ORF 418 : OPTIMAL LEARNING

LECTURE 1 : September 3, 2025

INTRODUCTION

1) Deterministic Optimal Control
2) Stochastic Optimal Control
3) Examples :
    a) Brachistrochrone
    b) Moon Landing

I. DECISION MAKING

a) Overview

Many optimal decision problems have the following components:

(i) an action or control to be chosen by the controller (or the decision maker) based on the information gathered up to the decision time;

(ii) a finite collection in a certain set (generally a vector with real-valued entries) that describes the state of the system that is influenced by the actions chosen by the controller;

(iii) a function of the state and the action summarizing the goal of the decision maker.

Then, the optimal control problem is to appropriately chose admissible actions so as to optimize the goal function.

(b) Difficulties:

(a) often state dynamics are not exactly known and they need to be learned.

(b) in most cases, the system is subject to random perturbations and their statistical properties must be estimated.

(c) the goal function is not always clear and reasonable choices have to be made taking tractability into account.

(c) role of optimal control

Optimal control has been an effective modelling tool in many diverse areas including learning problems in engineering, in economic and finance theories as well as many quantitative approaches in social sciences. Many learning algorithms were initiated by intuition developed through optimal control. Reinforcement learning is a defining example.

II. THIS COURSE

1) We start with the classical simple deterministic control problem called the Linear Quadratic Regulator (LQR). We develop a powerful solution technique: dynamic programming of Richard Bellman.

2) Then we add randomness and study the Linear Quadratic Gaussian (LQG) control to develop tools to handle random elements.

3) We recall Bayesian Updating from probability theory to tackle learning problems in a stochastic environment. The linear Kalman Filter includes estimation in the LQG setting.

4) We then consider Markov Decision Processes (MDP) to go beyond the linear, quadratic setting which is widely used due its tractability.

5) General dynamic programming and effective algorithms are discussed in the MDP setting.

6) We then introduce Q-learning algorithms for MDP combining both learning and optimization.

III. OPTIMAL CONTROL

We use the following notation throughout the course:

(i) time is discrete and labelled by integers:

$t = 0, 1, 2, \ldots$

Sometimes there is a finite horizon often called T.

(ii) x_t = state of the system at time t.

In most examples, x_t is a d-dimensional vector with real-valued entries. But discrete structures are also useful and will be used: such as "on", "off" or "full" "empty", etc.

(iii) u_t = action taken at time t.

Importantly, actions use information avaliable up to time t and do not foresee the future. The controller however can estimate the future using the information. The set of admissible controllers is denoted by $\mathcal{U}$.

a) Deterministic Setting

Here state dynamics is given by

$$x_{t+1} = f(t, x_0, \dots x_t, u_0, \dots, u_t), \quad t = 0, 1, \dots$$

and we assume that x_0 is known and fixed.

The function f is either known or to be learned.

There are two important classes of goal functions that we consider.

(i) finite horizon: We fix horizon T and define

$$J_T(x, u) := \sum_{t=0}^{T-1} g(t, x_t, u_t) + h(x_T)$$

Then, the goal is to find u^* so that

$$J_T(x^*, u^*) \leq J_T(x, u) \quad \forall u \in \mathcal{U},$$

and x^* is the state process corresponding to the control process u^*. We assume that g and h are either known or to be learned.

(ii) infinite horizon: We fix a discount factor $\rho > 0$ (often $\rho < 1$), and define

$$J_\infty(x, u) := \sum_{t=0}^{\infty} \rho^t g(x_t, u_t).$$

Again we look for a minimizer u^*.

b) Stochastic Setting.

We assume that there is a sequence of random variables $w_0, w_1, \dots$ whose statics are either known or to be learned, and the dynamics in given by

$$x_{t+1} = \tilde{f}(t, x_0, \dots x_t, u_0, \dots u_t, w_0, \dots w_t) \quad t = 0, 1, \dots$$

and the goal functions are:

<u>(i) finite horizon</u>: We fix a horizon T and define

$$J_T(x,u) := \mathbb{E}\left[\sum_{t=0}^{T-1} g(t, x_t, u_t) + h(x_T)\right]$$

where $\mathbb{E}(\cdot\cdot)$ is the expected value.

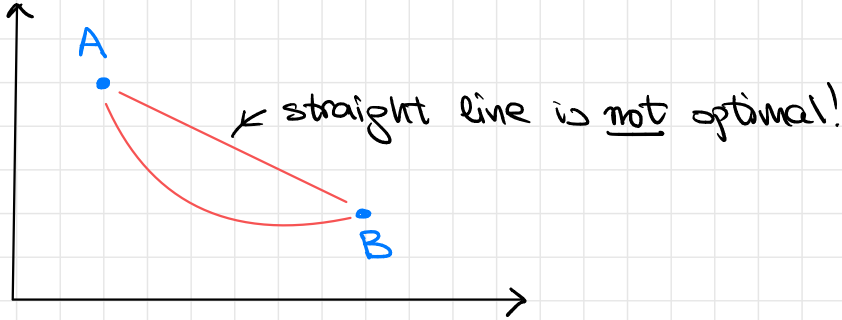
<u>(ii) infinite horizon</u>: We fix a discount factor $\rho > 0$ and define

$$J_\infty(x,u) := \mathbb{E}\left[\sum_{t=0}^{\infty} \rho^t g(x_t, u_t)\right].$$

IV. Examples.

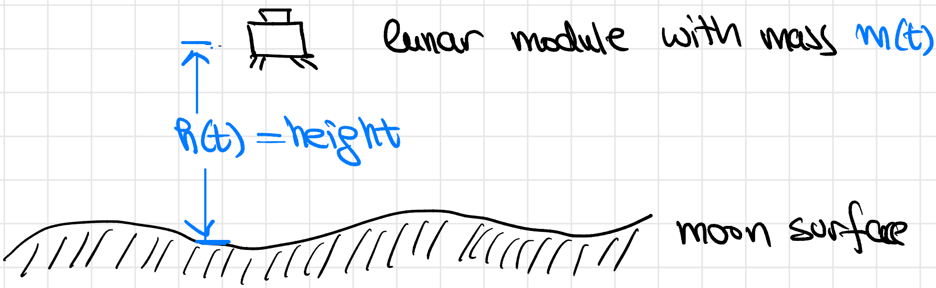
a) Brachistochrone Problem posed by Bernoulli

Find the path that connects an initial point A and a final point B so that a bead sliding on this path under gravity reaches B in shortest time.



See lecture notes for the solution. But we do not further study this example.

b) Moon landing problem was an important part of the success of the moon project of the 60's.



Newton law: $\quad m(t)\, h(t)'' = -g_{moon}\, m(t) + \alpha(t)$, where g_{moon} is the gravitational constant of moon and $\alpha(t)$ is the control.

Goal is to minimize total fuel $\int_0^{\tau} \alpha(t)\,dt$ and we want soft-landing: $v(\tau) = h'(\tau) = 0$ where τ is the landing time. See notes for the solution.