F.L. Lewis
UTA Research Institute (UTARI)
The University of Texas at Arlington, USA

Supported by
NSF, ARO, AFOSR

Optimal Control Introduction

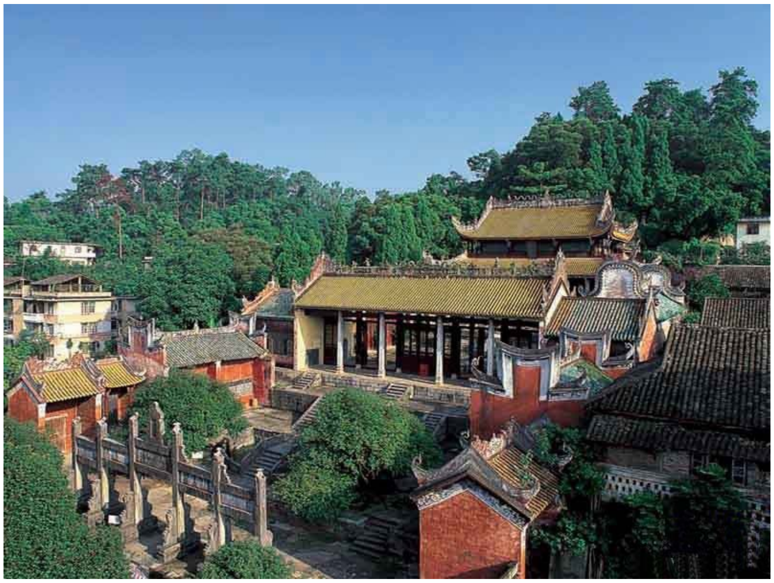
Meng Tz
500 BC

Mencius

He who exerts his mind to the utmost knows nature’s pattern.

The way of learning is none other than finding the lost mind.

Man’s task is to understand patterns in nature and society.



- Optimal Control
- Reinforcement learning
- Control of Robotic Devices
- Adaptive Learning of Human-Robot Interfaces
- Dynamic Games for HRI

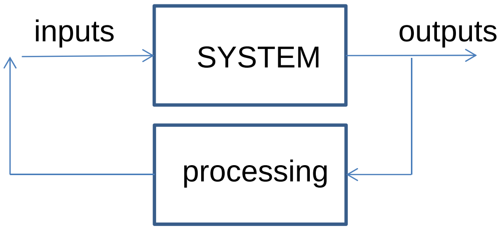
Importance of Feedback Control

Darwin- FB and natural selection
Volterra- FB and fish population balance
Adam Smith- FB and international economy
James Watt- FB and the steam engine
FB and cell homeostasis

The resources available to most species for their survival are meager and limited

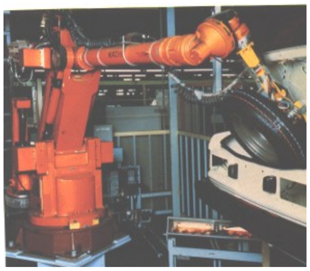
Nature uses Optimal control

Alfred North Whitehead
Von Bertalanffy
Systems Theory 1920s

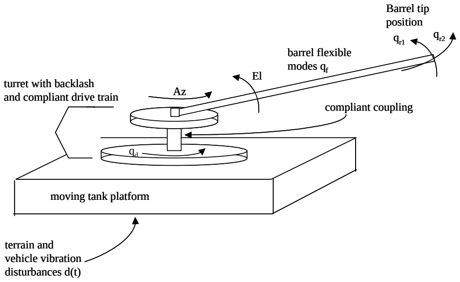


Relevance- Machine Feedback Control

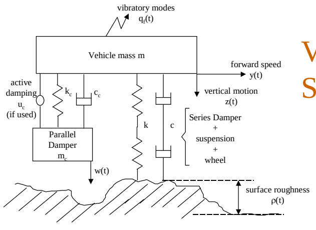
High-Speed Precision Motion Control with unmodeled dynamics, vibration suppression, disturbance rejection, friction compensation, deadzone/backlash control



Industrial Machines



Military Land Systems

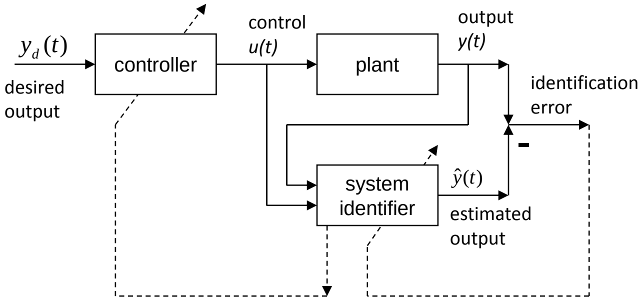


Single-Wheel/Terrain System with Nonlinearities

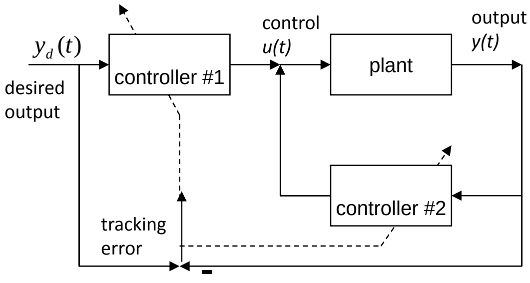
Vehicle Suspension

Aerospace

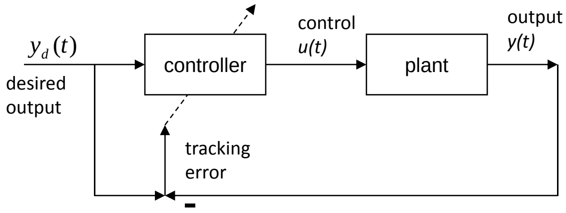
Fixed Controller Topologies



Indirect Scheme



Feedback/Feedforward Scheme

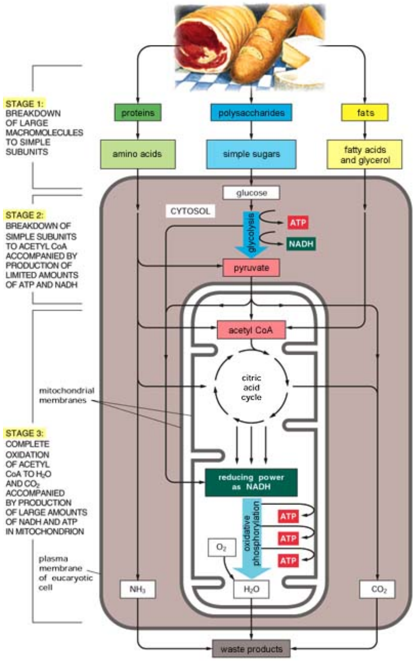


Direct Scheme

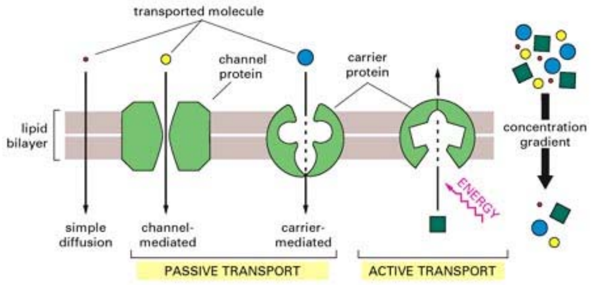
Optimality in Biological Systems

Cell Homeostasis

The individual cell is a complex feedback control system. It pumps ions across the cell membrane to maintain homeostatis, and has only limited energy to do so.



Cellular Metabolism



Permeability control of the cell membrane

http://www.accessexcellence.org/RC/VL/GG/index.html

R. Kalman 1960

Optimality in Control Systems Design

Rocket Orbit Injection

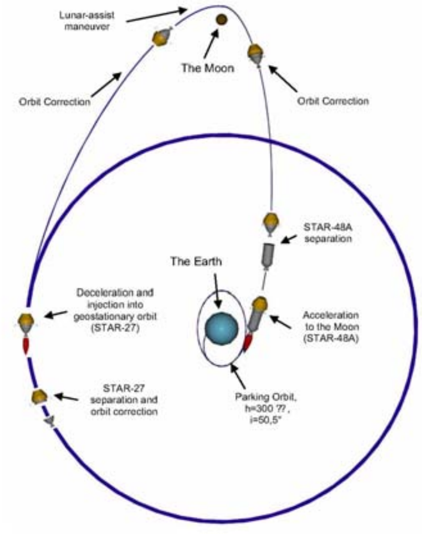


Fig. 1-1. Trajectory scheme

ISC Kosmotras Proprietary 9

http://microsat.sm.bmstu.ru/e-library/Launch/Dnepr_GEO.pdf

Dynamics

$$\dot{r} = w$$

$$\dot{w} = \frac{v^2}{r} - \frac{\mu}{r^2} + \frac{F}{m}\sin\phi$$

$$\dot{v} = \frac{-wv}{r} + \frac{F}{m}\cos\phi$$

$$\dot{m} = -Fm$$

Objectives

- Get to orbit in minimum time
- Use minimum fuel

Optimality and Games

Optimal Control is Effective for:
- Aircraft Autopilots
- Vehicle engine control
- Aerospace Vehicles
- Ship Control
- Industrial Process Control
- Robot Control

Optimal Control:
- Minimum time
- Minimum fuel
- Minimum energy
- Constrained control

Optimal control solutions are found by
- Offline solution of Matrix Design equations
- A full dynamical model of the system is needed

Optimal Control: Linear Quadratic Regulator (LQR)

System $\dot{x} = Ax + Bu$

Performance Index $V(x(t)) = \int_t^\infty (x^T Qx + u^T Ru)\, d\tau \qquad = x^T(t)Px(t)$

Leibniz's formula-

$$\dot{V} = -(x^T Qx + u^T Ru) = \frac{d}{dt}(x^T Px) = \dot{x}^T Px + x^T \dot{P}x + x^T P\dot{x} = (Ax + Bu)^T Px + x^T P(Ax + Bu)$$

Differential equivalent to PI is the Bellman equation

$$0 = H(x, \frac{\partial V}{\partial x}, u) = \dot{V} + x^T Qx + u^T Ru = 2\left(\frac{\partial V}{\partial x}\right)^T \dot{x} + x^T Qx + u^T Ru = 2x^T P(Ax + Bu) + x^T Qx + u^T Ru$$

Hamiltonian function $H(x, \frac{\partial V}{\partial x}, u)$

Stationarity Condition $\frac{d}{du} H(x, \frac{\partial V}{\partial x}, u) = \frac{d}{du}\left(2(Ax + Bu)^T Px + x^T Qx + u^T Ru\right) = 0$

$$2Ru + B^T Px = 0$$

Optimal Control is SVFB $u = -R^{-1}B^T Px = -Kx$

Algebraic Riccati equation

$$0 = PA + A^T P + Q - PBR^{-1}B^T P$$

Full system dynamics must be known
Off-line solution

Optimal Control: Linear Quadratic Regulator

System model $\dot{x} = Ax + Bu$

Performance Function $V(x(t)) = \int_t^{\infty} (x^T Qx + u^T Ru)\, d\tau \qquad = x^T(t)Px(t)$

Algebraic Riccati equation

$$0 = PA + A^T P + Q - PBR^{-1}B^T P$$

Optimal Control is

$$u = -R^{-1}B^T Px = -Kx$$

MATLAB Control Systems Toolbox

[K,P]=lqr(A,B,Q,R)

Full system dynamics must be known
Off-line solution
Cannot change performance objectives

Optimal Control: Linear Quadratic Regulator

System $\dot{x} = Ax + Bu$

Cost $V(x(t)) = \int_t^{\infty} (x^T Qx + u^T Ru)\, d\tau \qquad = x^T(t)Px(t)$

Leibniz's formula- Differential equivalent is the Bellman equation

$$0 = H(x, \frac{\partial V}{\partial x}, u) = \dot{V} + x^T Qx + u^T Ru = 2\left(\frac{\partial V}{\partial x}\right)^T \dot{x} + x^T Qx + u^T Ru = 2x^T P(Ax + Bu) + x^T Qx + u^T Ru$$

Given any stabilizing FB policy $u = -Kx$

The cost value is found by solving Lyapunov equation

$$0 = (A - BK)^T P + P(A - BK) + Q + K^T RK$$

$$\frac{d}{du} H(x, \frac{\partial V}{\partial x}, u) = \frac{d}{du}\left(2(Ax + Bu)^T Px + x^T Qx + u^T Ru\right) = 0$$

Optimal Control is

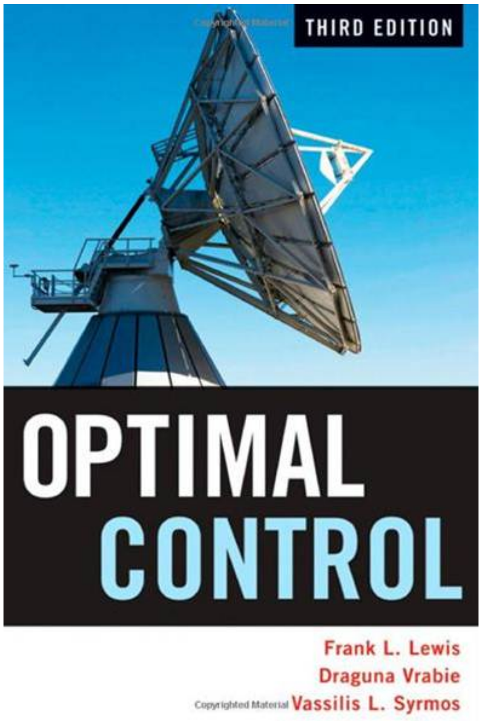
$$u = -R^{-1}B^T Px = -Kx$$

Algebraic Riccati equation

$$0 = PA + A^T P + Q - PBR^{-1}B^T P$$

Full system dynamics must be known
Off-line solution

Many Successful Design Applications of Optimal Control



Optimal Control: Linear Quadratic Regulator

System $\dot{x} = Ax + Bu$

Cost function $V(x(t)) = \int_t^\infty (x^T Qx + u^T Ru)\, d\tau = \int_t^\infty r(x,u)\, d\tau \qquad = x^T(t)Px(t)$

Optimal Control is

$$u = -R^{-1}B^T Px = -Kx$$

Algebraic Riccati equation

$$0 = PA + A^T P + Q - PBR^{-1}B^T P$$

MATLAB Control Systems Toolbox

Full system dynamics must be known
Off-line solution

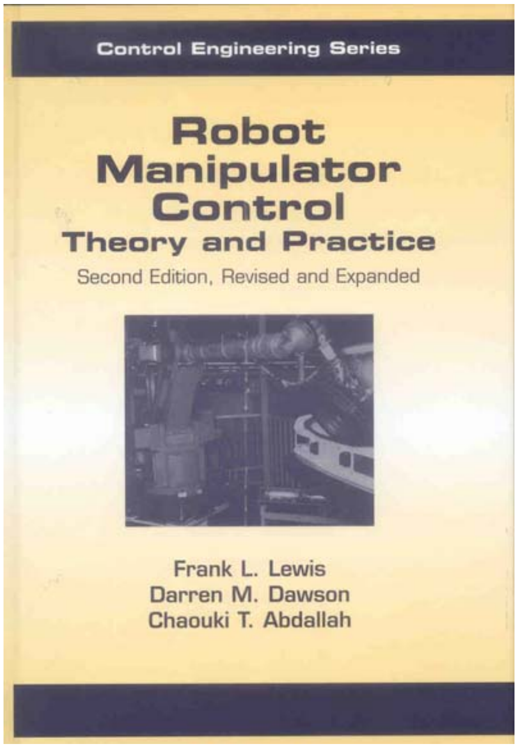
Chop off Tail of cost function $V(x(t)) = \int_t^{t+T} r(x,u)\, d\tau \quad + \int_{t+T}^{\infty} r(x,u)\, d\tau$

Bellman Equation $V(x(t)) = \int_t^{t+T} r(x,u)\, d\tau \quad + \quad V(x(t+T))$

Update Control using Hamiltonian Physics $u = -R^{-1}B^T Px = -Kx$

Reinforcement Learning- Policy Iteration

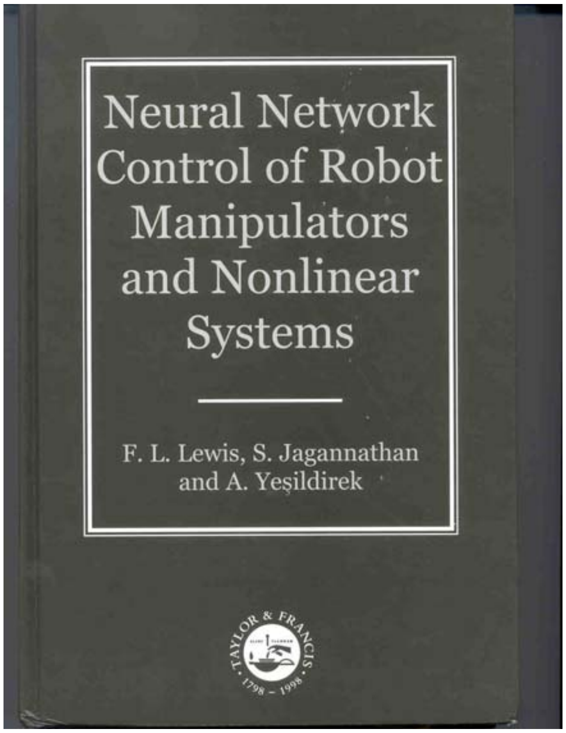
Online Learning Feedback Control
Synthesis of Robot Control and Biologically Inspired Learning



Control Engineering Series

Robot Manipulator Control
Theory and Practice
Second Edition, Revised and Expanded

Frank L. Lewis
Darren M. Dawson
Chaouki T. Abdallah

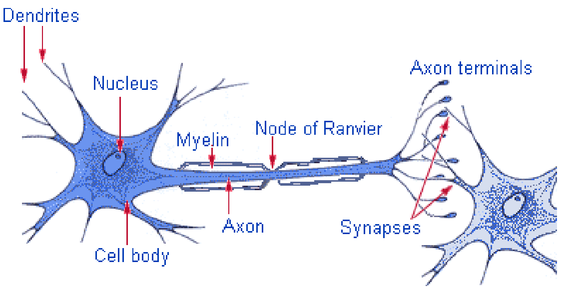


Neural Network Control of Robot Manipulators and Nonlinear Systems

F. L. Lewis, S. Jagannathan
and A. Yeşildirek

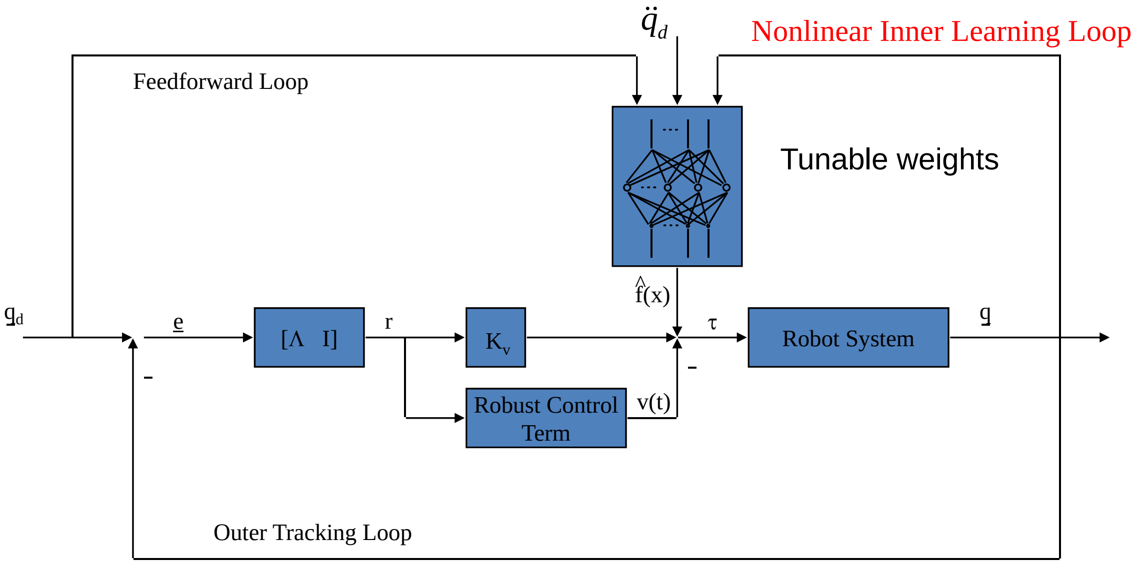
Neural Network Properties

- Learning
- Recall
- Function approximation
- Generalization
- Classification
- Association
- Pattern recognition
- Clustering
- Robustness to single node failure
- Repair and reconfiguration



Nervous system cell.
http://www.sirinet.net/~jgjohnso/index.html

Dynamic Neural Networks and Control Learning



NEW NN Tuning Laws for Control Learning

Theorem 1 (NN Weight Tuning for Stability)

Let the desired trajectory $q_d(t)$ and its derivatives be bounded. Let the initial tracking error be within a certain allowable set U. Let Z_M be a known upper bound on the Frobenius norm of the unknown ideal weights Z.

Take the control input as

$$\tau = \hat{W}^T \sigma(\hat{V}^T x) + K_v r - v \qquad \text{with} \qquad v(t) = -K_Z(\|Z\|_F + Z_M)r .$$

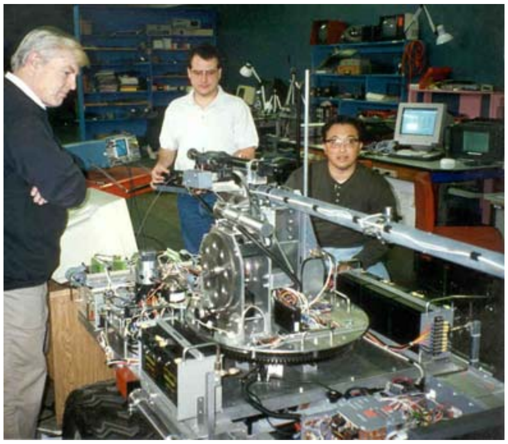
Let weight tuning be provided by

$$\boxed{\dot{\hat{W}} = F\hat{\sigma}r^T - F\hat{\sigma}'\hat{V}^T x r^T - \kappa F\|r\|\hat{W} ,} \quad \boxed{\dot{\hat{V}} = Gx(\hat{\sigma}'^T \hat{W} r)^T - \kappa G\|r\|\hat{V}}$$

with any constant matrices $F = F^T > 0, G = G^T > 0$, and scalar tuning parameter $\kappa > 0$. Initialize the weight estimates as $\hat{W} = 0, \hat{V} = random$.

Then the filtered tracking error $r(t)$ and NN weight estimates $\hat{W}, \hat{V}$ are uniformly ultimately bounded. Moreover, arbitrarily small tracking error may be achieved by selecting large control gains K_v.

Force Control



Flexible pointing systems
Simis labs, Inc. and US Army

SBIR Contracts

Vehicle active suspension
Leo Davis Technol.

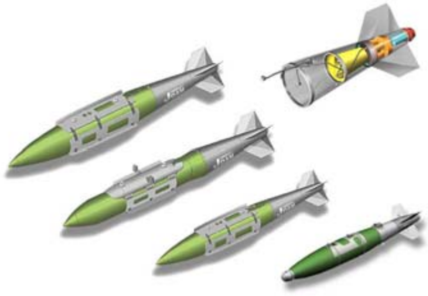
Learning NN Controller

Applications at Boeing Defense Space & Security

Kevin Wise and Eugene Lavretsky

Highly reliable adaptive uncertainty approximation compensators for flight control applications:
unmanned aircraft – Phantom Ray

Tailkit adaptive control systems for
Joint Direct Attack Munition (JDAM) munitions:
Mk-82, Mk-84, and Laser-guided variants
Fielded for defense in Iraq and Afghanistan.



BUT-
Nature is More than Stable
Nature is OPTIMAL and conserves energy

For Adaptive Man/Machine systems-

We want to Learn optimal control solutions
Online in real-time
Using adaptive control techniques
Without knowing the full dynamics

Optimality in Biological Systems

Every living organism improves its control actions based on rewards received from the environment

The resources available to living organisms are usually meager.
Nature uses optimal control.

Reinforcement Learning

1. Apply a control. Evaluate the benefit of that control.
2. Improve the control policy.

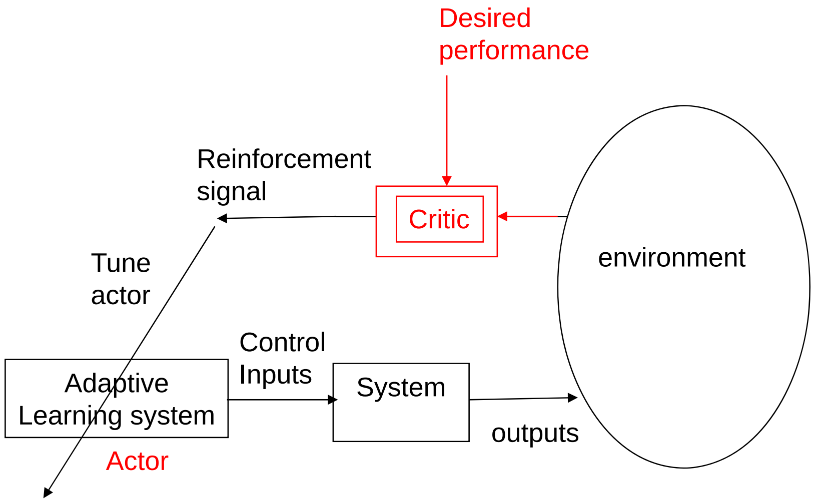
RL finds optimal policies by evaluating the effects of suboptimal policies

Different methods of learning

Reinforcement learning
Ivan Pavlov 1890s

We want OPTIMAL performance
- ADP- Approximate Dynamic Programming

Actor-Critic Learning



Desired performance

Reinforcement signal

Critic

environment

Tune actor

Control Inputs

Adaptive Learning system

Actor

System

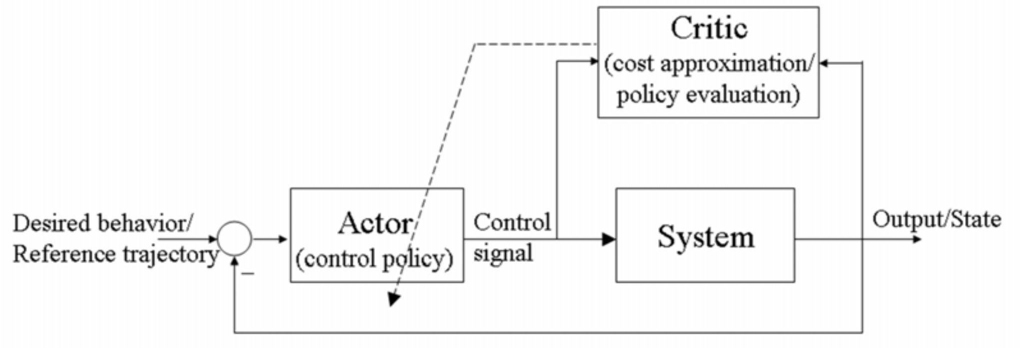
outputs

Sutton & Barto book

Adaptive Critic structure

Reinforcement learning

Slow learning Improvement loop



Fast inner control loops

Optimal Control: Linear Quadratic Regulator

System $\dot{x} = Ax + Bu$

Cost function $V(x(t)) = \int_t^{\infty} (x^T Qx + u^T Ru)\, d\tau = \int_t^{\infty} r(x,u)\, d\tau \quad = x^T(t)Px(t)$

Optimal Control is

$$u = -R^{-1}B^T Px = -Kx$$

Full system dynamics must be known

Off-line solution

Algebraic Riccati equation

$$0 = PA + A^T P + Q - PBR^{-1}B^T P$$

MATLAB Control Systems Toolbox

Chop off Tail of cost function $V(x(t)) = \int_t^{t+T} r(x,u)\, d\tau \quad + \int_{t+T}^{\infty} r(x,u)\, d\tau$

Bellman Equation $V(x(t)) = \int_t^{t+T} r(x,u)\, d\tau \quad + \quad V(x(t+T))$

Update Control using Hamiltonian Physics $u = -R^{-1}B^T Px = -Kx$

Reinforcement Learning- Policy Iteration

CT Policy Iteration – How to implement online?
Linear Systems Quadratic Cost- LQR

Value Function Approximation

Value function is quadratic $V(x(t)) = x^T(t)Px(t)$

Policy evaluation- solve IRL Bellman Equation

$$x^T(t)P_k x(t) = \int_t^{t+T} x^T(\tau)(Q + K_k{}^T RK_k)x(\tau)\, d\tau \quad + \quad x^T(t+T)P_k x(t+T)$$

$$x^T(t)P_k x(t) - x^T(t+T)P_k x(t+T) = \int_t^{t+T} x^T(\tau)(Q + K_k{}^T RK_k)x(\tau)\, d\tau$$

$$\bar{p}_k{}^T \phi(t) \equiv \quad \bar{p}_k{}^T \left[\bar{x}(t) - \bar{x}(t+T)\right] = \int_t^{t+T} x(\tau)^T (Q + K_k{}^T RK_k)x(\tau)d\tau$$

regression vector

Reinforcement on time interval *[t, t+T]*

Same form as standard System ID problems

Solve using RLS or batch LS

Union of Reinforcement Learning and Adaptive Control

Integral Reinforcement Learning (IRL)

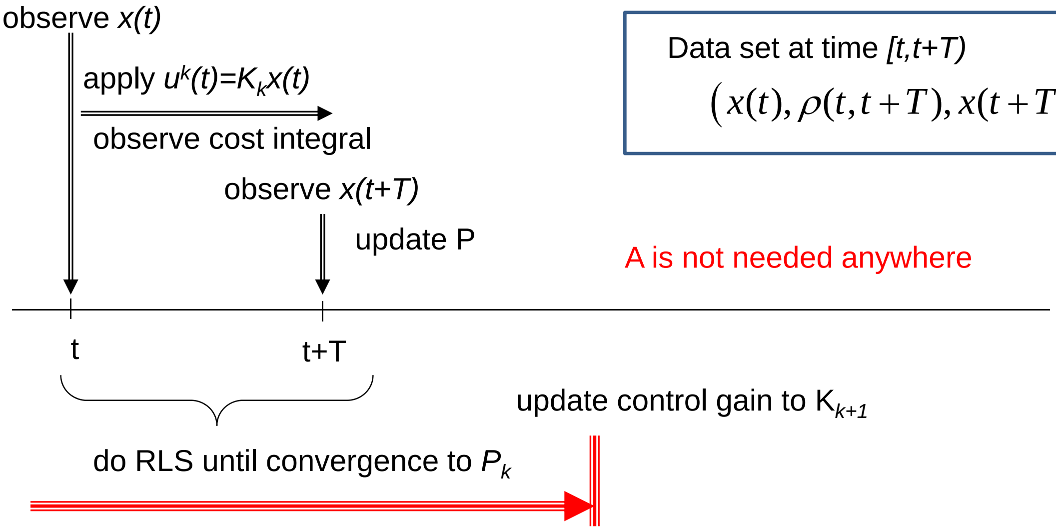
This is a data-based approach that uses measurements of *x(t), u(t)*
Instead of the plant dynamical model.

1. Select initial control policy
2. Find associated cost

Bellman Equation
Solves Lyapunov eq. without knowing dynamics

$$\overline{p}_k^{\,T}\left[\overline{x}(t)-\overline{x}(t+T)\right]=\int_t^{t+T} x(\tau)^T(Q+K_k^{\,T}RK_k)x(\tau)d\tau=\rho(t,t+T)$$

3. Improve control $K_{k+1}=R^{-1}B^TP_k$



observe *x(t)*

apply $u^k(t)=K_kx(t)$

observe cost integral

observe *x(t+T)*

update P

t

t+T

do RLS until convergence to P_k

update control gain to K_{k+1}

Data set at time *[t,t+T)*

$$\left(x(t),\rho(t,t+T),x(t+T)\right)$$

A is not needed anywhere

Oscillation is a fundamental property of neural tissue

Brain has multiple adaptive clocks with different timescales

gamma rhythms 30-100 Hz, hippocampus and neocortex
high cognitive activity.

- consolidation of memory
- spatial mapping of the environment – place cells

The high frequency processing is due to the large amounts of sensorial data to be processed

theta rhythm, Hippocampus, Thalamus, 4-10 Hz
sensory processing, memory and voluntary control of movement.

Spinal cord

Motor control 200 Hz

D. Vrabie, F. Lewis, and Dan Levine- RL for Continuous-Time Systems

Limbic system

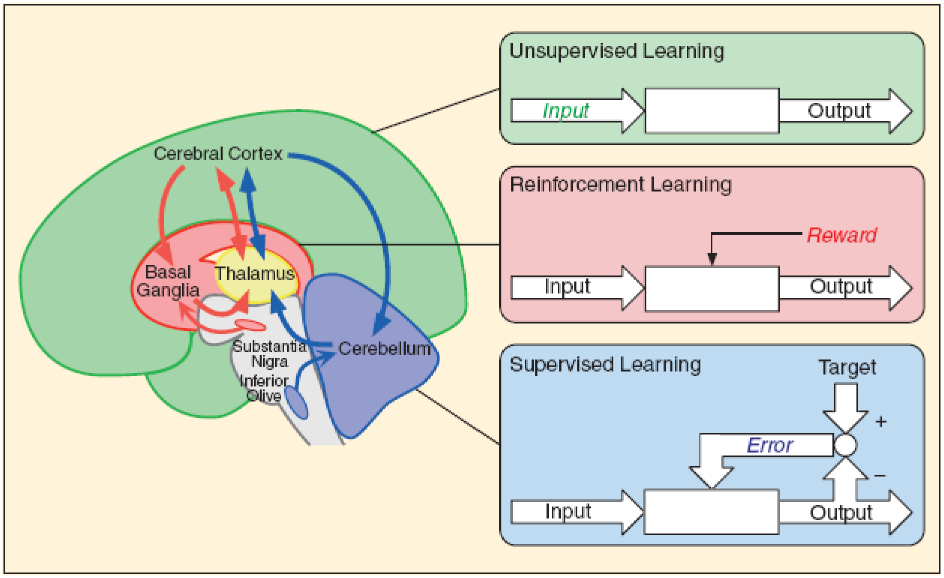
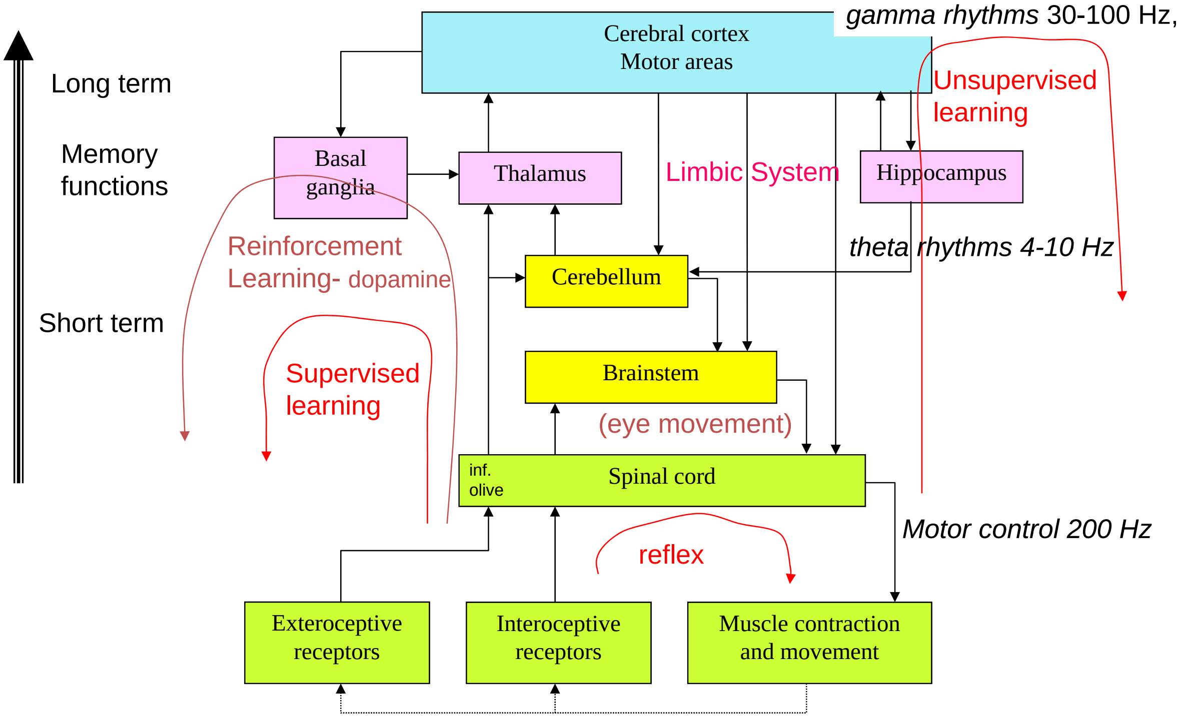


Figure 1. *Learning-oriented specialization of the cerebellum, the basal ganglia, and the cerebral cortex [1], [2]. The cerebellum is specialized for supervised learning based on the error signal encoded in the climbing fibers from the inferior olive. The basal ganglia are specialized for reinforcement learning based on the reward signal encoded in the dopaminergic fibers from the substantia nigra. The cerebral cortex is specialized for unsupervised learning based on the statistical properties of the input signal.*

Doya, Kimura, Kawato 2001

Summary of Motor Control in the Human Nervous System

picture by E. Stingu
D. Vrabie



gamma rhythms 30-100 Hz,

Long term

Memory functions

Short term

Cerebral cortex
Motor areas

Basal ganglia

Thalamus

Limbic System

Hippocampus

Unsupervised learning

Reinforcement Learning- dopamine

Cerebellum

theta rhythms 4-10 Hz

Supervised learning

Brainstem

(eye movement)

inf. olive

Spinal cord

Motor control 200 Hz

reflex

Exteroceptive receptors

Interoceptive receptors

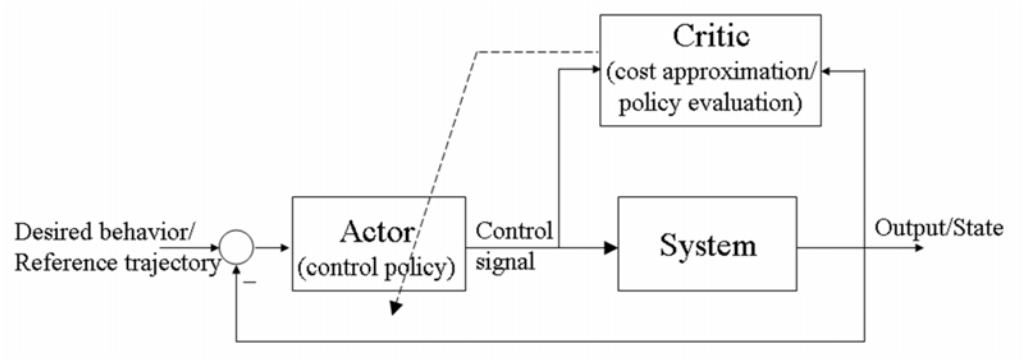
Muscle contraction and movement

Hierarchy of multiple parallel loops

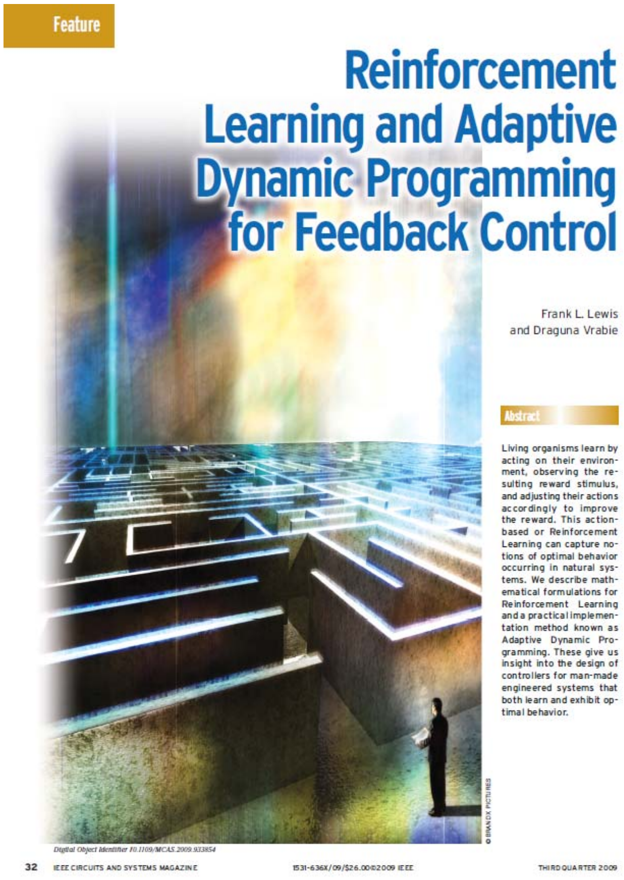
Adaptive Critic structure

Reinforcement learning

Theta waves 4-8 Hz



Motor control 200 Hz



Feature

Reinforcement Learning and Adaptive Dynamic Programming for Feedback Control

Frank L. Lewis
and Draguna Vrabie

Abstract

Living organisms learn by acting on their environment, observing the resulting reward stimulus, and adjusting their actions accordingly to improve the reward. This action-based or Reinforcement Learning can capture notions of optimal behavior occurring in natural systems. We describe mathematical formulations for Reinforcement Learning and a practical implementation method known as Adaptive Dynamic Programming. These give us insight into the design of controllers for man-made engineered systems that both learn and exhibit optimal behavior.

© BRAND X PICTURES

Digital Object Identifier 10.1109/MCAS.2009.933854

32 IEEE CIRCUITS AND SYSTEMS MAGAZINE 1531-636X/09/$26.00©2009 IEEE THIRD QUARTER 2009

F.L. Lewis and D. Vrabie, “Reinforcement learning and adaptive dynamic programming for feedback control,” IEEE Circuits & Systems Magazine, Invited Feature Article, pp. 32-50, Third Quarter 2009.

IEEE Control Systems Magazine “Reinforcement learning and feedback Control,” Dec. 2012

Books

F.L. Lewis, D. Vrabie, and V. Syrmos,
Optimal Control, third edition, John Wiley and Sons, New York, 2012.

New Chapters on:
- Reinforcement Learning
- Differential Games

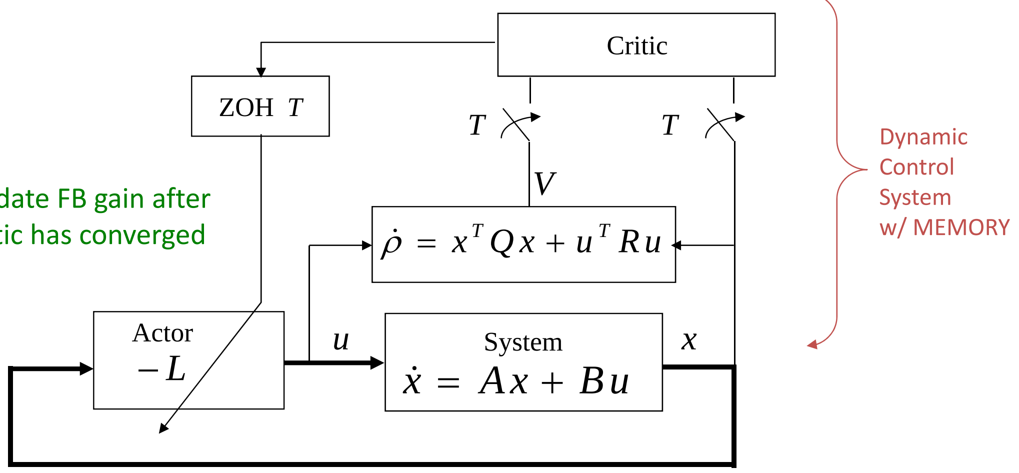
D. Vrabie, K. Vamvoudakis, and F.L. Lewis, *Optimal Adaptive Control and Differential Games by Reinforcement Learning Principles*, IET Press, 2012, to appear.

Draguna Vrabie

Direct Optimal Adaptive Controller

Solves Riccati Equation Online without knowing A matrix

Run RLS or use batch L.S.
To identify value of current control



Dynamic
Control
System
w/ MEMORY

Update FB gain after
Critic has converged

A hybrid continuous/discrete dynamic controller
whose internal state is the observed cost over the interval

Reinforcement interval T can be selected on line on the fly – can change

Optimal Control Design Allows a Lot of Design Freedom

$$V(x(t)) = \int_t^{\infty} r(x,u)\, d\tau$$

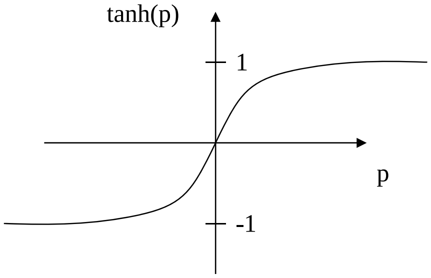
Tailor controls design by choosing utility $r(x,u)$

Minimum Energy
Minimum Fuel
Minimum Time
Control Actuator Constraints

Murad Abu-Khalaf

Optimal Control for Constrained Input Systems

Control constrained by saturation function $\sigma(.)$



Encode constraint into Value function

$$J(u,d) = \int_0^{\infty} \left(Q(x) + 2\int_0^{u} \sigma^{-T}(v)dv \right) dt$$

$$\|u\|_q^{\;2} = 2\int_0^{u} \sigma^{-T}(v)dv$$

(Used by Lyshevsky for H_2 control)

This is a quasi-norm

Weaker than a norm –
homogeneity property is replaced by the weaker symmetry property $\|x\|_q = \|-x\|_q$

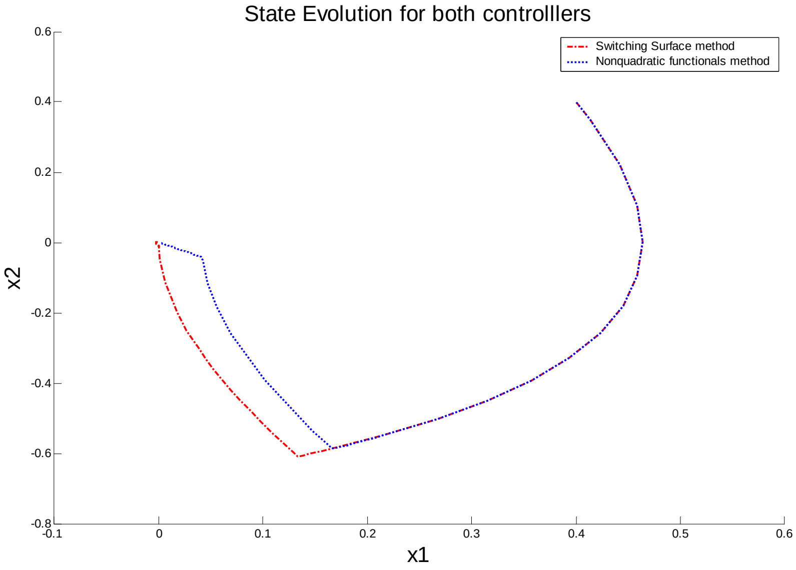
Then $u = -\sigma\left(R^{-1} g(x)^T \frac{\partial V}{\partial x} \right)$ Is BOUNDED

Murad Abu-Khalaf

Near Minimum-Time Control

Encode into Value Function

$$V = \int_0^{\infty} \left[\tanh(x^T Q\, x) + 2\int_0^{u} \left(\sigma^{-1}(\mu)\right)^T R d\mu \right] dt$$



SBIR Contracts

SBA Tibbets Award 1996 (led by TMAC)

Force Control
Intelligent Automation, Inc.

Flexible pointing systems
Simis labs, Inc. and US Army

Vehicle active suspension
Leo Davis Technol.

Learning NN Controller

Patents

- A. Yesildirek and F.L. Lewis, "Method for feedback linearization of neural networks and neural network incorporating same," U.S. Patent 5,943,660, awarded 24 August 1999.
- S. Jagannathan and F.L. Lewis, "Discrete-time tuning of neural network controllers for nonlinear dynamical systems," U.S. Patent 6,064,997, awarded 16 May 2000.
- R. Selmic, F.L. Lewis, A.J. Calise, and M.B. McFarland, "Backlash Compensation Using Neural Network," U.S. Patent 6,611,823, awarded 26 Aug. 2003.
- J. Campos and F.L. Lewis, "Method for Backlash Compensation Using Discrete-Time Neural Networks," U.S. Patent 7,080,055, awarded July 2006.
- K. Vamvoudakis, D. Vrabie, and F.L. Lewis, “Control methodology for online adaptation to optimal feedback controller using integral reinforcement learning,” provisional patent, filed March 2012.

Applications of our Algorithms at Boeing Defense Space & Security

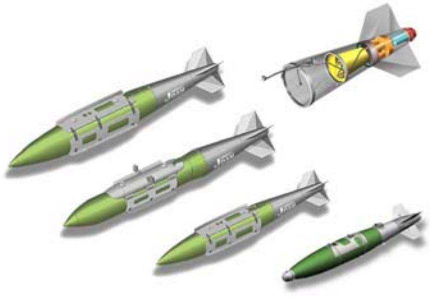
Kevin Wise and Eugene Lavretsky

Highly reliable adaptive uncertainty approximation compensators for flight control applications:

Unmanned aircraft – Phantom Ray

Tailkit adaptive control systems for
Joint Direct Attack Munition (JDAM) munitions:
Mk-82, Mk-84, and Laser-guided variants

Fielded for defense in Iraq and Afghanistan.
Saved Lives.



Applications of Our Algorithms to Auto Engine Control

Student S. Jagannathan
11 US patents

optimal engine controllers based on RL for Caterpillar, Ford (Zetec engine), and GM

8-10% improvement in fuel efficiency and a drastic reduction in NOx (90%), HC (30%) and CO (15%) by operating with adaptive exhaust gas recirculation.

Savings to Caterpillar were over $65,000 per component.

A Q-Learning Based Adaptive Optimal Controller Implementation for a Humanoid Robot Arm

Said Ghani Khan[1], Guido Herrmann[1], Tony Pipe[1], Chris Melhuish[1]

Bristol Robotics Laboratory
University of Bristol and University of the West of England, Bristol, UK

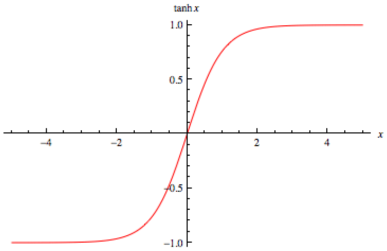
Conference on Decision and Control (CDC) 2011, Orlando, 11 December 2011

Introducing Constraints

The cost function is modified to include constraints

$$C(q) = \begin{cases} tan^2(\frac{q}{q_L} \times \frac{\pi}{2}), & if||q|| < q_L \times \lambda \\ tan^2(\lambda \times \frac{q}{q_L}), & if||q|| \geq q_L \times \lambda \end{cases}$$

$0<\lambda<1$



$q_L>0$ is the joint limit.

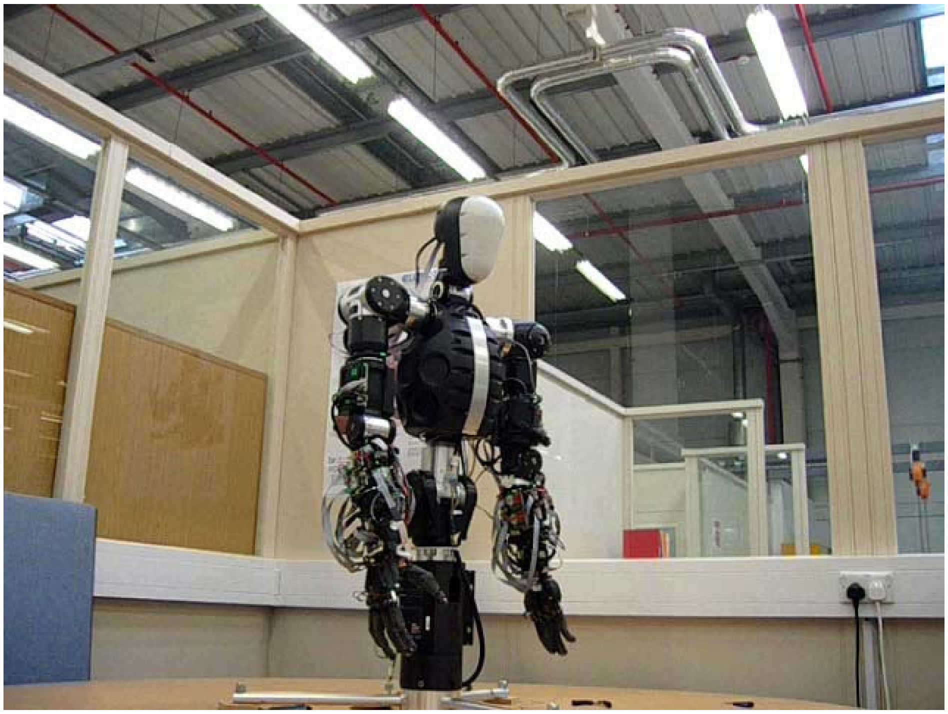
The new cost function....

$$r(x_k, u_k, d_k) = e_k^T Q_c e_k + (u_{k+1} - u_k)^T S(u_{k+1} - u_k) + (u_k)^T R(u_k) + \Lambda C(q)$$

where, Λ is a positive constant.

BRL BERT II ARM

Bristol Robotics Lab



Natural Human-Like Motion

Automation & Robotics Research Institute
University of Texas at Arlington

An Approximate Dynamic Programming Based Controller for an Underactuated 6DoF Quadrotor

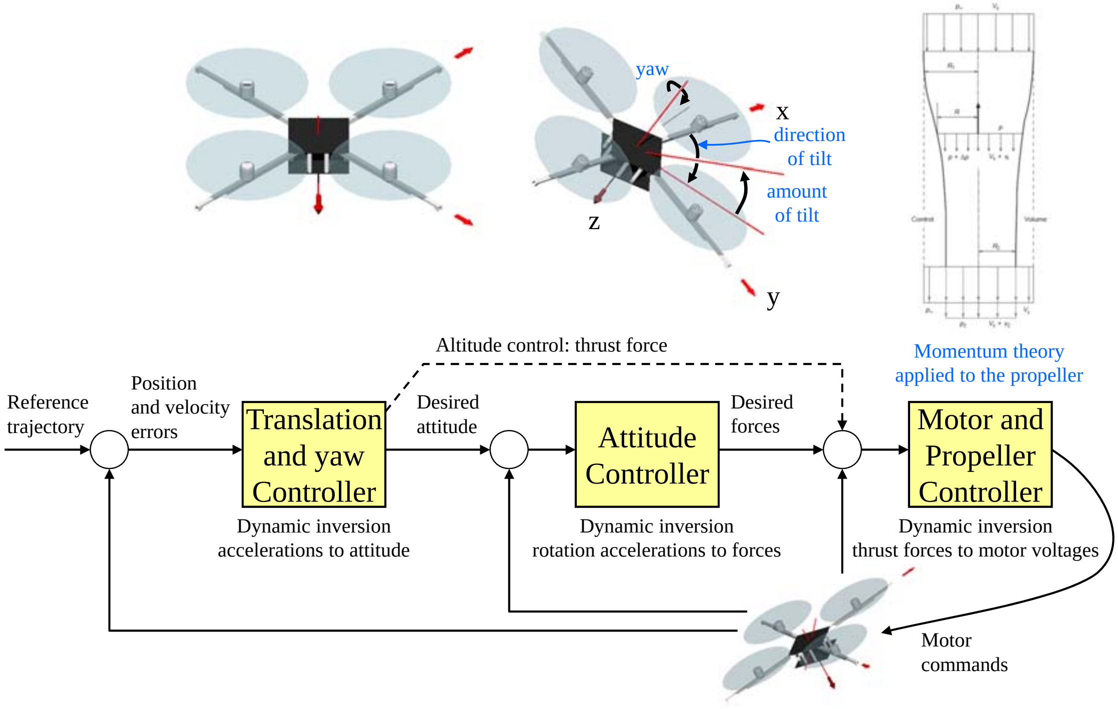
Emanuel Stingu

Frank Lewis

Supported by
ARO grant W91NF-05-1-0314
NSF grant ECCS-0801330

3 control loops

The quadrotor has 17 states and only 4 control inputs, thus it is very under-actuated.
Three control loops with dynamic inversion are used to generate the 4 control signals.



Momentum theory applied to the propeller

Altitude control: thrust force

Reference trajectory

Position and velocity errors

Translation and yaw Controller

Dynamic inversion accelerations to attitude

Desired attitude

Attitude Controller

Dynamic inversion rotation accelerations to forces

Desired forces

Motor and Propeller Controller

Dynamic inversion thrust forces to motor voltages

Motor commands

The compass is used such that the quadrotor goes in the correct direction independent of its relative orientation to the pilot.

Using standard controller

Using Reinforcement learning controller

The pilot will give heading commands most of the time to test the behavior of the quadrotor.

Potential New Applications in Prosthetics

With Sesh Commuri, Oklahoma University
And Muthu Wijesundara

Adaptive Reinforcement Learning of
Human/Prosthetic Ankle Interactions

Current methods use stored gait information

magellan | Microprocessor Foot + Ankle

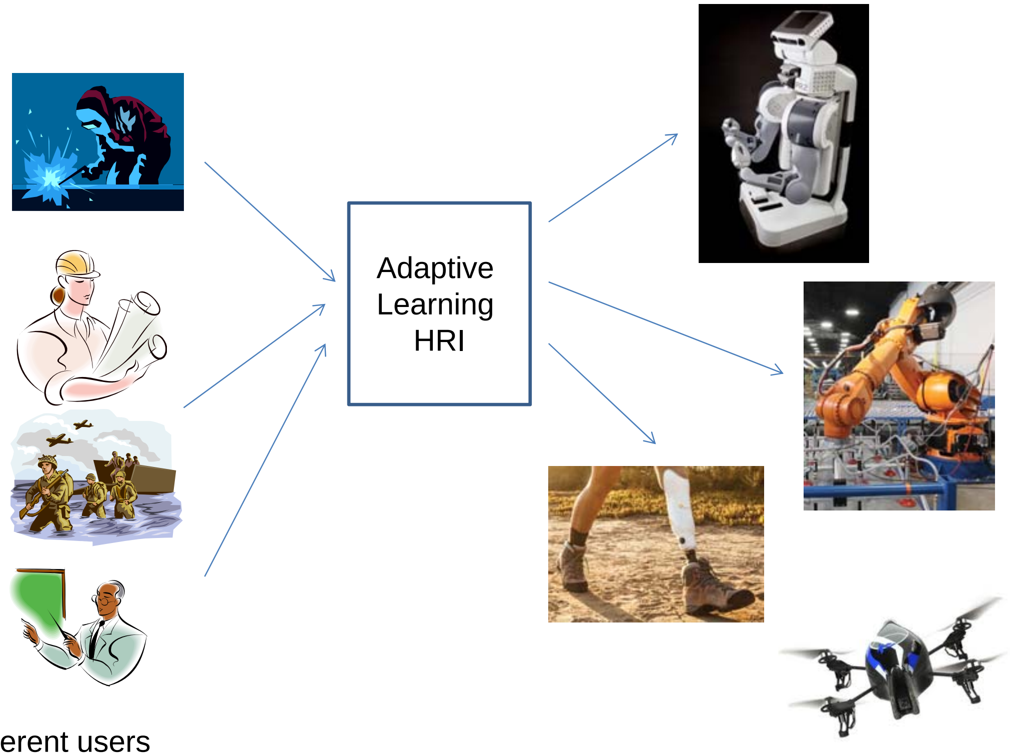
Microprocessor—The Brain

magellan's microprocessor intelligence "shifts gears" rapidly on all terrains, adjusting to changes more quickly than other devices. The advanced software is able to intuit almost any movement. The device's advanced hydraulic system and controls provide comfort, safety, first-step resistance modification and adaptive control for an easier walk.

The Nervous System

magellan's automatic intuitive adjustments require little conscious effort on the user's part. Orthocare's europa™ smart-sensing technology incorporates electronic force sensors into the structure of the prosthesis that measure, analyze and improve dynamic socket reaction forces.

What about Adaptive Human-Robot Interfaces (HRI)?



Adaptive Learning HRI

Different users

Different robotic devices

Interface learns so as to improve Human/Robot team performance

Tune parameters such as stiffness, speed of response, dynamic motion

There are 3 components
  Human
  Robot
  The Interface

They should all be intelligent

We want verifiable performance and straightforward tuning algorithms

Several Approaches to Adaptive HRI

1. Learning of Coordinate Transforms
2. Dynamic Motion Primitives to Modify Task Parameters
3. Adaptive Impedance Control

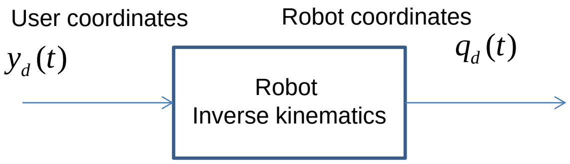
Current and Expected Funding

F.L. Lewis, “Adaptive Dynamic Programming for Real-Time Cooperative Multi-Player Games and Graphical Games,” NSF Grant, $272,000 for 3 years, July 2011.

D. Popa, Z. Celik-Butler, D. Butler, and F.L. Lewis, “NRI: Multi-Modal Skin and Garments for Healthcare and Home Robots,” NSF Grant, $1.3M for 4 years, Sept. 2012.

F.L. Lewis, “Games and Learning for Cooperative Nonlinear Systems and Internal Structure of Coalitions on Graphs, TARDEC/NAC grant, $81,000 for 1 year, Jan. 2013 expected.
IF NO FISCAL CLIFF!

Learning of Coordinate Transforms



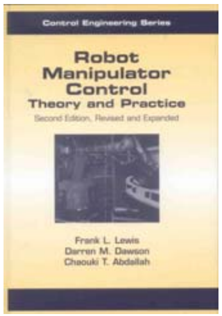
User coordinates

Robot coordinates

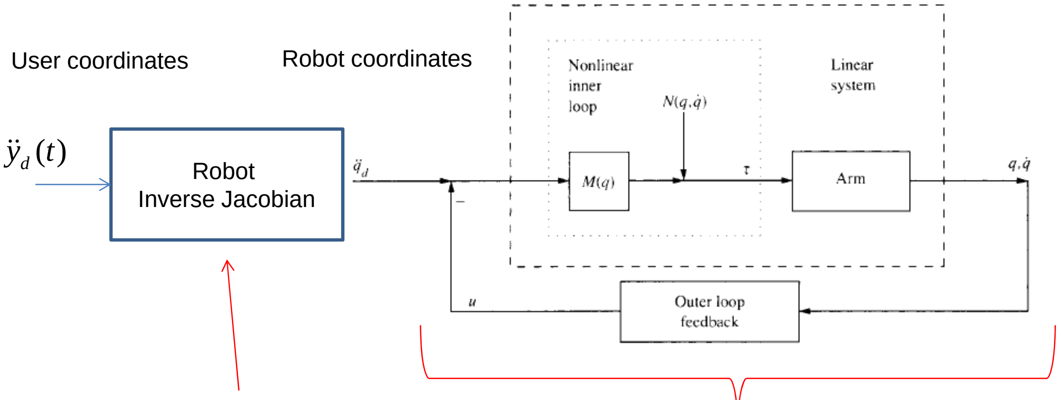
$y_d(t)$

Robot
Inverse kinematics

$q_d(t)$



$$\ddot{q}_d = J^{-1}\ddot{y}_d$$



User coordinates

Robot coordinates

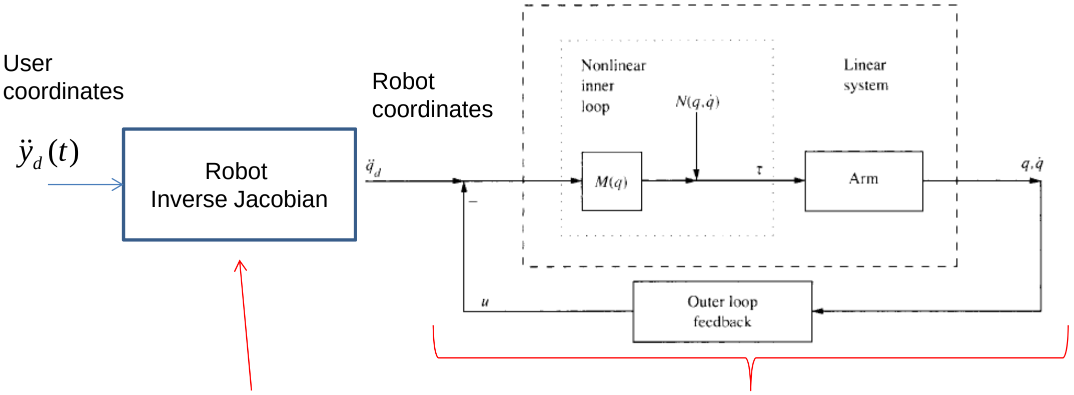
$\ddot{y}_d(t)$

Robot
Inverse Jacobian

Hard to find Accurately
Need dynamics model of robot
Does not generalize to different robots

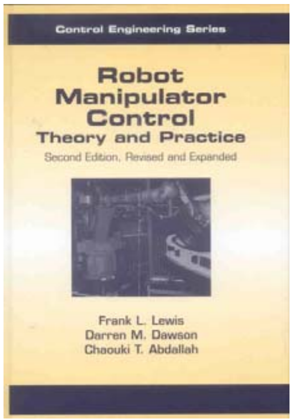
Standard robot controller

1. Learning of Coordinate Transforms



Standard robot controller

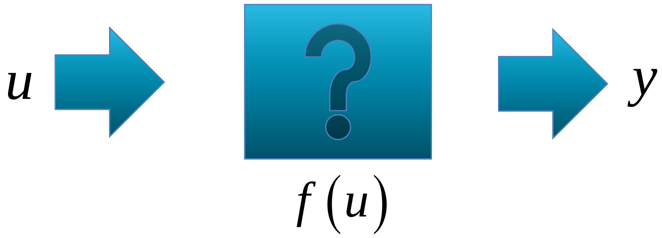
Hard to find accurately
Need dynamics model of robot
Does not generalize to different robots



Learning the Interface Mapping

Reinforcement Learning

Dan Popa



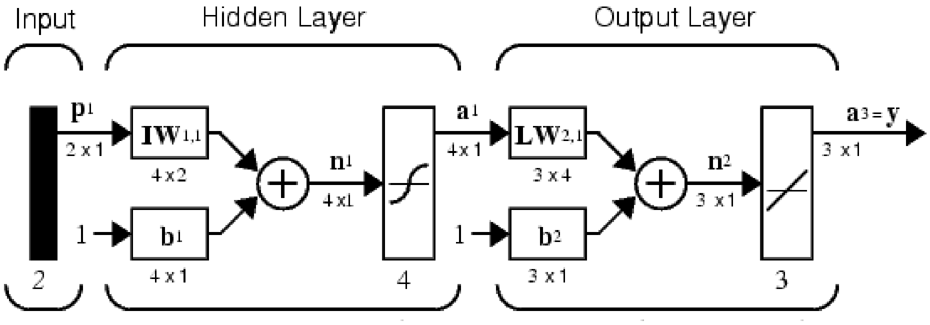
What can we do to get f(u)?

The simplest way is to obtain a set of inputs and a set of outputs and calculate the relationship (Curve Fitting)

Static nonlinear map $y = f(u) = \sum_{i=0}^{M} w_i \Phi_i\left(\sum_{j=0}^{P} w_{ij} u_i\right)$

NN Training

Dan Popa



Log sigmoid function is used as a neuron activation function

$$\sigma(.) = \frac{1}{1+e^{-x}}$$

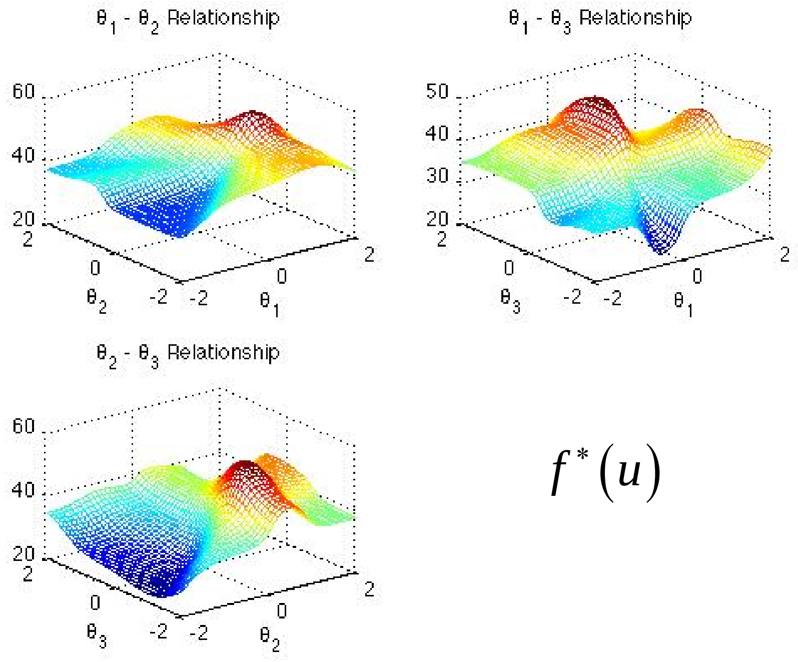
Train using MATLAB NN Toolbox very easily if input/output pairs are known.

Static Mapping Approach

Dan Popa

Found by training neural network with known input/output pairs

f(u) is a coordinate transformation



$$f^*(u)$$

Reinforcement Learning

Dan Popa

u

y

Curve Fitting Algorithm

$f^*(u)$

What if we cannot specify the desired output trajectory directly ?

Reinforcement Learning

With RL we do not have to specify a desired trajectory.

Instead, a Reward Function is used

How to find the Reward Function?

---

2. Learn and Modify Task Parameters

Model of Task

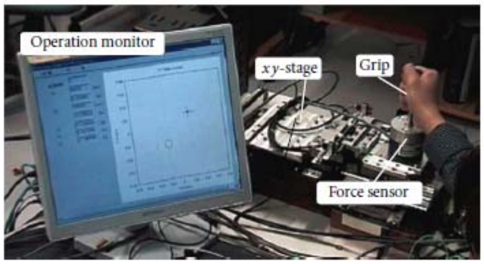
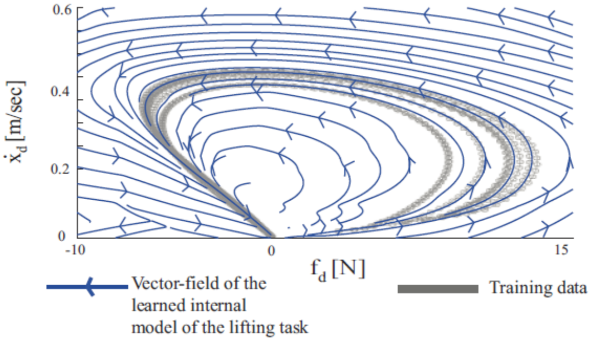


FIGURE 3: Haptic interface devices.

A task is a profile in velocity/force space
Different Curve for Different Tasks



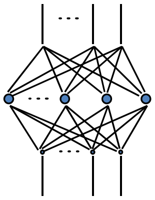
Dynamic Motion Primitives for Characterizing the Task

Tunable NN weights

$$\tau^2\ddot{x}+\tau D\dot{x}+Ks(g-x_0)-K(g-x)=KW^T\phi(V^Ts)$$

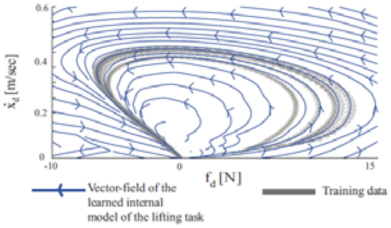
$$\tau\dot{s}=-\alpha s$$

Phase variable *s* makes DMP independent of time



Learning the task:
Kinesthetic demonstration to learn D, K, initial weights
Reward-based learning to tune the weights

Different NN weights give different task trajectories

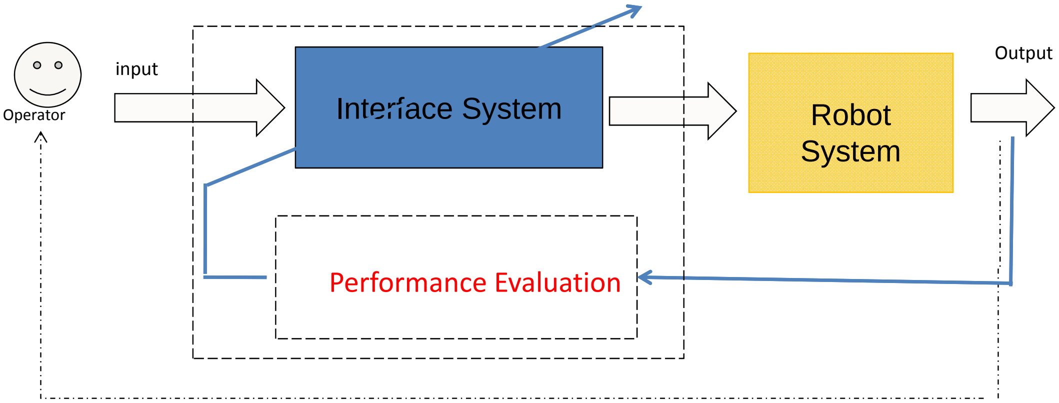


Adaptive Impedance Control

Reinforcement Learning for Human-Robot Interfaces

Dan Popa

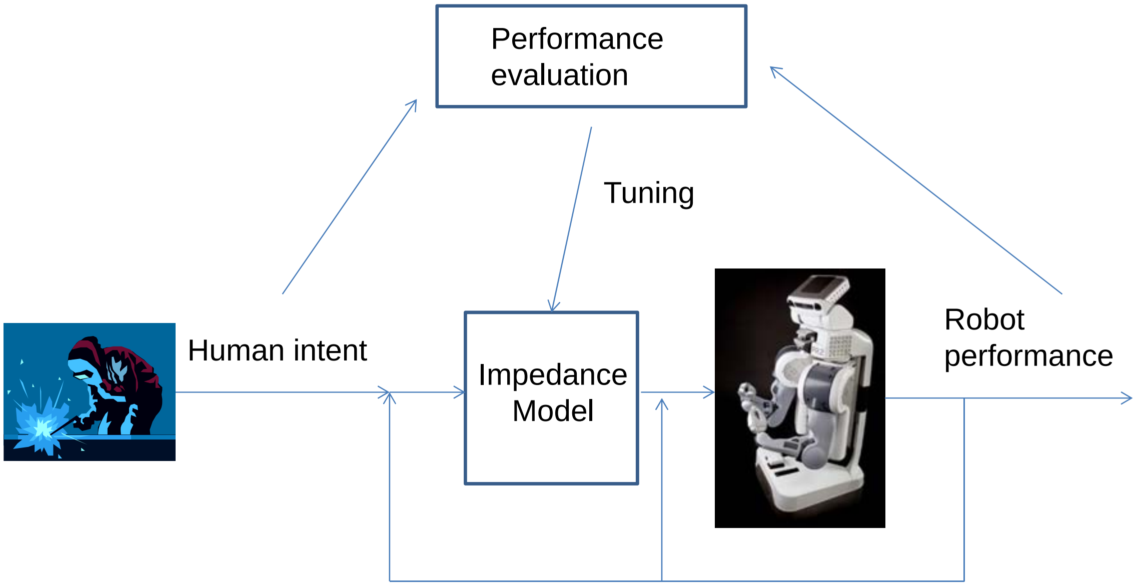
- Build an adaptive interface system that allows a single operator to manipulate a robots with multiple degrees of freedom and/or multiple robots
- The interface system should be intuitive, easy to use and can be learned quickly , and should be able to apply with different robot configurations



What to learn?
How to evaluate Performance?

3. Adaptive Impedance Control

Tune interface so robot learns to control the human model to improve task performance



Performance evaluation

Tuning

Human intent

Impedance Model

Robot performance

Inner loop FB controllers

Model of Human Behavior

Suzuki & Furuta 2012

For a given task, a skilled human operator adapts to learn-

1. An inverse model of the robot system to cancel nonlinearities
2. A feedforward predictive control based on the task

Frequency Crossover Theory -

Human operator adapts to make the overall closed-loop man-machine system

look like a high bandwidth frequency response

And remain invariant to a wide range of task variations and changes in operating condition

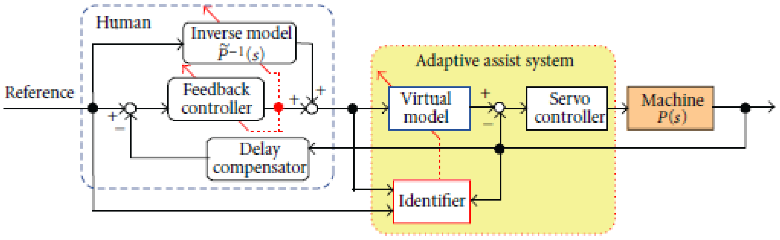
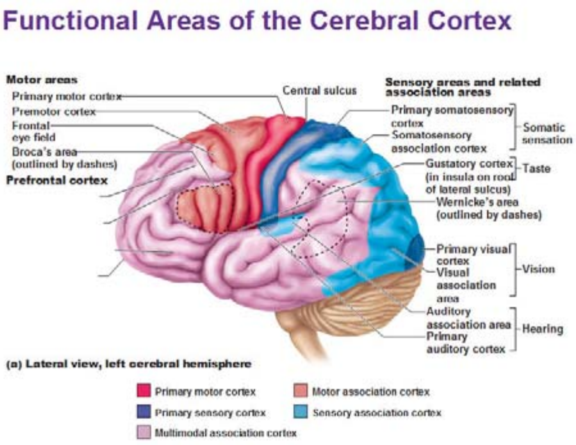
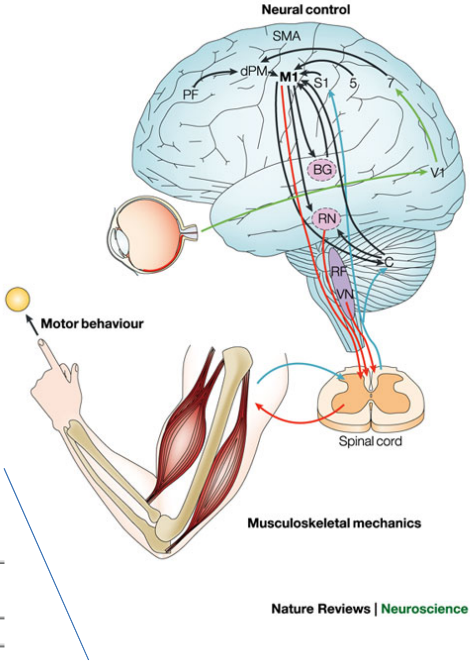
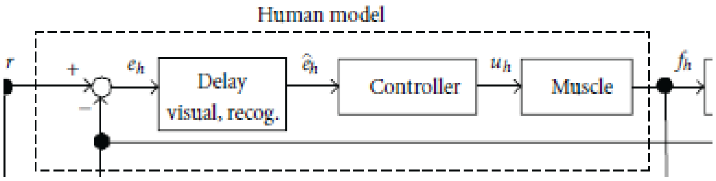


Figure 2: Structure of human assistive system with the adaptive impedance control.







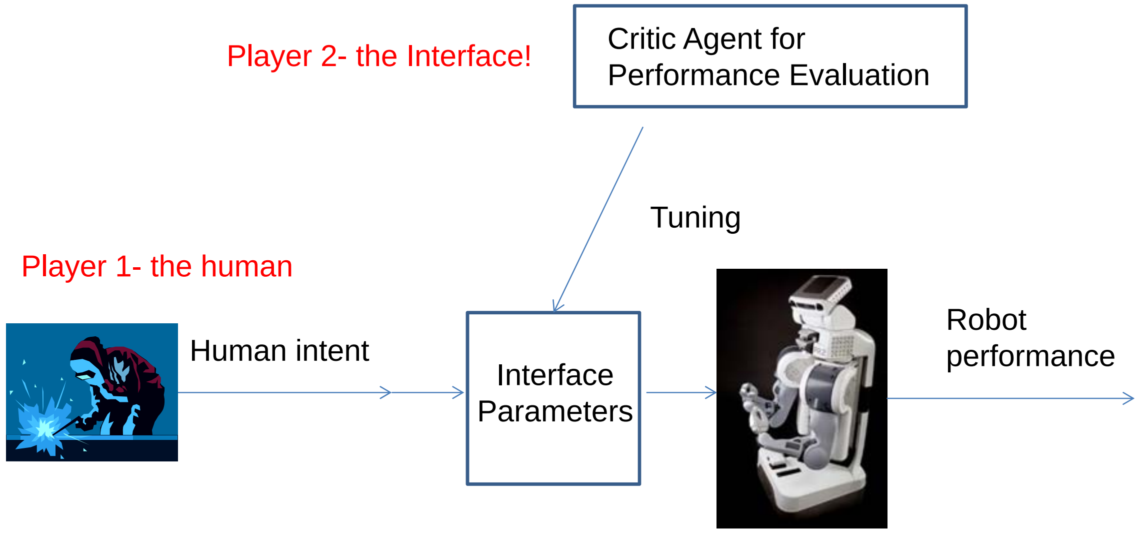
For a given task, after learning, the human model has three parts

1. Time delay due to vision processing and reaction time
2. A first-order filter due to neuromuscular dynamics
3. A PD controller in the cerebellum

$$G_h'(s) = \frac{K_d s + K_p}{Ts + 1} e^{-Ls},$$

Adaptive Impedance Control as a 2-Player Game

A 2-Player Dynamic Game



Player 2- the Interface!

Critic Agent for Performance Evaluation

Tuning

Player 1- the human

Human intent

Interface Parameters

Robot performance

Graphical Coalitional Games

500 BC

孙子兵法

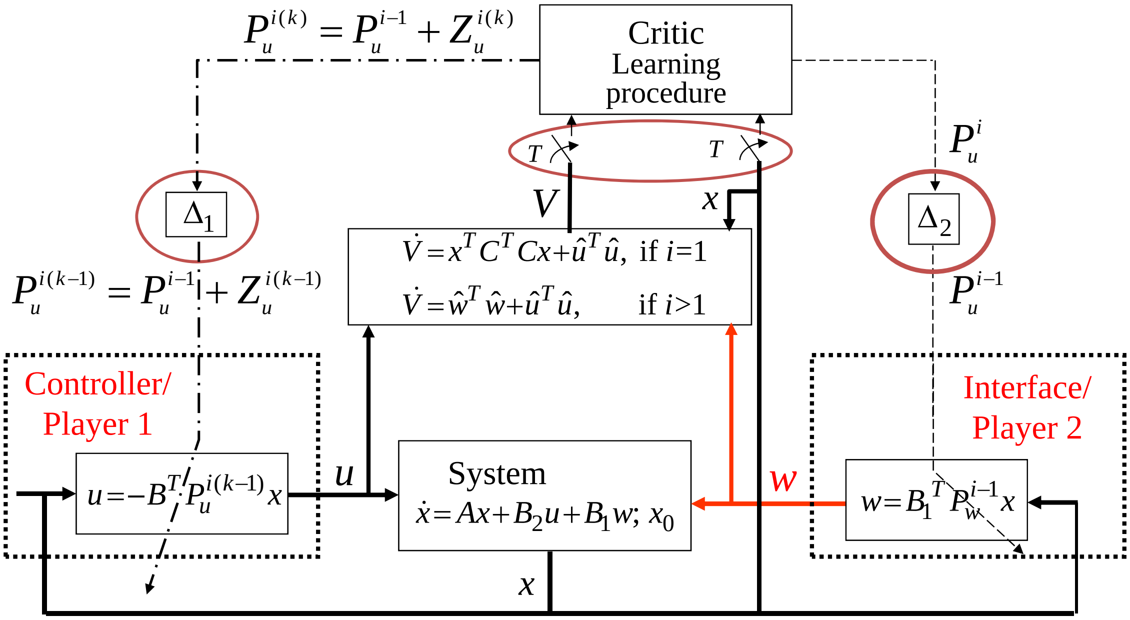
Sun Tz bin fa

Optimality and Games

Optimal Control is Effective for:
- Aircraft Autopilots
- Vehicle engine control
- Aerospace Vehicles
- Ship Control
- Industrial Process Control
- Robot Control

Multi-player Games Occur in:
- Economics
- Control Theory disturbance rejection
- Team games
- International politics
- Sports strategy

Actor-Critic Game structure - three time scales



$P_u^{i(k)} = P_u^{i-1} + Z_u^{i(k)}$

Critic
Learning
procedure

P_u^i

T
T
V
x

Δ_1

Δ_2

$\dot{V} = x^T C^T C x + \hat{u}^T \hat{u}$, if $i=1$

$\dot{V} = \hat{w}^T \hat{w} + \hat{u}^T \hat{u}$, if $i>1$

$P_u^{i(k-1)} = P_u^{i-1} + Z_u^{i(k-1)}$

P_u^{i-1}

Controller/
Player 1

$u = -B^T P_u^{i(k-1)} x$

u

System
$\dot{x} = Ax + B_2 u + B_1 w;\ x_0$

w

Interface/
Player 2

$w = B_1^T P_w^{i-1} x$

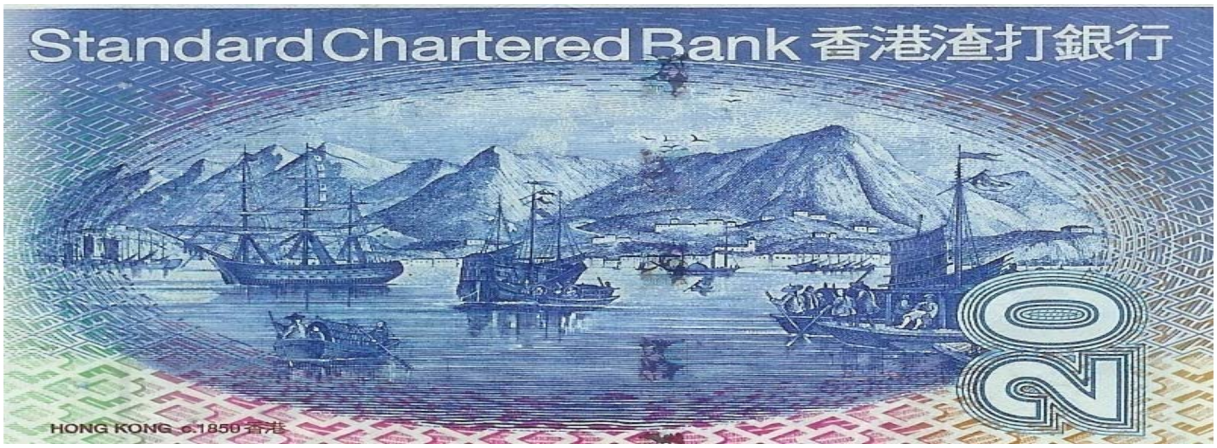
x

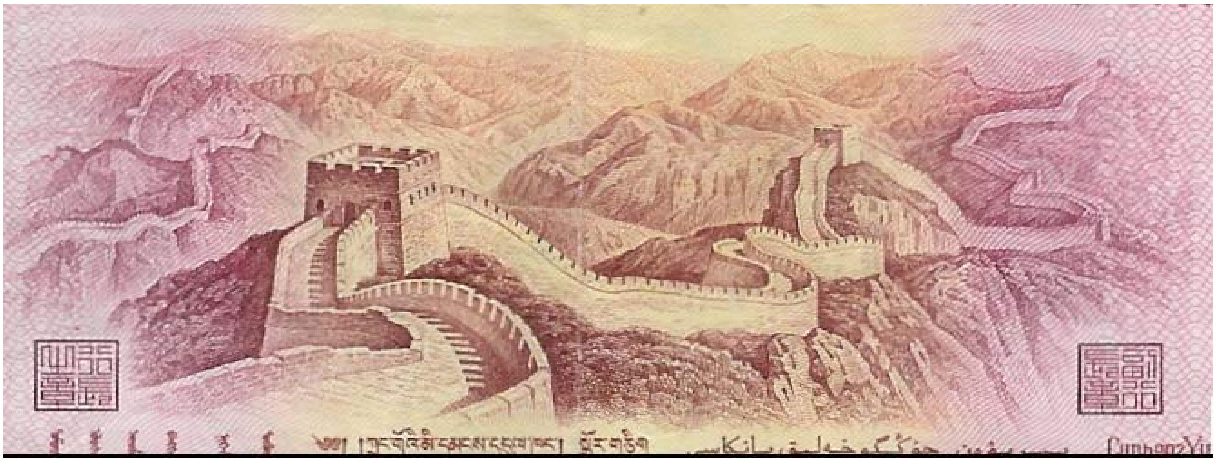
Our revels now are ended. These our actors,
As I foretold you, were all spirits, and
Are melted into air, into thin air.

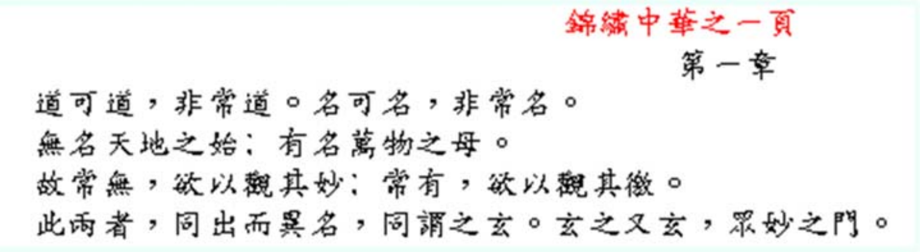
The cloud-capped towers, the gorgeous palaces,
The solemn temples, the great globe itself,
Yea, all which it inherit, shall dissolve,
And, like this insubstantial pageant faded,
Leave not a rack behind.

We are such stuff as dreams are made on,
and our little life is rounded with a sleep.

Prospero, in The Tempest, act 4, sc. 1, l. 152-6, Shakespeare


Standard Chartered Bank 香港渣打銀行
20
HONG KONG c.1850 香港





錦繡中華之一頁

第一章

道可道，非常道。名可名，非常名。
無名天地之始；有名萬物之母。
故常無，欲以觀其妙；常有，欲以觀其徼。
此兩者，同出而異名，同謂之玄。玄之又玄，眾妙之門。

The way that can be told is not the Constant Way
The name that can be named is not the Constant Name

For nameless is the true way
Beyond the myriad experiences of the world

To experience without intention is to sense the world

All experience is an arch
wherethrough gleams that untravelled land
whose margins fade forever as we move

Dao ke dao feichang dao
Ming ke ming feichang ming