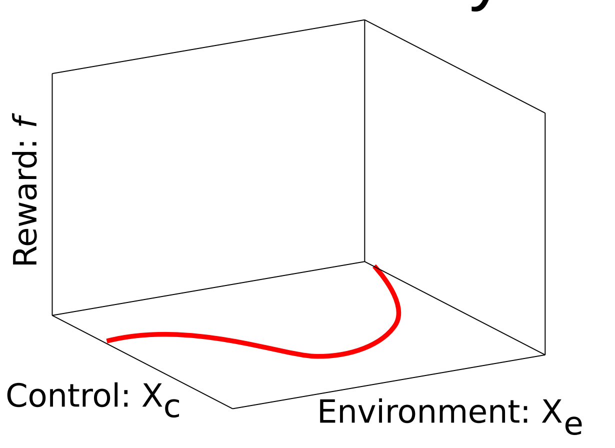
Adapting Control Policies for Expensive Systems to Changing Environments

Matthew Tesch, Jeff Schneider, and Howie Choset
Robotics Institute, Carnegie Mellon University
mtesch@cmu.edu

Policies, Scores, and the Optimal Policy

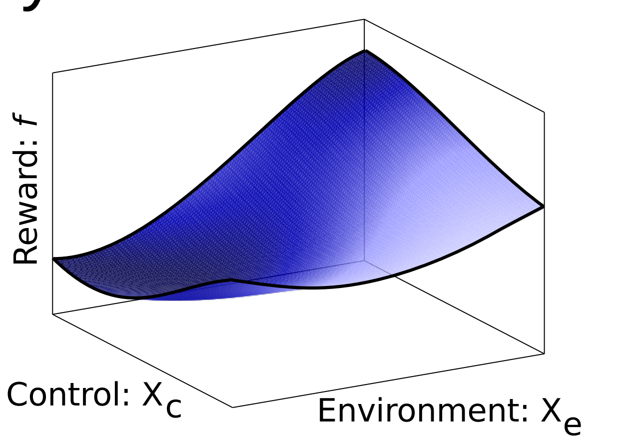
Some systems can operate in a range of conditions; a good controller should adapt to the changes. The goal of this work is to train a good ***control policy*** in the lab -- where environmental conditions can be set -- that has good performance when tested in the field -- where the environment is not controlled.

Control Policy



$\gamma : X_e \rightarrow X_c$

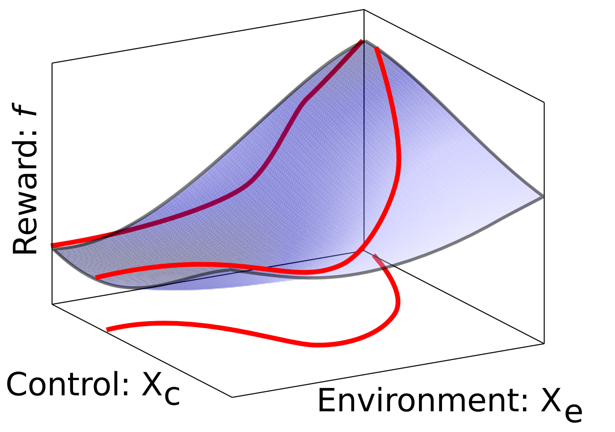
System Performance



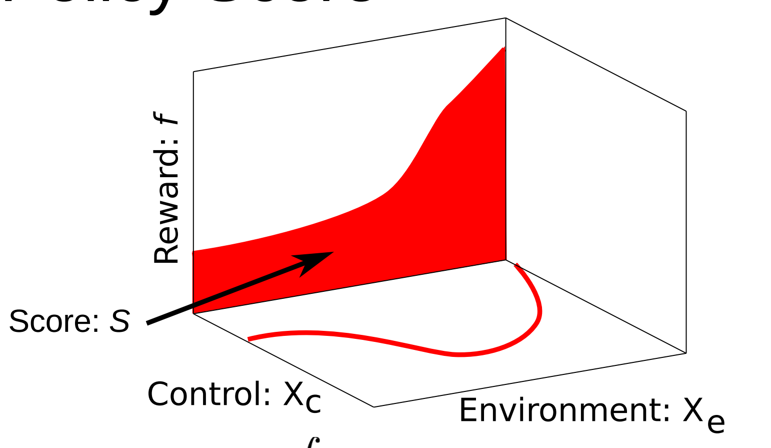
$f : X_e \times X_c \rightarrow \mathbb{R}$

Expensive *f*: limits number of queries

Control Policy Score



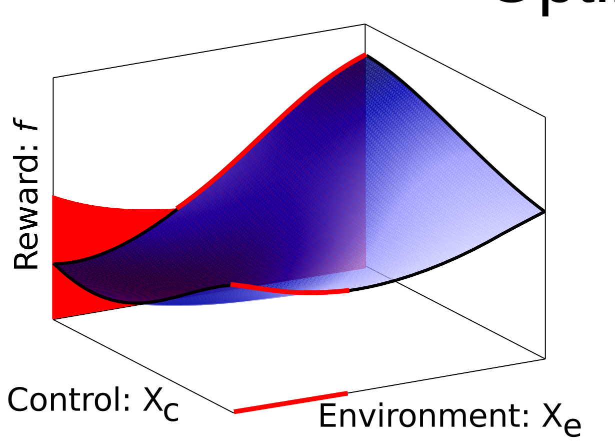
Project policy λ through reward *f*



$$\mathcal{S}(\gamma) = \int_{X_e} \omega(x_e) f(x_e, \gamma(x_e))\, dx_e$$

(expected policy performance)

Optimal Policy



$$\gamma^* = \underset{\gamma}{\operatorname{argmax}}\, \mathcal{S}(\gamma)$$

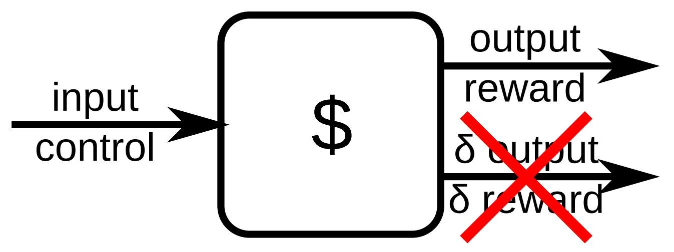
or

$$\gamma^*(x_e) = \underset{x_e \in X_c}{\operatorname{argmax}}\, f(x_e, x_c)$$

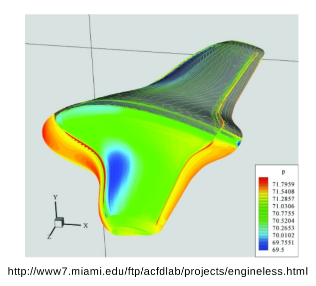
Recall *f* is expensive; can't calculate *S*(λ), let alone directly search for λ*

Expensive Optimization using Surrogate Functions

Expensive Function

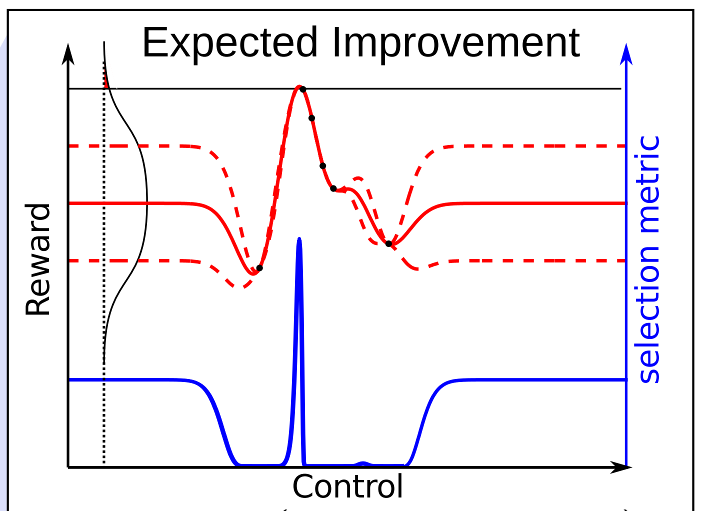


Expensive System Examples



Surrogate Function Method:

1) Initial Objection Samples
2) Fit Function (Gaussian process)
3) Use Fit to Select Next Sample
4) Repeat steps 2 and 3



$$EI(x) = \mathbb{E}\left(\max(y - \max(\tilde{Y}), 0)\right)$$

Problem Definition

Experiment Selection:
Sequentially select and evaluate a series of x^i in $X_e \times X_c$, so as to maximize performance *S* of the selected policy λ.

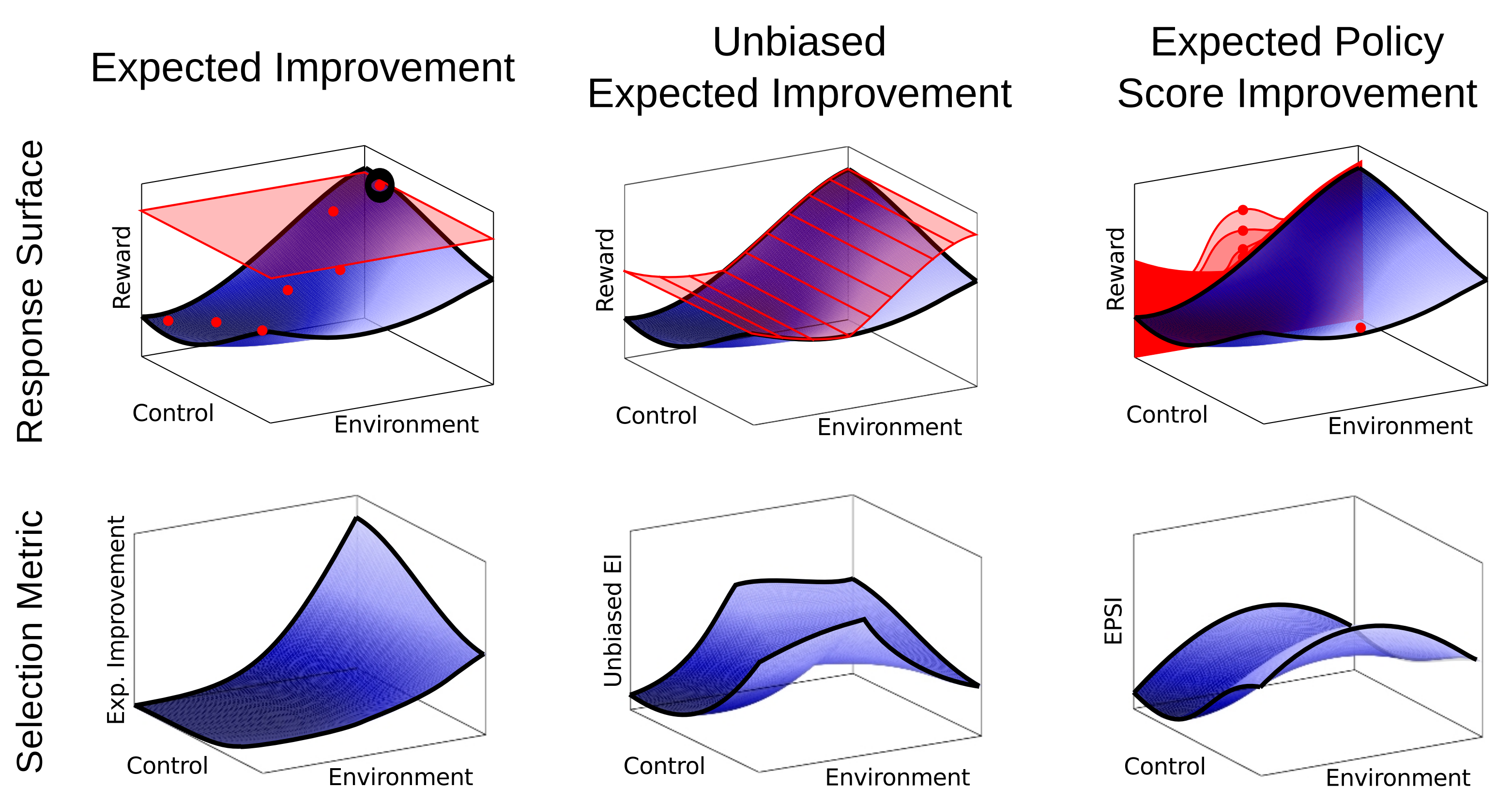
Policy Selection:
Given *n* evaluations of *f*, choose the policy λ with the highest expected performance.

Proposed Methods

Policy Selection

A Gaussian process is fit to the existing data, providing an estimate $\hat{f}$ of the true *f*. This $\hat{f}$ can be searched cheaply to optimize $\hat{S}(\lambda)$; alternatively a lower confidence bound can be used as *f* to produce a more robust policy.

Experiment Selection



$$EI(x) = \int_{\max(\tilde{Y})}^{\infty} (y - \max(\tilde{Y})) p_x(y)\, dy$$

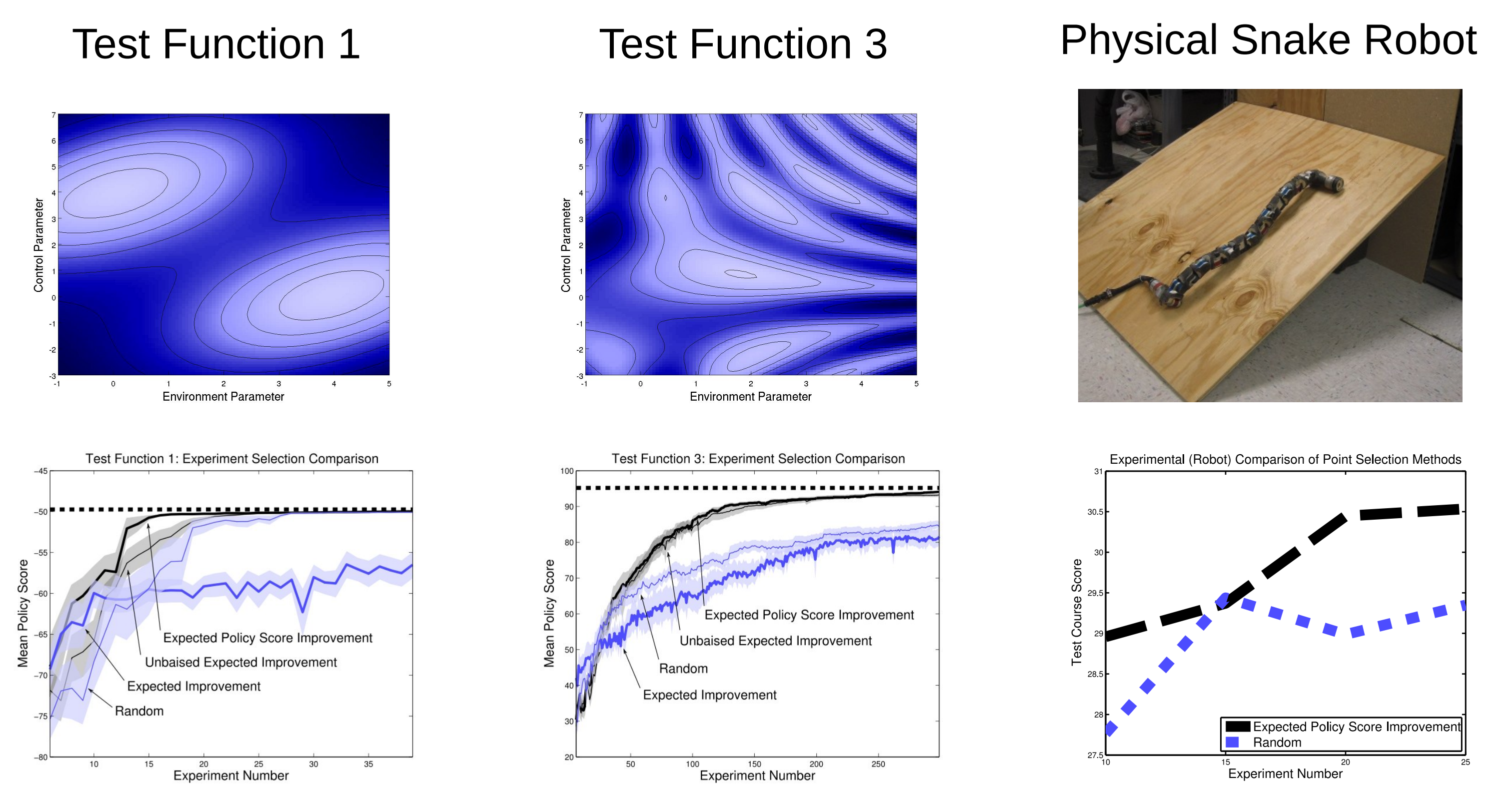
$$y^*_{x_e} = \max_{x_c \in X_c} (\hat{f}(x_c, x_e))$$

$$\text{UEI}(x) = \omega(x_e) \int_{y^*_{x_e}}^{\infty} (y - y^*_{x_e}) p_x(y)\, dy$$

$$\text{imp}^e_y = \max(\hat{f}_y(x_e, \hat{\gamma}^*_y(x_e)) - \hat{f}(x_e, \hat{\gamma}^*(x_e)), 0)$$

$$\text{EPSI}(x) = \int_{-\infty}^{\infty} p_x(y) \int_{X_e} \omega(x_e)\, \text{imp}^e_y\, dx_e\, dy$$

Experimental Results



Empirical Results: Compared score of policies generated from samples selected by each metric; average score over 20 trials for analytic test functions, and over 1 trial for robot test.

M. Tesch et al., Using Response Surfaces and Expected Improvement to Optimize Snake Robot Gait Parameters, *IROS 2011*
M. Tesch et al., Adapting Control Policies for Expensive Systems to Changing Environments, *IROS 2011*
M. Tesch et al., Parameterized and Scripted Gaits for Modular Snake Robots, *Advanced Robotics*

This work would not have been possible without the help of the Biorobotics Lab and Auton Lab at CMU, and the funding of the NDSEG fellowship.