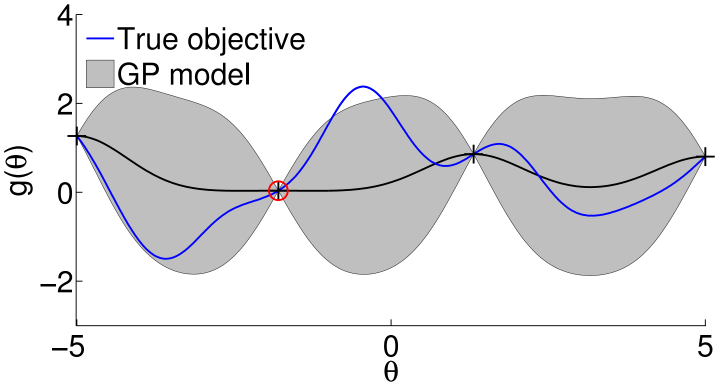
$$\theta^* \in \arg\min_{\theta} \quad \mathbb{E}[\tilde{g}(\theta)] - 2\sqrt{\mathbb{V}[\tilde{g}(\theta)]}$$

Bayesian Optimization: Illustration



- Upper-Confidence-Bound (UCB) criterion to select next point

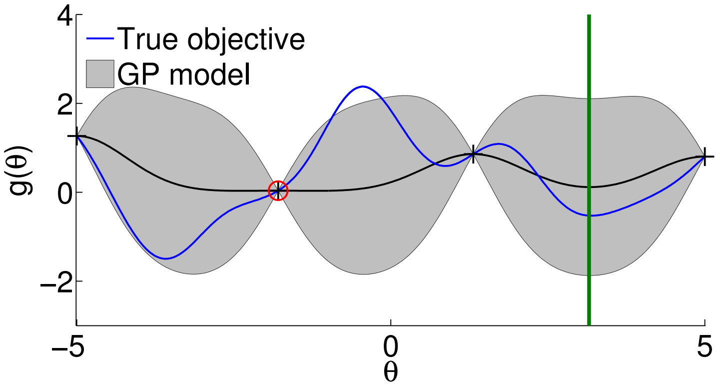
$$\theta^* \in \arg\min_{\theta} \quad \mathbb{E}[\tilde{g}(\theta)] - 2\sqrt{\mathbb{V}[\tilde{g}(\theta)]}$$



- Upper-Confidence-Bound (UCB) criterion to select next point

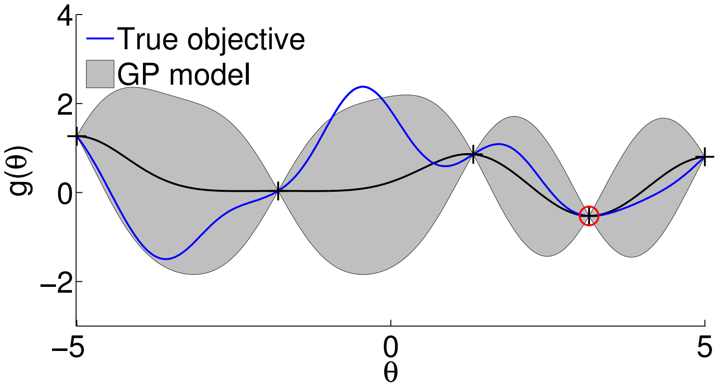
$$\theta^* \in \arg\min_{\theta} \quad \mathbb{E}[\tilde{g}(\theta)] - 2\sqrt{\mathbb{V}[\tilde{g}(\theta)]}$$

Bayesian Optimization: Illustration



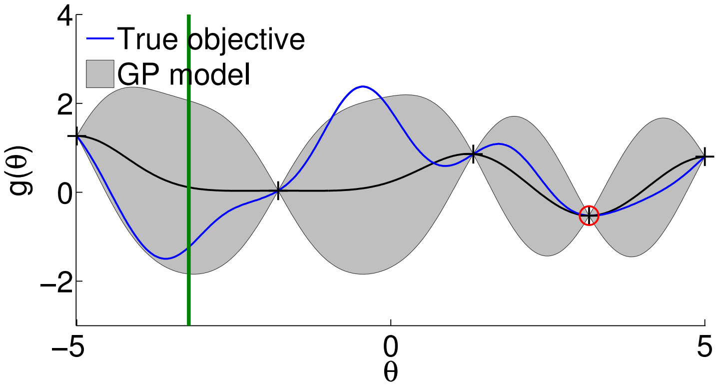
- Upper-Confidence-Bound (UCB) criterion to select next point

$$\theta^* \in \arg\min_{\theta} \quad \mathbb{E}[\tilde{g}(\theta)] - 2\sqrt{\mathbb{V}[\tilde{g}(\theta)]}$$



- Upper-Confidence-Bound (UCB) criterion to select next point

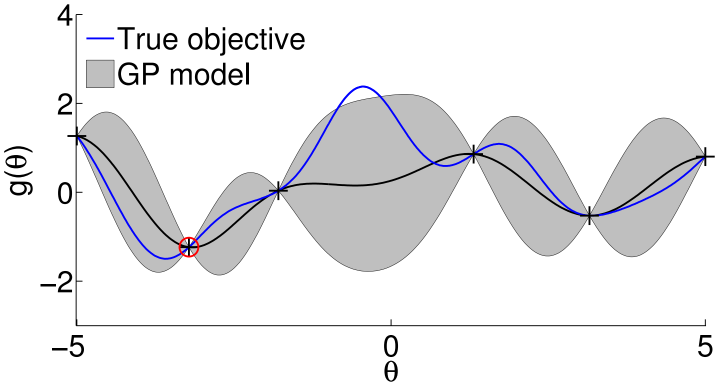
$$\theta^* \in \arg\min_{\theta} \quad \mathbb{E}[\tilde{g}(\theta)] - 2\sqrt{\mathbb{V}[\tilde{g}(\theta)]}$$



- Upper-Confidence-Bound (UCB) criterion to select next point

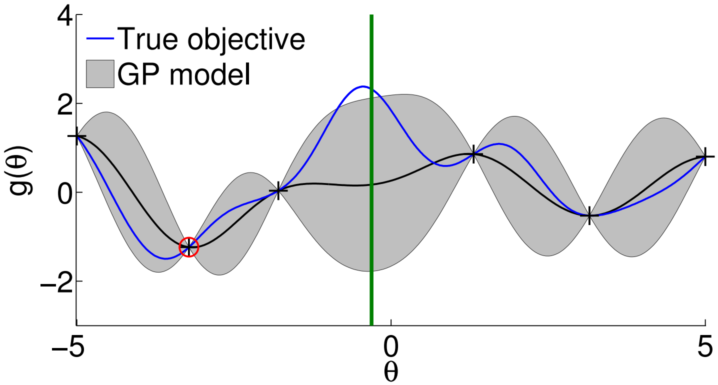
$$\theta^* \in \arg\min_{\theta} \quad \mathbb{E}[\tilde{g}(\theta)] - 2\sqrt{\mathbb{V}[\tilde{g}(\theta)]}$$

Bayesian Optimization: Illustration



- Upper-Confidence-Bound (UCB) criterion to select next point

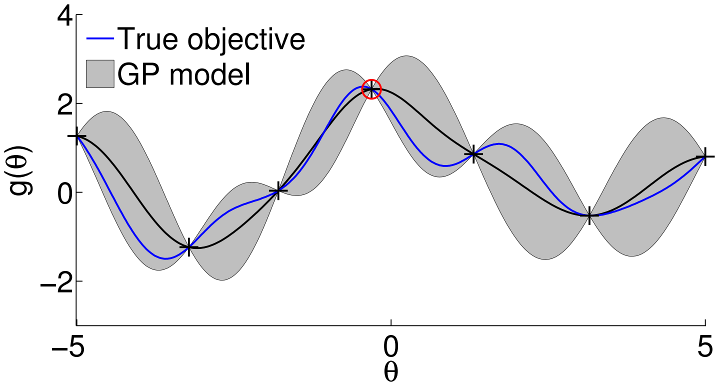
$$\theta^* \in \arg\min_{\theta} \quad \mathbb{E}[\tilde{g}(\theta)] - 2\sqrt{\mathbb{V}[\tilde{g}(\theta)]}$$



- Upper-Confidence-Bound (UCB) criterion to select next point

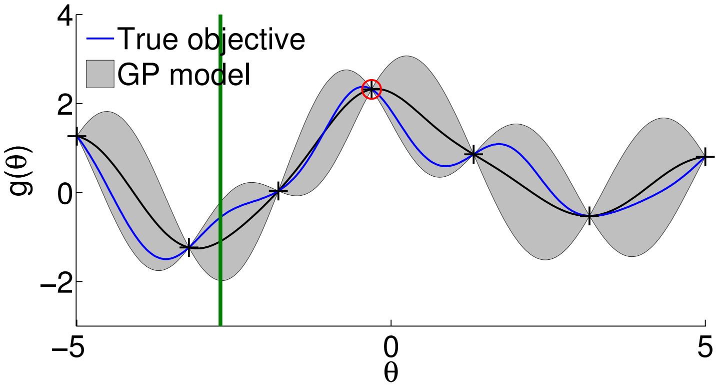
$$\theta^* \in \arg\min_{\theta} \quad \mathbb{E}[\tilde{g}(\theta)] - 2\sqrt{\mathbb{V}[\tilde{g}(\theta)]}$$

Bayesian Optimization: Illustration



- Upper-Confidence-Bound (UCB) criterion to select next point

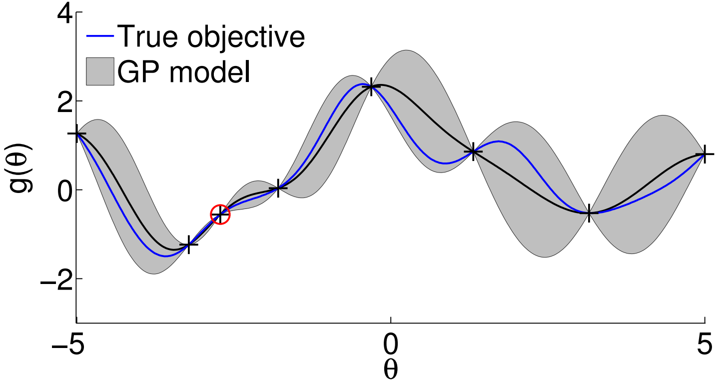
$$\theta^* \in \arg\min_{\theta} \quad \mathbb{E}[\tilde{g}(\theta)] - 2\sqrt{\mathbb{V}[\tilde{g}(\theta)]}$$



- Upper-Confidence-Bound (UCB) criterion to select next point

$$\theta^* \in \arg\min_{\theta} \quad \mathbb{E}[\tilde{g}(\theta)] - 2\sqrt{\mathbb{V}[\tilde{g}(\theta)]}$$

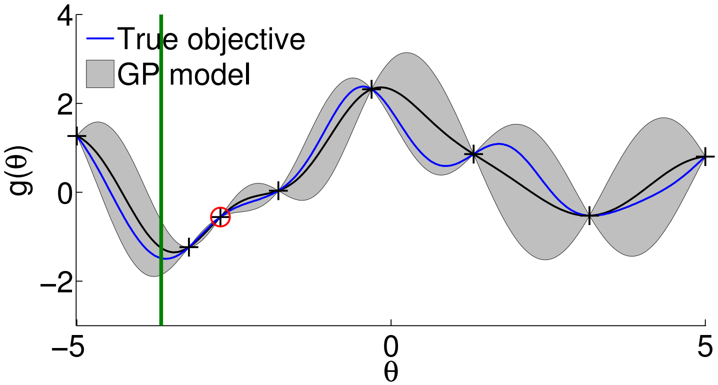
Bayesian Optimization: Illustration



- Upper-Confidence-Bound (UCB) criterion to select next point

$$\theta^* \in \arg\min_{\theta} \quad \mathbb{E}[\tilde{g}(\theta)] - 2\sqrt{\mathbb{V}[\tilde{g}(\theta)]}$$

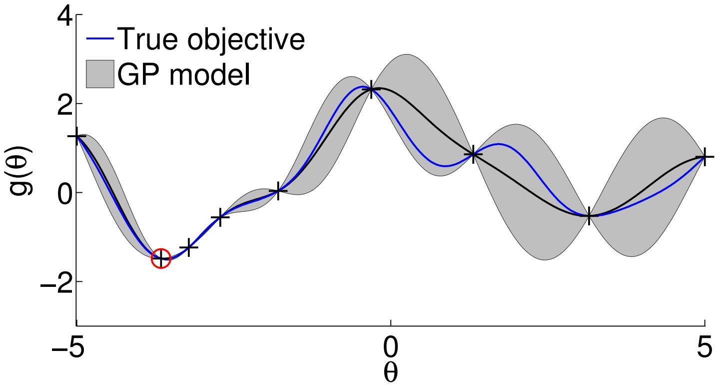
Bayesian Optimization: Illustration



- Upper-Confidence-Bound (UCB) criterion to select next point

$$\theta^* \in \arg\min_{\theta} \quad \mathbb{E}[\tilde{g}(\theta)] - 2\sqrt{\mathbb{V}[\tilde{g}(\theta)]}$$

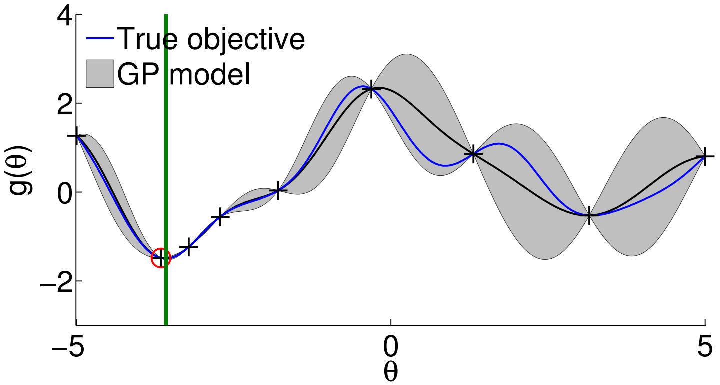
Bayesian Optimization: Illustration



- Upper-Confidence-Bound (UCB) criterion to select next point

$$\theta^* \in \arg\min_{\theta} \quad \mathbb{E}[\tilde{g}(\theta)] - 2\sqrt{\mathbb{V}[\tilde{g}(\theta)]}$$

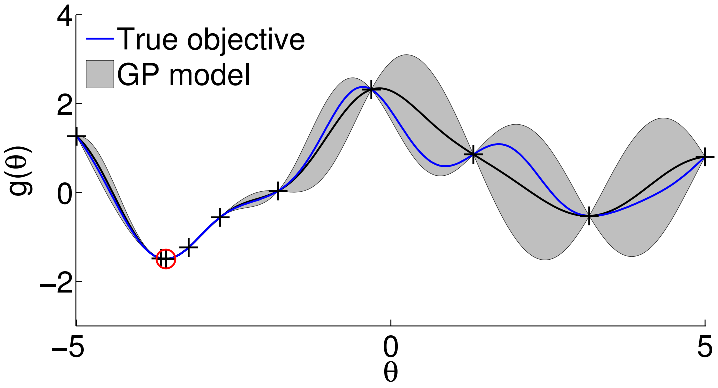
Bayesian Optimization: Illustration



- Upper-Confidence-Bound (UCB) criterion to select next point

$$\theta^* \in \arg\min_{\theta} \quad \mathbb{E}[\tilde{g}(\theta)] - 2\sqrt{\mathbb{V}[\tilde{g}(\theta)]}$$

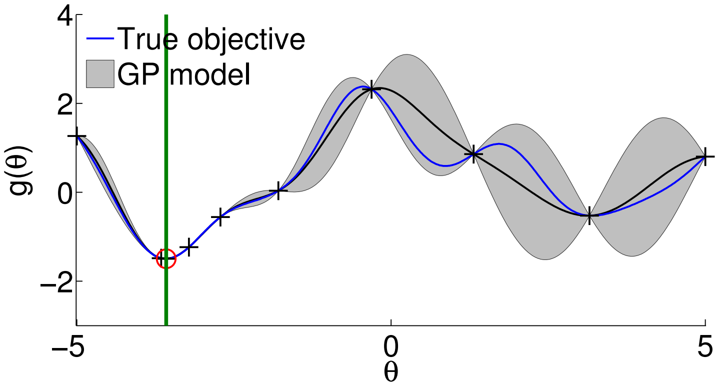
Bayesian Optimization: Illustration



- Upper-Confidence-Bound (UCB) criterion to select next point

$$\theta^* \in \arg\min_{\theta} \quad \mathbb{E}[\tilde{g}(\theta)] - 2\sqrt{\mathbb{V}[\tilde{g}(\theta)]}$$

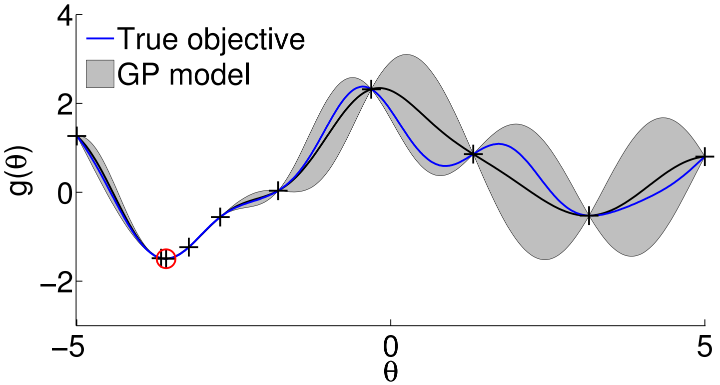
Bayesian Optimization: Illustration



- Upper-Confidence-Bound (UCB) criterion to select next point

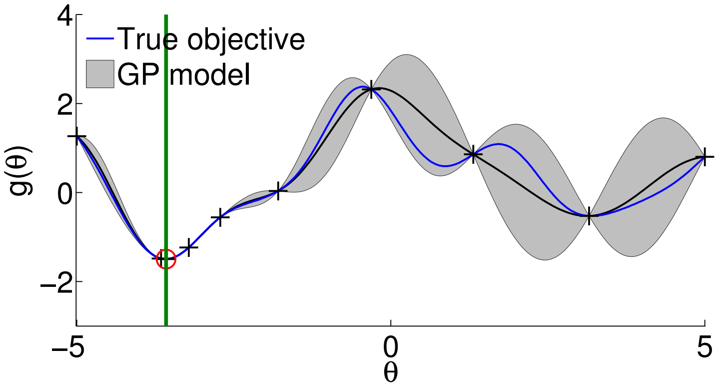
$$\theta^* \in \arg\min_{\theta} \quad \mathbb{E}[\tilde{g}(\theta)] - 2\sqrt{\mathbb{V}[\tilde{g}(\theta)]}$$

Bayesian Optimization: Illustration



- Upper-Confidence-Bound (UCB) criterion to select next point

$$\theta^* \in \arg\min_{\theta} \quad \mathbb{E}[\tilde{g}(\theta)] - 2\sqrt{\mathbb{V}[\tilde{g}(\theta)]}$$



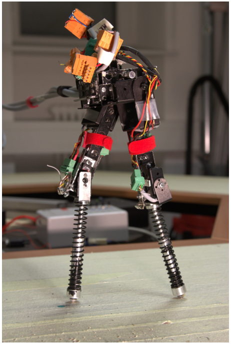
- Upper-Confidence-Bound (UCB) criterion to select next point

$$\theta^* \in \arg\min_{\theta} \quad \mathbb{E}[\tilde{g}(\theta)] - 2\sqrt{\mathbb{V}[\tilde{g}(\theta)]}$$

- Global minimum found after 10 function evaluations

Bayesian Gait Optimization for Legged Robots

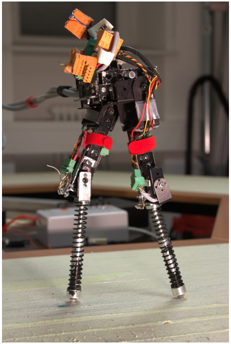
- Fragile biped
  - ⏩ Only few experiments feasible
- Maximize robustness and walking speed
- 4 motors:
  2 actuated hips + 2 actuated knees
- Controller implemented as a
  finite-state-machine (8 parameters)



Calandra et al. (LION, 2014): *Bayesian Gait Optimization for Bipedal Locomotion*
Calandra et al. (ICRA, 2014): *An Experimental Evaluation of Bayesian Optimization on Bipedal Locomotion*

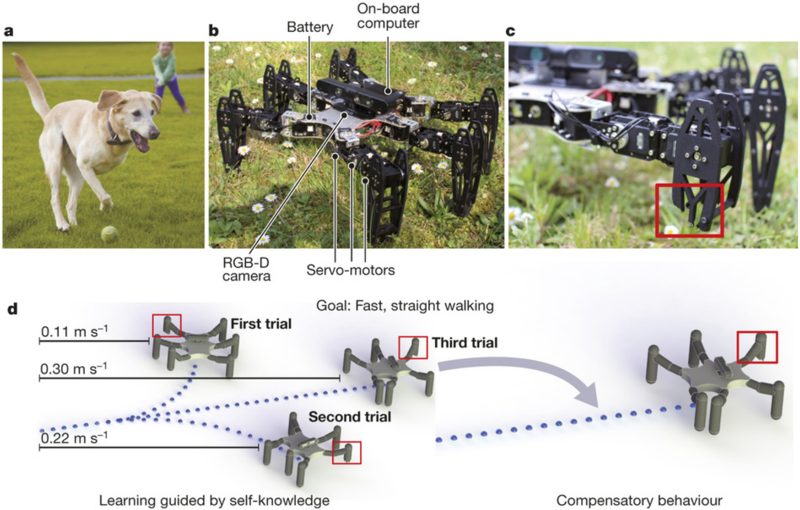
Bayesian Gait Optimization for Legged Robots

- Fragile biped
  - Only few experiments feasible
- Maximize robustness and walking speed
- 4 motors:
  2 actuated hips + 2 actuated knees
- Controller implemented as a finite-state-machine (8 parameters)
- Good parameters found after 80–100 experiments
- Substantial speed-up compared to manual parameter search



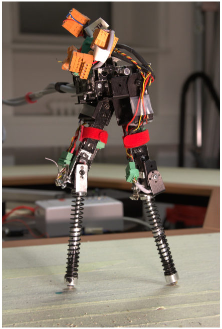
Calandra et al. (LION, 2014): *Bayesian Gait Optimization for Bipedal Locomotion*
Calandra et al. (ICRA, 2014): *An Experimental Evaluation of Bayesian Optimization on Bipedal Locomotion*

Robots That Learn to Recover from Damage



Cully et al. (2015)

Properties



Bayesian Optimization for Control

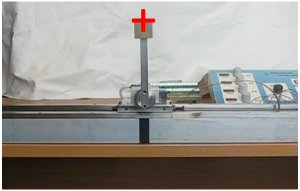
- Bayesian optimization for learning controllers in a few experiments
- General framework:
  No assumptions on dynamics, no explicit cost required
- Limited to few parameters ($\approx$ 10–20)

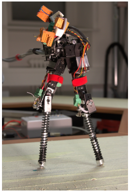
Two Approaches for Data-Efficient Learning

	Cost	Dynamics model	Policy learning	# Parameters
RL	✓	✓	✓	$\leqslant 10,000$
BO	✗	✗	✓	$\leqslant 20$

- If a good dynamics model can be learned and a cost function can be defined, RL-based methods provide more flexibility
- Bayesian optimization is a more general/easier framework for learning a few parameters, but it does not scale to many parameters

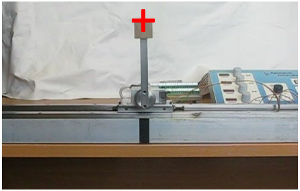
Wrap-up

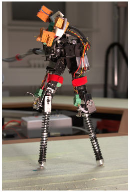




- Autonomous systems take humans out of the loop
- In robotics, data-efficient learning is critical
- Controller learning based using machine learning
  - Reinforcement Learning
  - Bayesian optimization
- Key to success: Uncertainty modeling and exploitation

Wrap-up





- Autonomous systems take humans out of the loop
- In robotics, data-efficient learning is critical
- Controller learning based using machine learning
  - Reinforcement Learning
  - Bayesian optimization
- Key to success: Uncertainty modeling and exploitation

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] B. Bischoff, D. Nguyen-Tuong, H. van Hoof, A. McHutchon, C. E. Rasmussen, A. Knoll, J. Peters, and M. P. Deisenroth. Policy Search For Learning Robot Control Using Sparse Data. In *Proceedings of the IEEE International Conference on Robotics and Automation*, 2014.

[5] E. Brochu, V. M. Cora, and N. de Freitas. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. Technical Report TR-2009-023, Department of Computer Science, University of British Columbia, 2009.

[6] A. Cully, J. Clune, D. Tarapore, and J.-B. Mouret. Robots That Can Adapt Like Animals. *Nature*, 521:503–507, 2015.

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

[8] 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.

[9] 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.

[10] 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*, 2011.

[11] 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.

References II

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

[13] A. Girard, C. E. Rasmussen, and R. Murray-Smith. Gaussian Process Priors with Uncertain Inputs: Multiple-Step Ahead Prediction. Technical Report TR-2002-119, University of Glasgow, 2002.

[14] 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.

[15] M. A. Osborne, R. Garnett, and S. J. Roberts. Gaussian Processes for Global Optimization. In *Proceedings of the International Conference on Learning and Intelligent Optimization*, 2009.

[16] J. Quiñonero-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*, volume 2, pages 701–704, Apr. 2003.

[17] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[18] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. De Freitas. Taking the Human out of the Loop: A Review of Bayesian Optimization. *Proceedings of the IEEE*, 104(1):148–175, 2016.