Robot Learning and Control #3

Linear Quadratic Gaussian Control

Kazuki Shibata

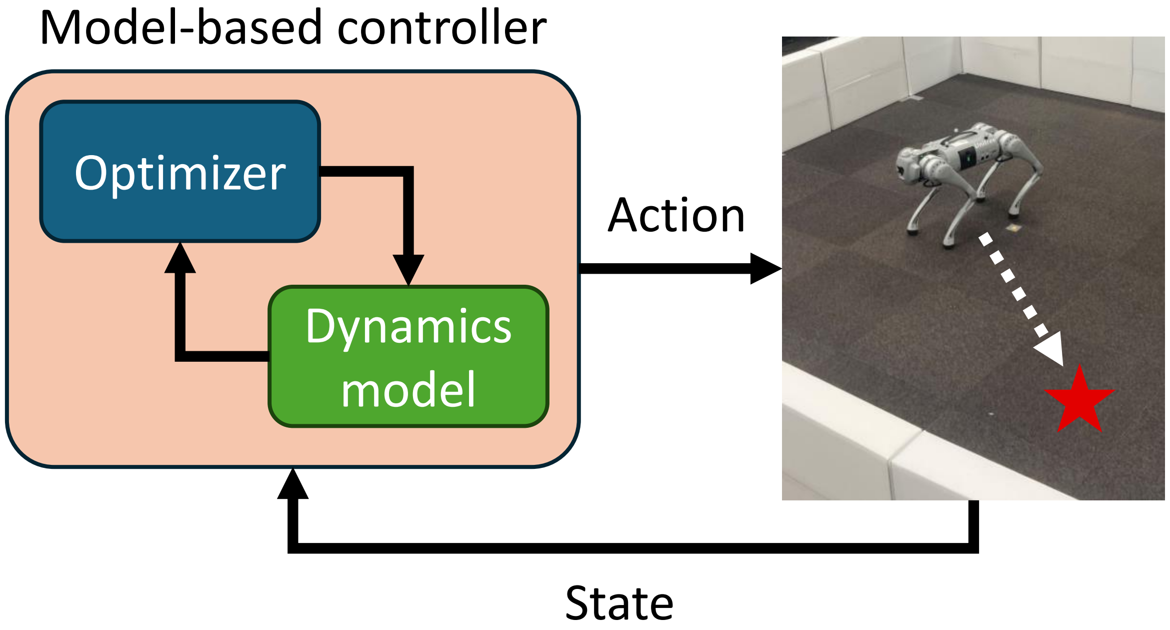
Robot Learning Lab

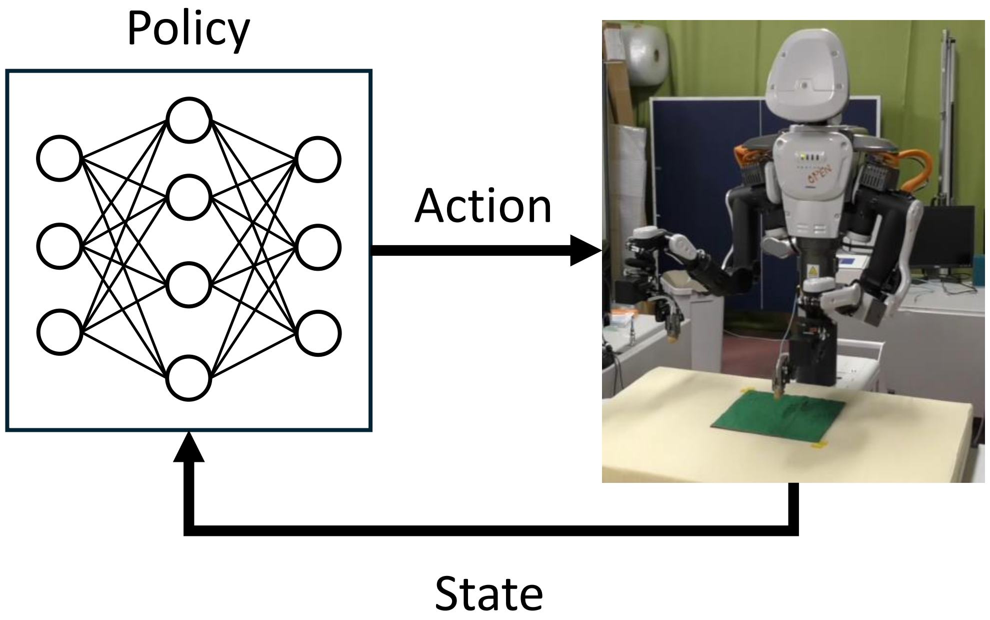
Nov. 22, 2024

General announcement

- The attendance is evaluated by **submitting the report paper**.
- The report is evaluated by **the exercises during the course**.
- Never give up!

Robot control





<u>Model-based approach</u>

✓ Provably guarantee the convergence

✗ Inapplicable without dynamics model

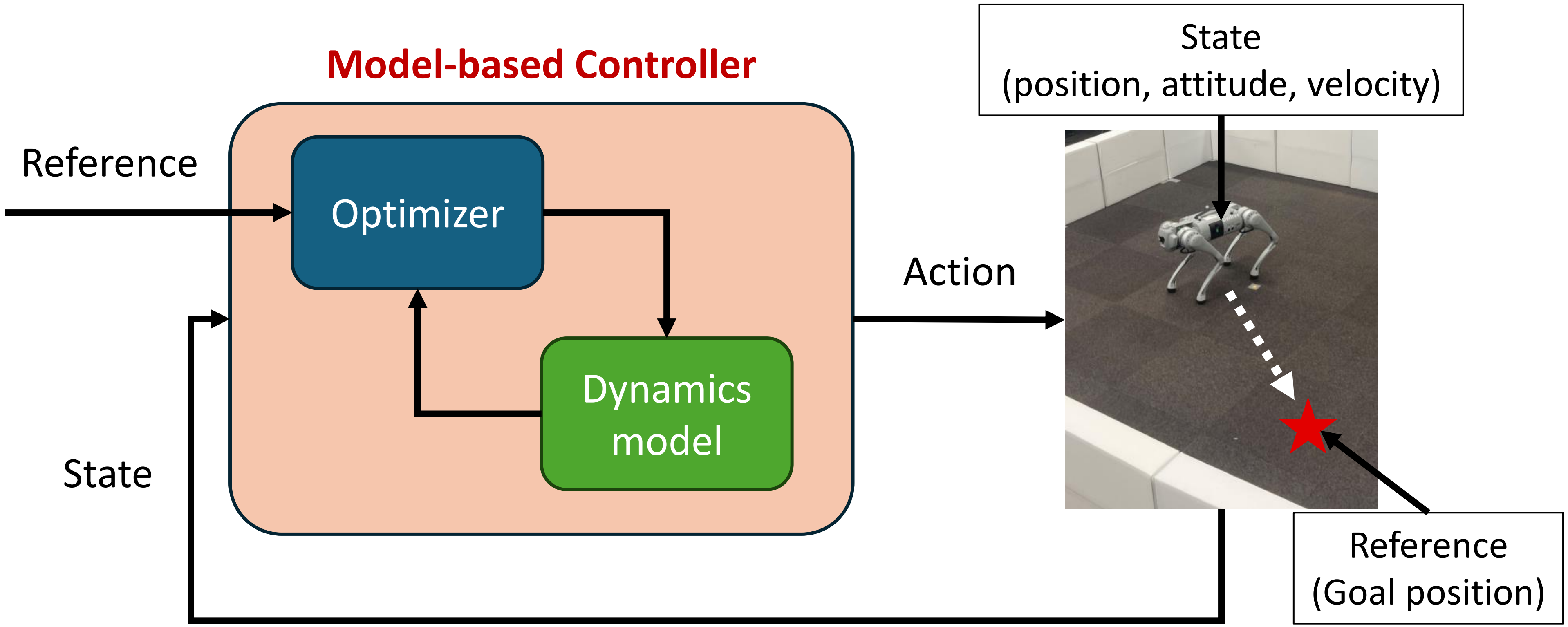
<u>Model-free approach</u>

✓ Applicable without dynamics model

✗ Unable to guarantee the convergence

This lecture focuses on the model-based approach based on the dynamics model.

Model-based Robot Control



How can we design robot controller by using dynamics model ?

Today´s notebook

This lecture utilizes Google Colab to run example codes developed by AP Sasaki.

Lecture3.ipynb

ファイル　編集　表示　挿入　ランタイム　ツール　ヘルプ　変更は保存されません

+ コード　+ テキスト　ドライブにコピー

Point mass control problem: task

```
[ ] warnings.simplefilter('ignore')
    env = gnwrapper.Monitor(gym.make('PosiEnv-v0'), directory="./random_action_video", video_callable=lambda a: True)
    # env = gym.make('PosiEnv-v0')
    env.name_prefix = "random_action"

    obs = env.reset()
    done = False

    while not done:
        env.render()
        a = env.action_space.sample()
        obs, r, done, _ = env.step(a)

    env.display()
```

URL: https://colab.research.google.com/drive/1jIUhqFEJ8mDMW-IF2GkCkytKvuDjvf7Y?usp=sharing

Table of Contents

Linear Gaussian Dynamics with Control

Consider a state-space model that involves zero-mean Gaussian noise:

$$\boldsymbol{x}_{t+1} = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{B}\boldsymbol{u}_t + \boldsymbol{C}\boldsymbol{w}_t \tag{1}$$

where, $\boldsymbol{x}$ is state, $\boldsymbol{u}$ is control input, and $\boldsymbol{w}$ is noise that follows $w \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$

It can also be written as follows:

$$p(\boldsymbol{x}_{t+1} | \boldsymbol{x}_t, \boldsymbol{u}_t) = \mathcal{N}(\boldsymbol{x}_{t+1};\ \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{B}\boldsymbol{u}_t, \sigma^2 \boldsymbol{C}\boldsymbol{C}^\top) \tag{2}$$

Let‘s confirm how to derive (2).

$$E[\boldsymbol{x}_{t+1}] = \int \boldsymbol{x}_{t+1}\, p(\boldsymbol{w}_t) d\boldsymbol{w}_t \tag{3}$$

$$= \int (\boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{B}\boldsymbol{u}_t + \boldsymbol{C}\boldsymbol{w}_t) p(\boldsymbol{w}_t) d\boldsymbol{w}_t \tag{4}$$

Since $\int p(\boldsymbol{w}_t) d\boldsymbol{w}_t = 1$ and $\int \boldsymbol{w}_t p(\boldsymbol{w}_t) d\boldsymbol{w}_t = E[\boldsymbol{w}_t] = 0$,

$$E[\boldsymbol{x}_{t+1}] = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{B}\boldsymbol{u}_t \tag{5}$$

Exercise 3-1: Linear Gaussian Dynamics with Control

Derive

$$V(\boldsymbol{x}_{t+1}) = \sigma^2 \boldsymbol{C}\boldsymbol{C}^\top$$

using $V(\boldsymbol{x}_{t+1}) = \int (\boldsymbol{x}_{t+1} - E[\boldsymbol{x}_{t+1}])(\boldsymbol{x}_{t+1} - E[\boldsymbol{x}_{t+1}])^\top p(\boldsymbol{w}_t) d\boldsymbol{w}_t$

- You now have 5 minutes to write up your report
- Write your name and student number in your report.

Exercise 3-1: Answer: Linear Gaussian Dynamics with Control

The variance of $\boldsymbol{x}_{t+1}$ is computed as follows:

$$V(\boldsymbol{x}_{t+1}) = \int (\boldsymbol{x}_{t+1} - E[\boldsymbol{x}_{t+1}])(\boldsymbol{x}_{t+1} - E[\boldsymbol{x}_{t+1}])^\top p(\boldsymbol{w}_t) d\boldsymbol{w}_t \tag{6}$$

$$= \boldsymbol{C} \left\{ \int \boldsymbol{w}_t \boldsymbol{w}_t^\top p(\boldsymbol{w}_t) d\boldsymbol{w}_t \right\} \boldsymbol{C}^\top \tag{7}$$

The variance of $\boldsymbol{w}_t$ is computed as follows:

$$V[\boldsymbol{w}_t] = \int (\boldsymbol{w}_t - E[\boldsymbol{w}_t])(\boldsymbol{w}_t - E[\boldsymbol{w}_t])^\top p(\boldsymbol{w}_t) d\boldsymbol{w}_t \tag{8}$$

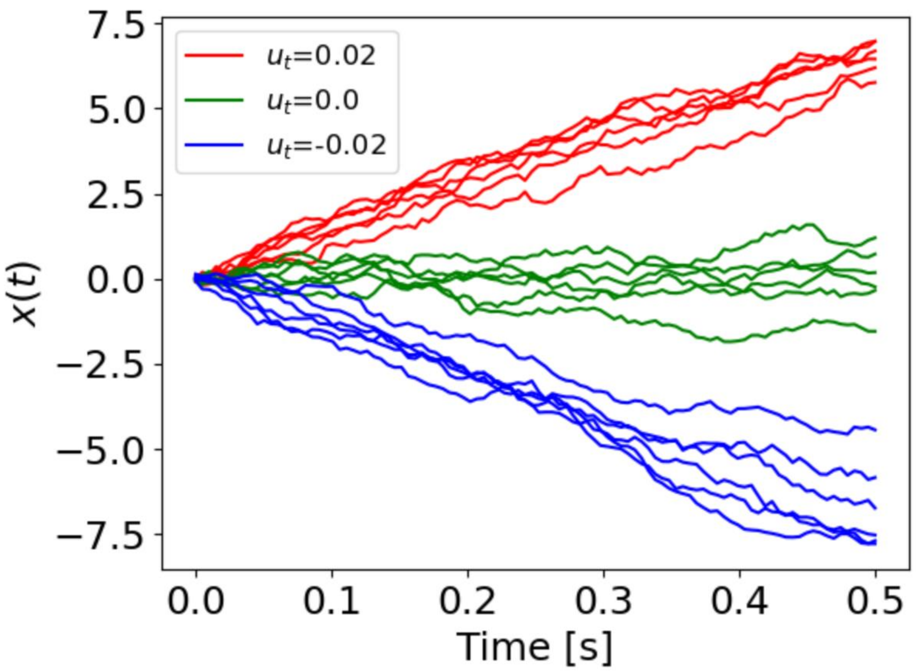
$$= \int \boldsymbol{w}_t \boldsymbol{w}_t^\top p(\boldsymbol{w}_t) d\boldsymbol{w}_t = \sigma^2 \mathbf{I} \tag{9}$$

Therefore, we obtain

$$V(\boldsymbol{x}_{t+1}) = \sigma^2 \boldsymbol{C} \boldsymbol{C}^\top \tag{10}$$

Simulation: Linear Gaussian Dynamics with Control

Let's see Google Colab to confirm the behavior of linear Gaussian dynamics.



How can we design actions to control the state to the desired position?

Table of Contents

Optimal Control Problem

Optimal control problem involves finding a control law for a dynamical system over a given period such that an objective function is optimized.

Derive a control sequence $\boldsymbol{u}_{1:T} \coloneqq (u_1, \cdots, u_T)$ that minimizes

$$J_0(\boldsymbol{u}_{1:T}) = E\left[\sum_{t=1}^{T} c(\boldsymbol{x}_t, \boldsymbol{u}_t)\right] \tag{11}$$

where $c(\boldsymbol{x}_t, \boldsymbol{u}_t)$ is immediate cost function and the state follows $p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, \boldsymbol{u}_t)$.

<u>How can we solve the optimal control problem?</u>

Naive Solutions

- **Random search**: generate random candidates of $\boldsymbol{u}_{1:T}$ and evaluate J_0 for each trial, and find the best one.
- **Direct search**: take the partial derivative of $\frac{\partial J_0}{\partial \boldsymbol{u}_{1:T}}$ and find $\boldsymbol{u}_{1:T}$ that makes it zero.

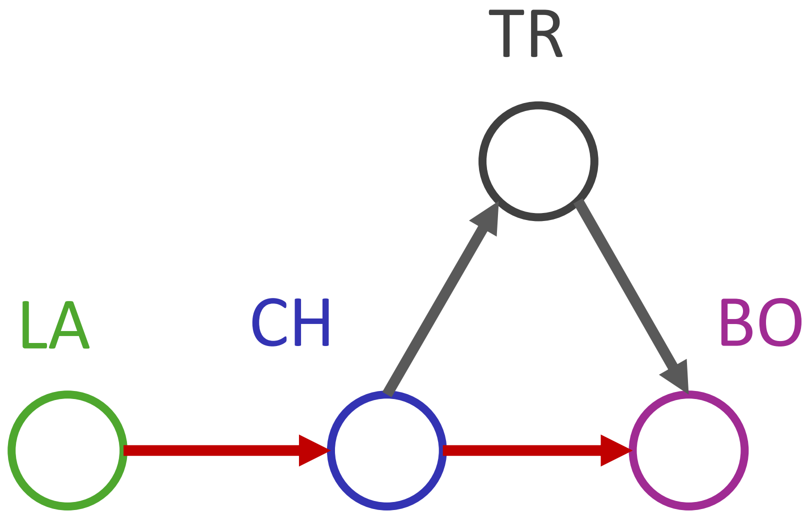
Both methods are not scalable for a large length of horizon T...

Table of Contents

Principle of Optimality

Sub-solutions of a whole optimal solution of the problem are themselves optimal solutions for their Sub-problems.

- Intuitive Example:
  Let's consider the optimal route from LA to BO.
  - Whole optimal solution is LA → CH → BO
  - If we know the shortest path from CH to BO, we could more easily find the shortest path from LA to BO.



We can solve the optimal problem more easily by considering it backward in time!

Dynamic Programming Algorithm

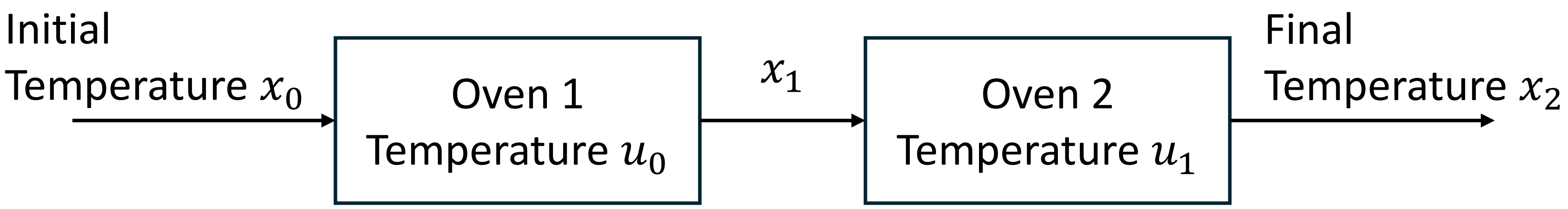
- Dynamic Programming:
  - if a problem can be solved optimally by breaking it into sub-problems, an efficient solution method is **recursively** finding the optimal solutions to the sub-problems
- Bellman equation:
  - consider a value function (total cost)

$$v_t(\boldsymbol{x}_t) = \min_{\boldsymbol{u}_{t:T}} E\left[\sum_{k=t}^{T} c(\boldsymbol{x}_k, \boldsymbol{u}_k)\right] \tag{8}$$

  - recursion formula of value function backward in time is as follows:

$$v_t(\boldsymbol{x}_t) = \min_{\boldsymbol{u}_t}[c(\boldsymbol{x}_t, \boldsymbol{u}_t) + E[v_{t+1}(\boldsymbol{x}_{t+1})]] \tag{9}$$

Exercise 3-2: Two-ovens Problem (deterministic)



The temperature of the material evolves according to:

$$x_{t+1} = (1-a)x_t + \ au_t \tag{10}$$

where a is a scalar in $0 < a < 1$, and $t = 0,1$.
The objective is to get the final temperature x_2 close to a given target T, while reducing energy.
This is expressed by a cost function of the form

$$c_2(x_2, u_2) = \ c_2(x_2) = r(x_2 - T)^{\ 2}, c_1(x_1, u_1) = \ c_1(u_1) = \ u_1^2, c_0 = \ u_0^2 \tag{11}$$

where r is a coefficient.
Please find optimal u_0 and u_1 by Bellman equation (10min)

Dimitri P. Bertsekas 2005

Exercise 3-2: Answer: Two-ovens Problem (deterministic)

Since $t = 2$ is the terminal time, the value function $v_2(x_2)$ is equal to cost $c_2(x_2)$:

$$v_2(x_2) = r(x_2 - T)^2 \tag{12}$$

Find the optimal control input at $t = 1$:

$$v_1(x_1) = \min_{u_1}\left[u_1^2 + v_2(x_2)\right] \tag{13}$$

$$= \min_{u_1}\left[u_1^2 + r\{(1-a)x_1 + au_1 - T\}^2\right] \tag{14}$$

con’t

Take derivative w.r.t. u_1 and set it 0:

$$\frac{\partial v_1}{\partial u_1} = 2u_1 + 2ra\{(1-a)x_1 + au_1 - T\}$$

$$= 2\left[\left(1+ra^2\right)u_1 - ra\{T-(1-a)x_1\}\right] = 0$$

Then, we obtain the optimal control input as follows:

$$u_1^* = \frac{ra(T\ -(1-a)x_1)}{1+ra^2} \tag{15}$$

con't

By substituting u_1^* into $v_1(x_1)$, we obtain $v_1(x_1)$ as follows:

$$v_1(x_1) = u_1^* + r\{(1-a)x_1 + au_1^* - T\}^2$$

$$\begin{aligned}
(1-a)x_1 + au_1^* - T &= (1-a)x_1 + a\frac{ra(T-(1-a)x_1)}{1+ra^2} - T \\
&= \frac{(1-a)\left(1+ra^2\right)x_1 + ra^2(T-(1-a)x_1)}{1+ra^2} - T \\
&= \frac{ra^2T + (1+ra^2-a-ra^2-ra^3+ra^3)x_1}{1+ra^2} - T \\
&= \frac{ra^2T + (1-a)x_1}{1+ra^2} - T \\
&= \frac{(1-a)x_1 - T}{1+ra^2}
\end{aligned}$$

con't

$$
\begin{aligned}
v_1(x_1) &= u_1^* + r\frac{\{(1-a)x_1 - T\}^2}{(1+ra^2)^2} \\
&= \frac{r^2a^2\left((1-a)x_1 - T\right)^2}{(1+ra^2)^2} + r\frac{\{(1-a)x_1 - T\}^2}{(1+ra^2)^2} \\
&= \frac{r\left(1+ra^2\right)\{(1-a)x_1 - T\}^2}{(1+ra^2)^2} \\
&= \frac{r\{(1-a)x_1 - T\}^2}{1+ra^2} \qquad (16)
\end{aligned}
$$

con't

Consider one-step backward problem at $t = 0$:

$$v_0(x_0) = \min_{u_0}\left[u_0^2 + \frac{r\{(1-a)x_1 - T\}^2}{1+ ra^2}\right]$$

$$= \min_{u_0}\left[u_0^2 + \frac{r((1-a)^2 x_0 + (1-a)au_0 - T)^2}{1+ ra^2}\right] \tag{17}$$

Take derivative w.r.t. u_0 and set it 0:

$$2u_0^* + \frac{2}{1+ra^2}(1-a)ar\{(1-a)^2 x_0 + (1-a)au_0^* - T\} = 0$$

$$\left\{1 + \frac{(1-a)^2 ra^2}{1+ra^2}\right\}u_0^* = \frac{r(1-a)a\{T-(1-a)^2 x_0\}}{1+ra^2}$$

$$u_0^* = \frac{r(1-a)a\left(T - (1-a)^2 x_0\right)}{1+ ra^2(1+(1-a)^2)} \tag{18}$$

Table of Contents

Linear Quadratic Gaussian Control (LQG)

Linear Quadratic Gaussian Control is one of the most fundamental optimal control methods. It concerns

- linear Gaussian dynamics

$$\boldsymbol{x}_{t+1} = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{B}\boldsymbol{u}_t + \boldsymbol{C}\boldsymbol{w}_t, \qquad \boldsymbol{w}_t \sim \mathcal{N}(0, \boldsymbol{I}) \tag{19}$$

- quadratic cost

$$c(\boldsymbol{x}_t, \boldsymbol{u}_t) = \boldsymbol{x}_t^\top \boldsymbol{Q} \boldsymbol{x}_t + \boldsymbol{u}_t^\top \boldsymbol{R} \boldsymbol{u}_t \tag{20}$$

Derive a control sequence $\boldsymbol{u}_{1:T}$ that minimizes

$$J_0(\boldsymbol{x}) = E\left[\sum_{t=1}^{T} c(\boldsymbol{x}_t, \boldsymbol{u}_t)\right] \tag{21}$$

Exercise 3-3: LQG

Assume the optimal value function is a second-order polynomial function of state:

$$v_{t+1}(\boldsymbol{x}_{t+1}) = \boldsymbol{x}_{t+1}^{\top}\boldsymbol{S}_{t+1}\boldsymbol{x}_{t+1} + s_{t+1} \tag{22}$$

where $\boldsymbol{S}_{t+1}^{\top} = \boldsymbol{S}_{t+1}$.

Derive the optimal control input $\boldsymbol{u}_t^*$ using Bellman equation.

- You now have 10 minutes to write up your report

Use the following formula:

$$E\left[\boldsymbol{w}_t^{\top}\boldsymbol{M}\boldsymbol{w}_t\right] = \mathrm{tr}(\boldsymbol{M})$$

Exercise 3-3: Answer: LQG

Apply Bellman equation:

$$v_t(\boldsymbol{x}_t) = \min_{\boldsymbol{u}_t} \left[c(\boldsymbol{x}_t, \boldsymbol{u}_t) + E[v_{t+1}(\boldsymbol{x}_{t+1})]\right] \tag{23}$$

$$= \min_{\boldsymbol{u}_t} \left[\boldsymbol{x}_t^\top \boldsymbol{Q} \boldsymbol{x}_t + \boldsymbol{u}_t^\top \boldsymbol{R} \boldsymbol{u}_t + E\left[\boldsymbol{x}_{t+1}^\top \boldsymbol{S}_{t+1} \boldsymbol{x}_{t+1} + s_{t+1}\right]\right] \tag{24}$$

Substitute $\boldsymbol{x}_{t+1} = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{B}\boldsymbol{u}_t + \boldsymbol{C}\boldsymbol{w}_t$

$$E\left[\boldsymbol{x}_{t+1}^\top \boldsymbol{S}_{t+1} \boldsymbol{x}_{t+1}\right] = E[(\boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{B}\boldsymbol{u}_t + \boldsymbol{C}\boldsymbol{w}_t)^\top \boldsymbol{S}_{t+1} (\boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{B}\boldsymbol{u}_t + \boldsymbol{C}\boldsymbol{w}_t)] \tag{25}$$

$$\begin{aligned} &= E[\boldsymbol{x}_t^\top \boldsymbol{A}^\top \boldsymbol{S}_{t+1} \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{x}_t^\top \boldsymbol{A}^\top \boldsymbol{S}_{t+1} \boldsymbol{B}\boldsymbol{u}_t + \boldsymbol{x}_t^\top \boldsymbol{A}^\top \boldsymbol{S}_{t+1} \boldsymbol{C}\boldsymbol{w}_t \\ &\quad + \boldsymbol{u}_t^\top \boldsymbol{B}^\top \boldsymbol{S}_{t+1} \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{u}_t^\top \boldsymbol{B}^\top \boldsymbol{S}_{t+1} \boldsymbol{B}\boldsymbol{u}_t + \boldsymbol{u}_t^\top \boldsymbol{B}^\top \boldsymbol{S}_{t+1} \boldsymbol{C}\boldsymbol{w}_t \\ &\quad + \boldsymbol{w}_t^\top \boldsymbol{C}^\top \boldsymbol{S}_{t+1} \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{w}_t^\top \boldsymbol{C}^\top \boldsymbol{S}_{t+1} \boldsymbol{B}\boldsymbol{u}_t + \boldsymbol{w}_t^\top \boldsymbol{C}^\top \boldsymbol{S}_{t+1} \boldsymbol{C}\boldsymbol{w}_t] \end{aligned} \tag{26}$$

Since $\boldsymbol{w}_t \sim \mathcal{N}(0, \boldsymbol{I})$,

$$E\left[\boldsymbol{x}_t^\top \boldsymbol{A}^\top \boldsymbol{S}_{t+1} \boldsymbol{C}\boldsymbol{w}_t\right] = \boldsymbol{x}_t^\top \boldsymbol{A}^\top \boldsymbol{S}_{t+1} \boldsymbol{C} E[\boldsymbol{w}_t] = 0$$

$$E\left[\boldsymbol{u}_t^\top \boldsymbol{B}^\top \boldsymbol{S}_{t+1} \boldsymbol{C}\boldsymbol{w}_t\right] = \boldsymbol{u}_t^\top \boldsymbol{B}^\top \boldsymbol{S}_{t+1} \boldsymbol{C} E[\boldsymbol{w}_t] = 0$$

$$E\left[\boldsymbol{w}_t^\top \boldsymbol{C}^\top \boldsymbol{S}_{t+1} \boldsymbol{A}\boldsymbol{x}_t\right] = \boldsymbol{x}_t^\top \boldsymbol{A}^\top \boldsymbol{S}_{t+1} \boldsymbol{C} E[\boldsymbol{w}_t] = 0$$

$$E\left[\boldsymbol{w}_t^\top \boldsymbol{C}^\top \boldsymbol{S}_{t+1} \boldsymbol{B}\boldsymbol{u}_t\right] = \boldsymbol{u}_t^\top \boldsymbol{B}^\top \boldsymbol{S}_{t+1} \boldsymbol{C} E[\boldsymbol{w}_t] = 0$$

Exercise 3-3: Answer: LQG

Using the formula $E\left[\boldsymbol{w}_t^\top \boldsymbol{M} \boldsymbol{w}_t\right] = \mathrm{tr}(\boldsymbol{M})$,

$$E\left[\boldsymbol{w}_t^\top \boldsymbol{M} \boldsymbol{w}_t\right] = \mathrm{tr}(\boldsymbol{C}^\top \boldsymbol{S}_{t+1} \boldsymbol{C}) \tag{27}$$

Therefore, we obtain

$$E\left[\boldsymbol{x}_{t+1}^\top \boldsymbol{S}_{t+1} \boldsymbol{x}_{t+1}\right] = \boldsymbol{x}_t^\top \boldsymbol{A}^\top \boldsymbol{S}_{t+1} \boldsymbol{A} \boldsymbol{x}_t + 2\boldsymbol{x}_t^\top \boldsymbol{A}^\top \boldsymbol{S}_{t+1} \boldsymbol{B} \boldsymbol{u}_t + \boldsymbol{u}_t^\top \boldsymbol{B}^\top \boldsymbol{S}_{t+1} \boldsymbol{B} \boldsymbol{u}_t + \mathrm{tr}(\boldsymbol{C}^\top \boldsymbol{S}_{t+1} \boldsymbol{C}) \tag{28}$$

Take derivative w.r.t. $\boldsymbol{u}_t$ and set it 0:

$$\frac{\partial}{\partial \boldsymbol{u}_t}(c(\boldsymbol{x}_t, \boldsymbol{u}_t) + E[v_{t+1}(\boldsymbol{x}_{t+1})]) = 2\boldsymbol{R}\boldsymbol{u}_t + 2\boldsymbol{B}^\top \boldsymbol{S}_{t+1} \boldsymbol{A} \boldsymbol{x}_t + 2\boldsymbol{B}^\top \boldsymbol{S}_{t+1} \boldsymbol{B} \boldsymbol{u}_t = 0 \tag{29}$$

$$\boldsymbol{u}_t^* = -\boldsymbol{L}_t \boldsymbol{x}_t \tag{30}$$

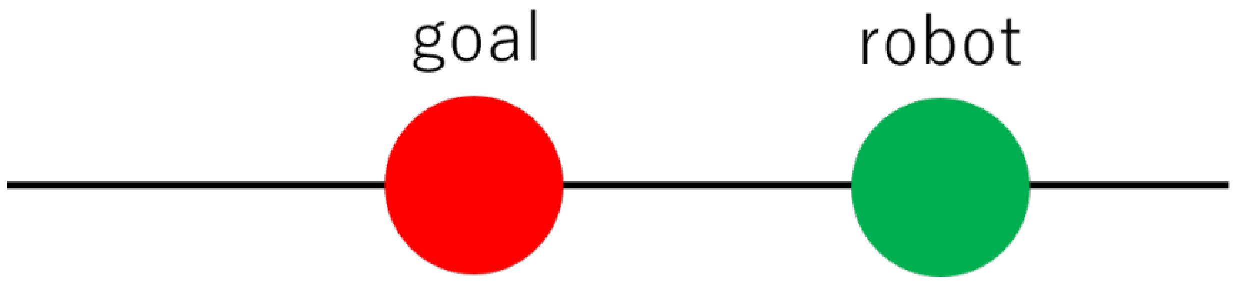
where

$$\boldsymbol{L}_t = (\boldsymbol{B}^\top \boldsymbol{S}_{t+1} \boldsymbol{B} + \boldsymbol{R})^{-1} (\boldsymbol{B}^\top \boldsymbol{S}_{t+1} \boldsymbol{A}) \tag{31}$$

Therefore, we obtain the optimal control input $\boldsymbol{u}_t^*$.

Simulation: A point-mass control problem

- $\boldsymbol{x} = \begin{bmatrix} pos \\ vel \end{bmatrix}$
- $\boldsymbol{u} = [acc]$
- $\boldsymbol{A} = \begin{bmatrix} 1 & 0.3 \\ 0 & 1 \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} 0 \\ 0.3 \end{bmatrix}, \boldsymbol{C} = \begin{bmatrix} 0.05 & 0 \\ 0 & 0.01 \end{bmatrix}$
- $\boldsymbol{Q} = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.08 \end{bmatrix}, R = 0.02$



Let's see Google Colab to confirm the control result using LQG.

Linear to Nonlinear?

Robot dynamics model often follows nonlinear dynamics.

Consider the following extended model:

$$\boldsymbol{x}_{t+1} = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t) + C\boldsymbol{w}_t \tag{32}$$

where $\boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t) \neq \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{B}\boldsymbol{u}_t$

Can we solve the optimal control for such nonlinear systems?

→ it is mostly impossible in exact, but possible in approximation

Iterative Linear Quadratic Gaussian Control (iLQG)

The algorithms begins with an open-loop control sequences $\bar{\boldsymbol{u}}_{1:T}$, and the corresponding "zero-noise" trajectories $\bar{\boldsymbol{x}}_{1:T}$ obtained by $\bar{\boldsymbol{x}}_{t+1} = \boldsymbol{f}(\bar{\boldsymbol{x}}_t, \bar{\boldsymbol{u}}_t)$.

The state and control input is deviated by $\delta\boldsymbol{x}_t$ and $\delta\boldsymbol{u}_t$:

$$\boldsymbol{x}_t = \bar{\boldsymbol{x}}_t + \delta\boldsymbol{x}_t, \quad \boldsymbol{u}_t = \bar{\boldsymbol{u}}_t + \delta\boldsymbol{u}_t \tag{33}$$

$$\boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t) = \boldsymbol{f}(\bar{\boldsymbol{x}}_t + \delta\boldsymbol{x}_t, \bar{\boldsymbol{u}}_t + \delta\boldsymbol{u}_t)(\text{Tayler expansion}) \tag{34}$$

$$\approx \boldsymbol{f}(\bar{\boldsymbol{x}}_t, \bar{\boldsymbol{u}}_t) + \left.\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\right|_{\bar{\boldsymbol{x}}_t} \delta\boldsymbol{x}_t + \left.\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{u}}\right|_{\bar{\boldsymbol{u}}_t} \delta\boldsymbol{u}_t \tag{35}$$

$$= \bar{\boldsymbol{x}}_{t+1} + \boldsymbol{A}_t\delta\boldsymbol{x}_t + \boldsymbol{B}_t\delta\boldsymbol{u}_t \tag{36}$$

Then, we can derive a linear system dynamics of deviation state and action at around $\bar{\boldsymbol{x}}$, $\bar{\boldsymbol{u}}$ as follows:

$$\boldsymbol{x}_{t+1} = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t) + C\boldsymbol{w}_t \tag{37}$$

$$\bar{\boldsymbol{x}}_{t+1} + \delta\boldsymbol{x}_{t+1} \approx \bar{\boldsymbol{x}}_{t+1} + \boldsymbol{A}_t\delta\boldsymbol{x}_t + \boldsymbol{B}_t\delta\boldsymbol{u}_t + C\boldsymbol{w}_t \tag{38}$$

$$\delta\boldsymbol{x}_{t+1} = \boldsymbol{A}_t\delta\boldsymbol{x}_t + \boldsymbol{B}_t\delta\boldsymbol{u}_t + C\boldsymbol{w}_t \tag{39}$$

Todorov and Li, ACC2005

Exercise 3-4: iLQG

The cost function is also expanded as follows:

$$c(\boldsymbol{x}_t, \boldsymbol{u}_t) = \boldsymbol{x}_t^\top \boldsymbol{Q} \boldsymbol{x}_t + \boldsymbol{u}_t^\top \boldsymbol{R} \boldsymbol{u}_t \tag{40}$$

$$= (\bar{\boldsymbol{x}}_t + \delta \boldsymbol{x}_t)^\top \boldsymbol{Q} (\bar{\boldsymbol{x}}_t + \delta \boldsymbol{x}_t) + (\bar{\boldsymbol{u}}_t + \delta \boldsymbol{u}_t)^\top \boldsymbol{R} (\bar{\boldsymbol{u}}_t + \delta \boldsymbol{u}_t) \tag{41}$$

$$= \bar{\boldsymbol{x}}_t^\top \boldsymbol{Q} \bar{\boldsymbol{x}}_t + 2\bar{\boldsymbol{x}}_t^\top \boldsymbol{Q} \delta \boldsymbol{x}_t + \delta \boldsymbol{x}_t^\top \boldsymbol{Q} \delta \boldsymbol{x}_t + \bar{\boldsymbol{u}}_t^\top \boldsymbol{R} \bar{\boldsymbol{u}}_t$$

$$+ 2\bar{\boldsymbol{u}}_t^\top \boldsymbol{R} \delta \boldsymbol{u}_t + \delta \boldsymbol{u}_t^\top \boldsymbol{R} \delta \boldsymbol{u}_t \tag{42}$$

Assume the following value function:

$$v_t(\delta \boldsymbol{x}_t) = s_t + \delta \boldsymbol{x}_t^\top \boldsymbol{s}_t + \delta \boldsymbol{x}_t^\top \boldsymbol{S}_t \delta \boldsymbol{x}_t \tag{43}$$

Derive the optimal control input $\delta \boldsymbol{u}_t^*$ using Bellman equation.

- You now have 10 minutes to write up your report

Exercise 3-4: Answer: iLQG

Apply Bellman equation:

$$v_t(\delta \boldsymbol{x}_t) = \min_{\delta \boldsymbol{u}_t} \left[c(\boldsymbol{x}_t, \boldsymbol{u}_t) + E[v_{t+1}(\delta \boldsymbol{x}_{t+1})]\right] \tag{44}$$

$$= \min_{\delta \boldsymbol{u}_t} \left[c(\boldsymbol{x}_t, \boldsymbol{u}_t) + E\left[s_{t+1} + \delta \boldsymbol{x}_{t+1}^\top \boldsymbol{s}_{t+1} + \delta \boldsymbol{x}_{t+1}^\top \boldsymbol{S}_{t+1} \delta \boldsymbol{x}_{t+1}\right]\right] \tag{45}$$

Substitute $\delta \boldsymbol{x}_{t+1} = \boldsymbol{A}_t \delta \boldsymbol{x}_t + \boldsymbol{B}_t \delta \boldsymbol{u}_t + \boldsymbol{C} \boldsymbol{w}_t$

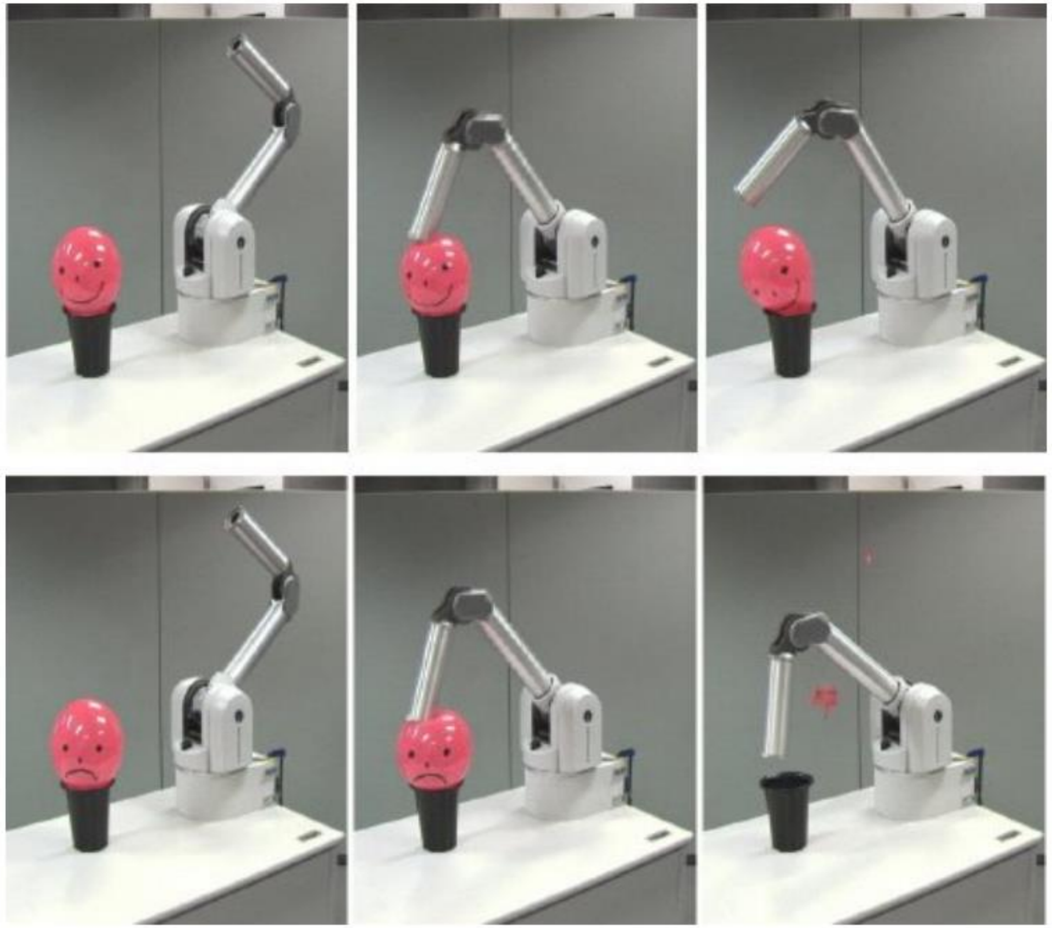
$$\begin{aligned} E[v_{t+1}(\delta \boldsymbol{x}_{t+1})] = {} & s_{t+1} + \delta \boldsymbol{x}_t^\top \boldsymbol{A}_t^\top \boldsymbol{s}_{t+1} + \delta \boldsymbol{u}_t^\top \boldsymbol{B}_t^\top \boldsymbol{s}_{t+1} + \delta \boldsymbol{x}_t^\top \boldsymbol{A}_t^\top \boldsymbol{S}_{t+1} \boldsymbol{A}_t \delta \boldsymbol{x}_t \\ & + \delta \boldsymbol{u}_t^\top \boldsymbol{B}_t^\top \boldsymbol{S}_{t+1} \boldsymbol{B}_t \delta \boldsymbol{u}_t + 2\delta \boldsymbol{x}_t^\top \boldsymbol{A}_t^\top \boldsymbol{S}_{t+1} \boldsymbol{B}_t \delta \boldsymbol{u}_t + \mathrm{tr}(\boldsymbol{C}^\top \boldsymbol{S}_{t+1} \boldsymbol{C}) \end{aligned} \tag{46}$$

From (42), (45) and (46), we obtain

$$\frac{\partial v_t(\delta \boldsymbol{x}_t)}{\partial \delta \boldsymbol{u}_t} = 2\boldsymbol{R}(\bar{\boldsymbol{u}}_t + \delta \boldsymbol{u}_t) + \boldsymbol{B}_t^\top \boldsymbol{s}_{t+1} + 2\boldsymbol{B}_t^\top \boldsymbol{S}_{t+1} \boldsymbol{B}_t \delta \boldsymbol{u}_t + 2\boldsymbol{B}_t^\top \boldsymbol{S}_{t+1} \boldsymbol{A}_t \delta \boldsymbol{x}_t = 0 \tag{47}$$

$$\delta \boldsymbol{u}_t^* = -\frac{1}{2}\left(\boldsymbol{R} + \boldsymbol{B}_t^\top \boldsymbol{S}_{t+1} \boldsymbol{B}_t\right)^{-1}\left(2\boldsymbol{R}\bar{\boldsymbol{u}}_t + \boldsymbol{B}_t^\top \boldsymbol{s}_{t+1}\right) - \left(\boldsymbol{R} + \boldsymbol{B}_t^\top \boldsymbol{S}_{t+1} \boldsymbol{B}_t\right)^{-1} \boldsymbol{B}_t^\top \boldsymbol{A}_t \delta \boldsymbol{x}_t \tag{48}$$

Robot Application



Mitrovic+, ICRA2010

Let’s see youtube video

References

[1] Y. Tsurumine, Y. Cui, E. Uchibe and T. Matsubara, "Deep dynamic policy programming for robot control with raw images," IROS2017, pp. 1545-1550.

[2] Dimitri P. Bertsekas: Dynamic programming and optimal control, 3rd Edition. Athena Scientific 2005, ISBN 1886529264, pp. I-XV, 1-543

[3] Emanuel Todorov: Optimal Control Theory, December 2006, DOI: 10.7551/mitpress/9780262042383.003.0012

[4] Emanuel Todorov, Weiwei Li: A generalized iterative LQG method for locally-optimal feedback control of constrained nonlinear stochastic systems. ACC 2005: 300-306vol.1

[5] Djordje Mitrovic, Sho Nagashima, Stefan Klanke, Takamitsu Matsubara, Sethu Vijayakumar: Optimal Feedback Control for anthropomorphic manipulators. ICRA 2010: 4143-4150