Control theory

The sensori-motor problem

Brain is a sensori-motor machine:

- perception
- action
- perception causes action, action causes perception
- much of this is learned

The sensori-motor problem

Brain is a sensori-motor machine:

- perception
- action
- perception causes action, action causes perception
- much of this is learned

Separately, we understand perception and action (somewhat):

- Perception is (Bayesian) statistics, information theory, max entropy

The sensori-motor problem

Brain is a sensori-motor machine:

- perception
- action
- perception causes action, action causes perception
- much of this is learned

Separately, we understand perception and action (somewhat):

- Perception is (Bayesian) statistics, information theory, max entropy
- Learning is parameter estimation

The sensori-motor problem

Brain is a sensori-motor machine:

- perception
- action
- perception causes action, action causes perception
- much of this is learned

Separately, we understand perception and action (somewhat):

- Perception is (Bayesian) statistics, information theory, max entropy
- Learning is parameter estimation
- Action is control theory?
  - limited use of adaptive control theory
  - intractability of optimal control theory
    - computing 'backward in time'.
    - representing control policies
    - model based vs. model free

The sensori-motor problem

Brain is a sensori-motor machine:

- perception
- action
- perception causes action, action causes perception
- much of this is learned

We seem to have no good theories for the combined sensori-motor problem.

- Sensing depends on actions
- Features depend on task(s)
- Action hierarchies, multiple tasks

The two realities of the brain

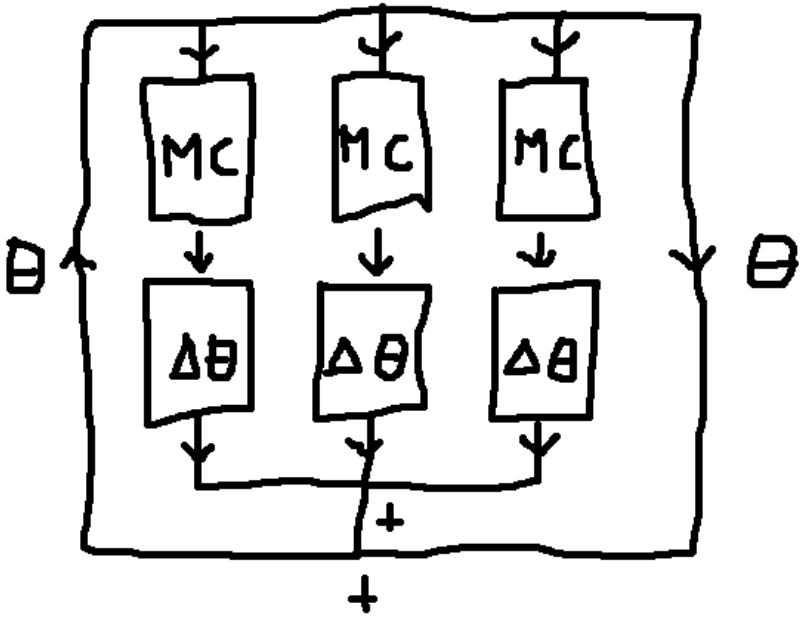
The neural activity of the brain simulates two realities:

- the physical world that enters through our senses
  - 'world' is everything outside the brain
  - neural activity depends on stimuli and internal model (perception, Bayesian inference, ...)
- the inner world that the brain simulates through its own activity
  - 'spontaneous activity', planning, thinking, 'what if...', etc.
  - neural activity is autonomous, depends on internal model

Integrating control, inference and learning

The inner world computation serves three purposes:

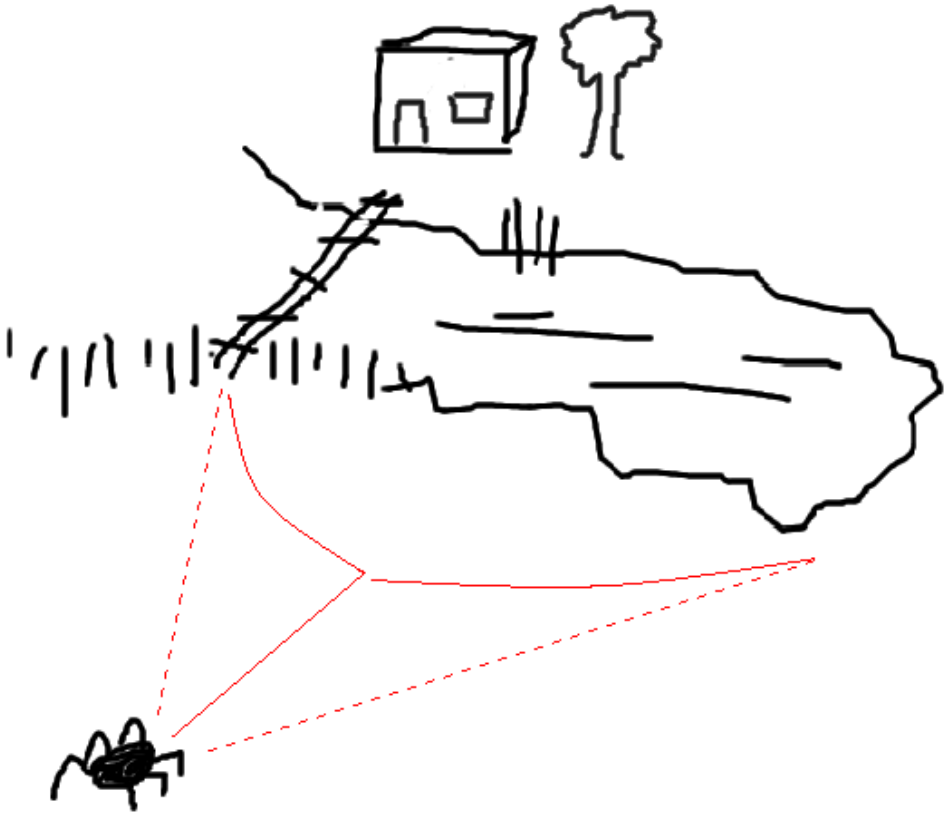
- the spontaneous activity is a type of Monte Carlo sampling
- Planning: compute actions for the current situation x from these samples
- Learning: improves the sampler using these samples



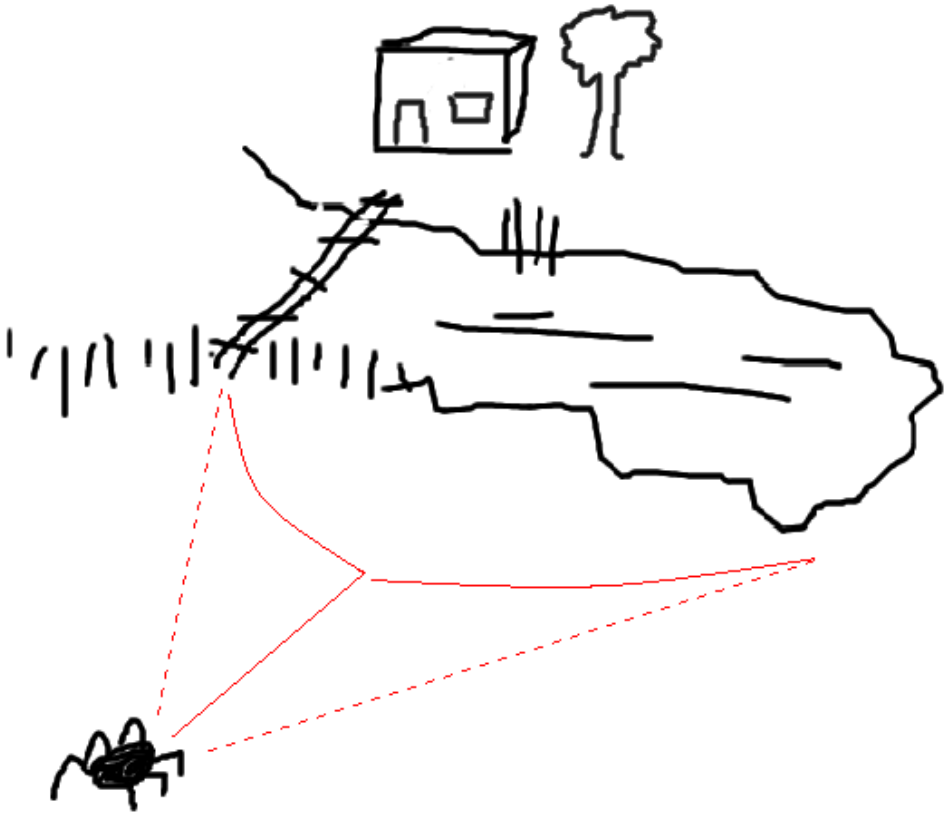
Optimal control theory

Given a current state and a future desired state, what is the best/cheapest/fastest way to get there.

Why stochastic optimal control?



Why stochastic optimal control?



Exploration
Learning

Optimal control theory

Hard problems:
- a learning and exploration problem
- a stochastic optimal control computation
- a representation problem $u(x,t)$

The idea: Control, Inference and Learning

Path integral control theory

Express a control computation as an inference computation.

Compute optimal control using MC sampling

The idea: Control, Inference and Learning

Path integral control theory

Express a control computation as an inference computation.

Compute optimal control using MC sampling

Importance sampling

Accellerate with importance sampling (=a state-feedback controller)

Optimal importance sampler is optimal control

The idea: Control, Inference and Learning

Path integral control theory

Express a control computation as an inference computation.

Compute optimal control using MC sampling

Importance sampling

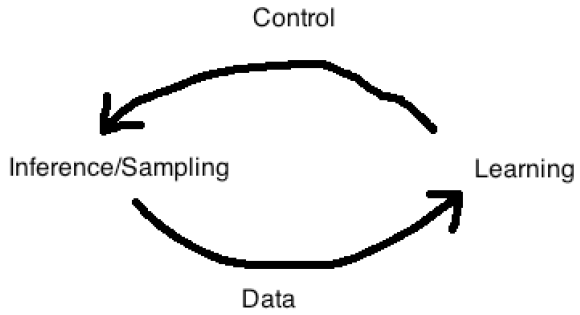
Accellerate with importance sampling (=a state-feedback controller)

Optimal importance sampler is optimal control

Learning

Learn the controller from self-generated data

Use Cross Entropy method for parametrized controller



Outline

Optimal control theory, discrete time
- Introduction of delayed reward problem in discrete time;
- Dynamic programming solution

Optimal control theory, continuous time
- Pontryagin maximum principle;

Stochastic optimal control theory
- Stochastic differential equations
- Kolmogorov and Fokker-Plack equations
- Hamilton-Jacobi-Bellman equation
- LQ control, Ricatti equation;
- Portfolio selection

Path integral/KL control theory
- Importance sampling
- KL control theory

Material

- H.J. Kappen. Optimal control theory and the linear Bellman Equation. In *Inference and Learning in Dynamical Models (Cambridge University Press 2010)*, edited by David Barber, Taylan Cemgil and Sylvia Chiappa
  `http://www.snn.ru.nl/~bertk/control/timeseriesbook.pdf`
- Dimitri Bertsekas, Dynamic programming and optimal control
- `http://www.snn.ru.nl/~bertk/machinelearning/`

Introduction

Optimal control theory: Optimize sum of a path cost and end cost. Result is optimal control sequence and optimal trajectory.

Input: Cost function.
Output: Optimal trajectory and controls.

Introduction

Control problems are delayed reward problems:

- Motor control: devise a sequece of motor commands to reach a goal
- finance: devise a sequence of buy/sell commands to maximize profit
- Learning, exploration vs. exploitation

Types of optimal control problems

Finite horizon (fixed horizon time):

- Dynamics and environment may depend explicitly on time.
- Optimal control depends explicitly on time.

Finite horizon (moving horizon):

- Dynamics and environment are static.
- Optimal control is time independent.

Infinite horizon:

- discounted reward, Reinforcement learning
- total reward, absorbing states
- average reward

Other issues:

- discrete vs. continuous state
- discrete vs. continuous time
- observable vs. partial observable
- noise

Discrete time control

Consider the control of a discrete time deterministic dynamical system:

$$x_{t+1} = x_t + f(t, x_t, u_t), \quad t = 0, 1, \ldots, T-1$$

x_t describes the *state* and u_t specifies the *control* or *action* at time t.

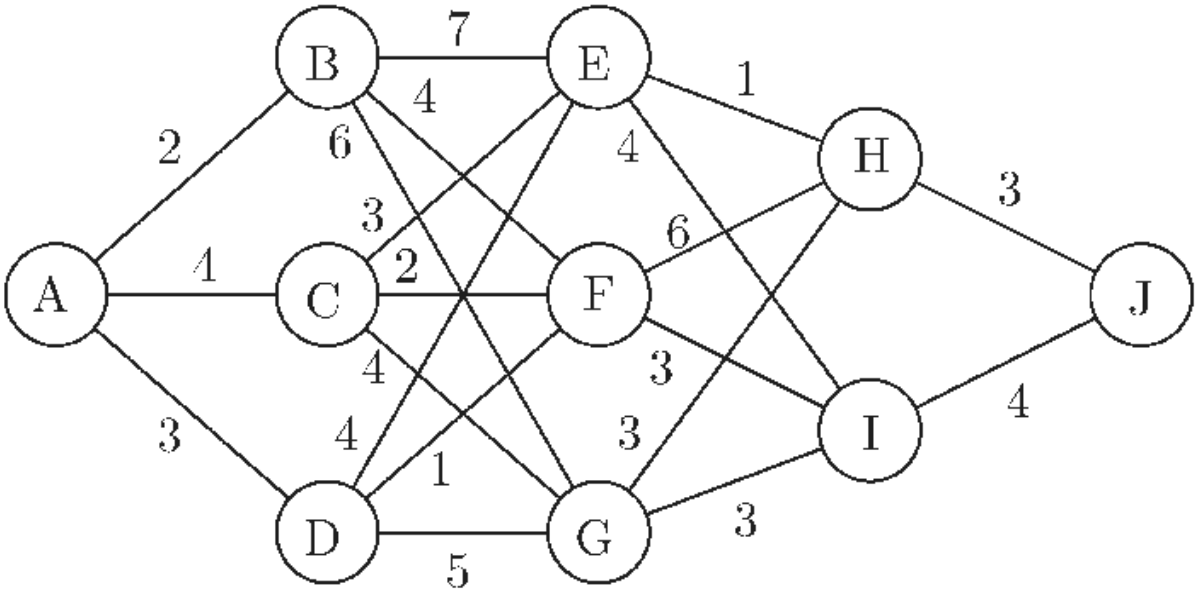
Given $x_{t=0} = x_0$ and $u_{0:T-1} = u_0, u_1, \ldots, u_T - 1$, we can compute $x_{1:T}$.

Define a cost for each sequence of controls:

$$C(x_0, u_{0:T-1}) = \phi(x_T) + \sum_{t=0}^{T-1} R(t, x_t, u_t)$$

The problem of optimal control is to find the sequence $u_{0:T-1}$ that minimizes $C(x_0, u_{0:T-1})$.

Dynamic programming



Find the minimal cost path from A to J.

$$C(J) = 0, C(H) = 3, C(I) = 4$$

$$C(F) = \min(6 + C(H), 3 + C(I))$$

Discrete time control

The optimal control problem can be solved by dynamic programming. Introduce the *optimal cost-to-go*:

$$J(t, x_t) = \min_{u_{t:T-1}} \left(\phi(x_T) + \sum_{s=t}^{T-1} R(s, x_s, u_s) \right)$$

which solves the optimal control problem from an intermediate time t until the fixed end time T, for all intermediate states x_t.

Then,

$$\begin{aligned} J(T, x) &= \phi(x) \\ J(0, x) &= \min_{u_{0:T-1}} C(x, u_{0:T-1}) \end{aligned}$$

Discrete time control

One can recursively compute $J(t, x)$ from $J(t+1, x)$ for all x in the following way:

$$\begin{aligned}
J(t, x_t) &= \min_{u_{t:T-1}} \left(\phi(x_T) + \sum_{s=t}^{T-1} R(s, x_s, u_s) \right) \\
&= \min_{u_t} \left(R(t, x_t, u_t) + \min_{u_{t+1:T-1}} \left[\phi(x_T) + \sum_{s=t+1}^{T-1} R(s, x_s, u_s) \right] \right) \\
&= \min_{u_t} \left(R(t, x_t, u_t) + J(t+1, x_{t+1}) \right) \\
&= \min_{u_t} \left(R(t, x_t, u_t) + J(t+1, x_t + f(t, x_t, u_t)) \right)
\end{aligned}$$

This is called the *Bellman Equation*.

Computes u as a function of x, t for all intermediate t and all x.

Discrete time control

The algorithm to compute the optimal control $u^*_{0:T-1}$, the optimal trajectory $x^*_{1:T}$ and the optimal cost is given by

1. Initialization: $J(T, x) = \phi(x)$
2. Backwards: For $t = T-1, \ldots, 0$ and for all x compute

$$\begin{aligned} u^*_t(x) &= \arg\min_u \{R(t, x, u) + J(t+1, x + f(t, x, u))\} \\ J(t, x) &= R(t, x, u^*_t) + J(t+1, x + f(t, x, u^*_t)) \end{aligned}$$

3. Forwards: For $t = 0, \ldots, T-1$ compute

$$x^*_{t+1} = x^*_t + f(t, x^*_t, u^*_t(x^*_t))$$

NB: the backward computation requires $u^*_t(x)$ for all x.

Stochastic case

$$x_{t+1} = x_t + f(t, x_t, u_t, w_t) \qquad t = 0, \ldots, T-1$$

At time t, w_t is a random value drawn from a probability distribution $p(w)$.

For instance,

$$\begin{aligned} x_{t+1} &= x_t + w_t, \qquad x_0 = 0 \\ w_t &= \pm 1, \qquad p(w_t = 1) = p(w_t = -1) = 1/2 \\ x_t &= \sum_{s=0}^{t-1} w_s \end{aligned}$$

Thus, x_t random variable and so is the cost

$$C(x_0) = \phi(x_T) + \sum_{t=0}^{T-1} R(t, x_t, u_t, \xi_t)$$

Stochastic case

$$
\begin{aligned}
C(x_0) &= \left\langle \phi(x_T) + \sum_{t=0}^{T-1} R(t, x_t, u_t, \xi_t) \right\rangle \\
&= \sum_{w_{0:T-1}} \sum_{\xi_{0:T-1}} p(w_{0:T-1}) p(\xi_{0:T-1}) \left(\phi(x_T) + \sum_{t=0}^{T-1} R(t, x_t, u_t, \xi_t) \right)
\end{aligned}
$$

with ξ_t, x_t, w_t random. Closed loop control: find *functions* $u_t(x_t)$ that minimizes the remaining expected cost when in state x at time t. $\pi = \{u_0(\cdot), \ldots, u_{T-1}(\cdot)\}$ is called a policy.

$$
\begin{aligned}
x_{t+1} &= x_t + f(t, x_t, u_t(x_t), w_t) \\
C_\pi(x_0) &= \left\langle \phi(x_T) + \sum_{t=0}^{T-1} R(t, x_t, u_t(x_t), \xi_t) \right\rangle
\end{aligned}
$$

$\pi^* = \text{argmin}_\pi C_\pi(x_0)$ is optimal policy.

Stochastic Bellman Equation

$$
\begin{aligned}
J(t, x_t) &= \min_{u_t} \langle R(t, x_t, u_t, \xi_t) + J(t+1, x_t + f(t, x_t, u_t, w_t)) \rangle \\
J(T, x) &= \phi(x)
\end{aligned}
$$

u_t is optimized for each x_t separately. $\pi = \{u_0, \ldots, u_{T-1}\}$ is optimal a policy.

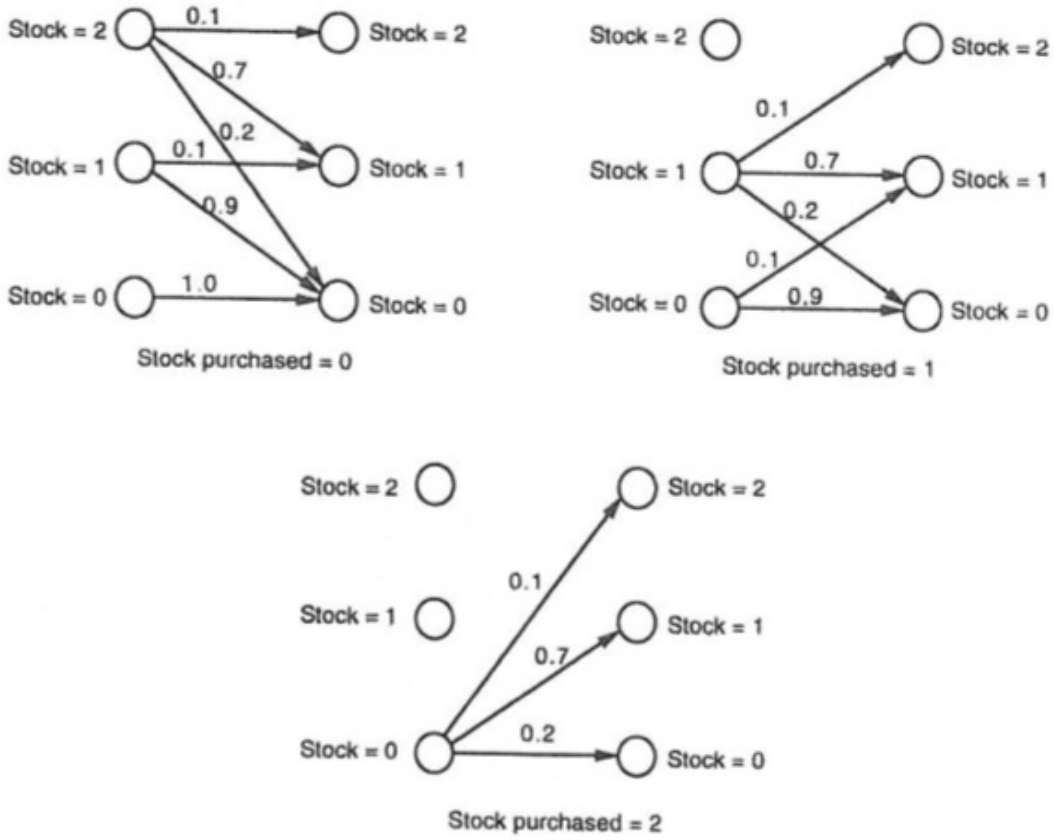
Inventory problem

- $x_t = 0, 1, 2$ stock available at the beginning of period t.
- u_t stock ordered at the beginning of period t. Maximum storage is 2: $u_t \leq 2 - x_t$.
- $w_t = 0, 1, 2$ demand during period t with $p(w = 0, 1, 2) = (0.1, 0.7, 0.2)$; excess demand is lost.
- u_t is the cost of purchasing u_t units. $(x_t + u_t - w_t)^2$ is cost of stock at end of period t.

$$
\begin{aligned}
x_{t+1} &= \max(0, x_t + u_t - w_t) \\
C(x_0, u_{0:T-1}) &= \left\langle \sum_{t=0}^{t=2} u_t + (x_t + u_t - w_t)^2 \right\rangle
\end{aligned}
$$

Planning horizon $T = 3$.

Inventory problem



Apply Bellman Equation

$$
\begin{aligned}
J_t(x_t) &= \min_{u_t} \langle R(x_t, u_t, w_t) + J_{t+1}(f(x_t, u_t, w_t)) \rangle \\
R(x, u, w) &= u + (x + u - w)^2 \\
f(x, u, w) &= \max(0, x + u - w)
\end{aligned}
$$

Start with $J_3(x_3) = 0, \forall x_3$.

Dynamic programming in action

Assume we are at stage $t = 2$ and the stock is x_2. The cost-to-go is what we order u_2 and how much we have left at the end of period $t = 2$.

$$\begin{aligned} J_2(x_2) &= \min_{0 \le u_2 \le 2 - x_2} u_2 + \left\langle (x_2 + u_2 - w_2)^2 \right\rangle \\ &= \min_{0 \le u_2 \le 2 - x_2} \Big(u_2 + 0.1 * (x_2 + u_2)^2 + 0.7 * (x_2 + u_2 - 1)^2 \\ &+ 0.2 * (x_2 + u_2 - 2)^2 \Big) \\ J_2(0) &= \min_{0 \le u_2 \le 2} \left(u_2 + 0.1 * u_2^2 + 0.7 * (u_2 - 1)^2 + 0.2 * (u_2 - 2)^2 \right) \end{aligned}$$

$$\begin{aligned} u_2 = 0 \quad &: \quad rhs = 0 + 0.7 * 1 + 0.2 * 4 = 1.5 \\ u_2 = 1 \quad &: \quad rhs = 1 + 0.1 * 1 + 0.2 * 1 = 1.3 \\ u_2 = 2 \quad &: \quad rhs = 2 + 0.1 * 4 + 0.7 * 1 = 3.1 \end{aligned}$$

Thus, $u_2(x_2 = 0) = 1$ and $J_2(x_2 = 0) = 1.3$

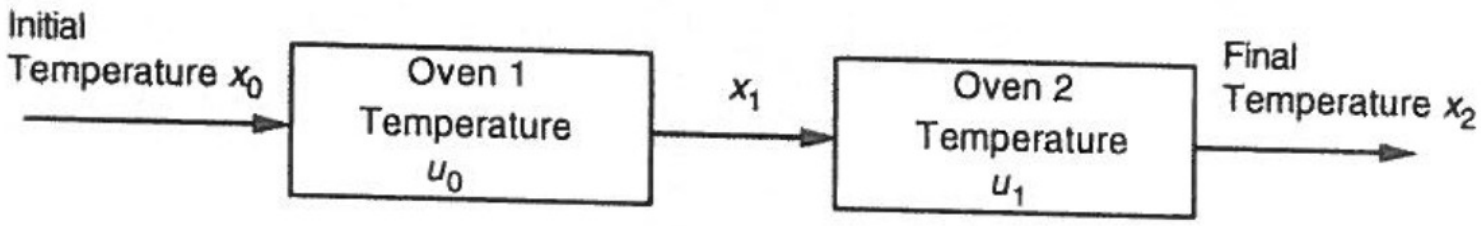
Inventory problem

The computation can be repeated for $x_2 = 1$ and $x_2 = 2$, completing stage 2 and subsequently for stage 1 and stage 0.

Stock	Stage 0 Cost-to-go	Stage 0 Optimal stock to purchase	Stage 1 Cost-to-go	Stage 1 Optimal stock to purchase	Stage 2 Cost-to-go	Stage 2 Optimal stock to purchase
0	3.67	1	2.5	1	1.3	1
1	2.67	0	1.2	0	0.3	0
2	2.608	0	1.68	0	1.1	0

Exercise: Two ovens

A certain material is passed through a sequence of two ovens. Aim is to reach pre-specified final product temperature x^* with minimal oven energy.



$x_{0,1,2}$ are the product temperatures initially, after pasing through oven 1 and after passing through oven 2. $u_{0,1}$ are the oven temperatures. The dynamics is

$$\begin{aligned} x_{t+1} &= (1-a)x_t + au_t \qquad t = 0, 1 \\ C &= r(x_2 - x^*)^2 + u_0^2 + u_1^2 \end{aligned}$$

- Find the optimal control solution u_0, u_1.
- Show that adding mean zero noise to the dynamics ($x_{t+1} = (1-a)x_t + au_t + w_t$ with $\langle w_t \rangle = 0$), does not change the optimal control solution.

Example: Two ovens

End cost-to-go is $J(2, x_2) = r(x_2 - x^*)^2$.

$$
\begin{aligned}
J(1, x_1) &= \min_{u_1} \left(u_1^2 + J(2, x_2) \right) = \min_{u_1} \left(u_1^2 + r((1-a)x_1 + au_1 - x^*)^2 \right) \\
u_1 &= \mu_1(x_1) = \frac{ra(x^* - (1-a)x_1)}{1 + ra^2} \\
J(1, x_1) &= \frac{r((1-a)x_1 - x^*)^2}{1 + ra^2} \\
J(0, x_0) &= \min_{u_0} \left(u_0^2 + J(1, x_1) \right) = \min_{u_0} \left(u_0^2 + \frac{r((1-a)x_1 - x^*)^2}{1 + ra^2} \right) \\
&= \min_{u_0} \left(u_0^2 + \frac{r((1-a)((1-a)x_0 + au_0) - x^*)^2}{1 + ra^2} \right) \\
u_0 &= \mu_0(x_0) = \frac{r(1-a)a(x^* - (1-a)^2 x_0)}{1 + ra^2(1 + (1-a)^2)} \\
J(0, x_0) &= \frac{r((1-a)^2 x_0 - x^*)^2}{1 + ra^2(1 + (1-a)^2)}
\end{aligned}
$$

Comments

- **Linear Quadratic Control:** Solution can be obtained in closed form because problem is linear quadratic.
- **Certainty equivalence:** Optimal control solution is unaffected by noise:

$$\begin{aligned} x_{t+1} &= (1-a)x_t + au_t + w_t \qquad t = 0, 1 \\ C &= r(x_2 - x^*)^2 + u_0^2 + u_1^2 \end{aligned}$$

with $\langle w_t \rangle = 0$.Then

$$\begin{aligned} J(1, x_1) &= \min_{u_1} \left(u_1^2 + \left\langle r((1-a)x_1 + au_1 + w_1 - x^*)^2 \right\rangle \right) \\ &= \min_{u_1} \left(u_1^2 + r((1-a)x_1 + au_1 - x^*)^2 + r \langle w_1 \rangle^2 \right) \end{aligned}$$

Continuous limit

Replace $t+1$ by $t+dt$ with $dt \to 0$.

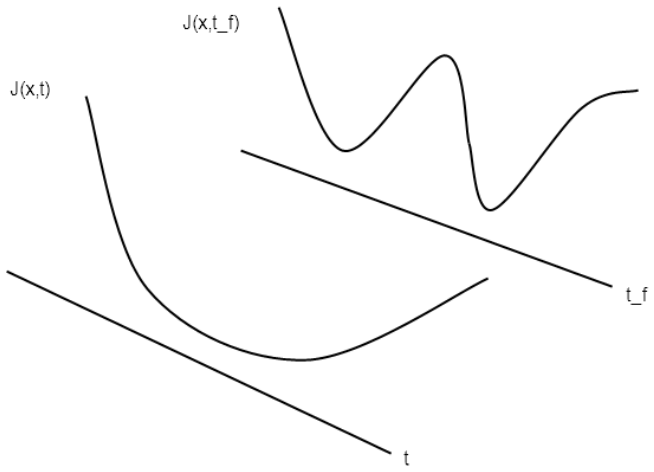
$$\begin{aligned} x_{t+dt} &= x_t + f(x_t, u_t, t)dt \\ C(x_0, u_{0\to T}) &= \phi(x_T) + \int_0' d\tau R(\tau, x(\tau), u(\tau)) \end{aligned}$$

Assume $J(x,t)$ is smooth.

$$\begin{aligned} J(t,x) &= \min_u (R(t,x,u)dt + J(t+dt, x+f(x,u,t)dt)) \\ &\approx \min_u (R(t,x,u)dt + J(t,x) + \partial_t J(t,x)dt + \partial_x J(t,x) f(x,u,t)dt) \\ -\partial_t J(t,x) &= \min_u (R(t,x,u) + f(x,u,t)\partial_x J(x,t)) \end{aligned}$$

with boundary condition $J(x,T) = \phi(x)$.

Continuous limit



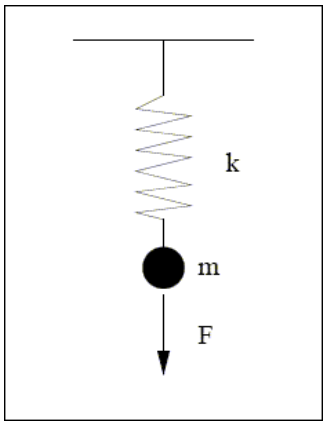
$$-\partial_t J(t,x) \quad = \quad \min_u \left(R(t,x,u) + f(x,u,t)\partial_x J(x,t)\right)$$

with boundary condition $J(x,T) = \phi(x)$.

This is called the *Hamilton-Jacobi-Bellman Equation*.

Computes the *anticipated potential* $J(t,x)$ from the future potential $\phi(x)$.

Example: Mass on a spring



The spring force $F_z = -z$ towards the rest position and control force $F_u = u$.

Newton's Law

$$F = -z + u = m\ddot{z}$$

with $m = 1$.

Control problem: Given initial position and velocity $z(0) = \dot{z}(0) = 0$ at time $t = 0$, find the control path $-1 < u(0 \to T) < 1$ such that $z(T)$ is maximal.

Example: Mass on a spring

Introduce $x_1 = z, x_2 = \dot{z}$, then

$$\begin{aligned} \dot{x_1} &= x_2 \\ \dot{x_2} &= -x_1 + u \end{aligned}$$

The end cost is $\phi(x) = -x_1$; path cost $R(x, u, t) = 0$.

The HJB takes the form:

$$\begin{aligned} -\partial_t J &= \min_u \left(x_2 \frac{\partial J}{\partial x_1} - x_1 \frac{\partial J}{\partial x_2} + \frac{\partial J}{\partial x_2} u \right) \\ &= x_2 \frac{\partial J}{\partial x_1} - x_1 \frac{\partial J}{\partial x_2} - \left| \frac{\partial J}{\partial x_2} \right|, \quad u = -\text{sign}\left(\frac{\partial J}{\partial x_2} \right) \end{aligned}$$

Example: Mass on a spring

We try $J(t,x) = \psi_1(t)x_1 + \psi_2(t)x_2 + \alpha(t)$. The HJBE reduces to the ordinary differential equations

$$\begin{aligned} \dot{\psi}_1 &= \psi_2 \\ \dot{\psi}_2 &= -\psi_1 \\ \dot{\alpha} &= -|\psi_2| \end{aligned}$$

These equations must be solved for all t, with final boundary conditions $\psi_1(T) = -1, \psi_2(T) = 0$ and $\alpha(T) = 0$.

Note, that the optimal control only requires $\partial_x J(x,t)$, which in this case is $\psi(t)$ and thus we do not need to solve α. The solution for ψ is

$$\begin{aligned} \psi_1(t) &= -\cos(t-T) \\ \psi_2(t) &= \sin(t-T) \end{aligned}$$

Example: Mass on a spring

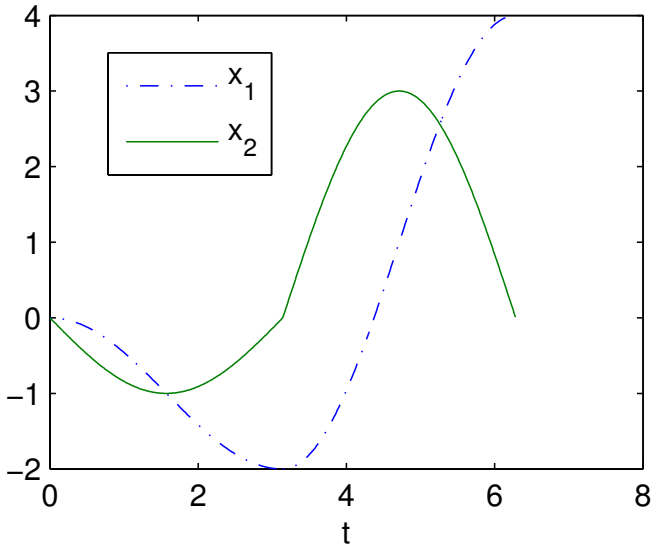
The optimal control is

$$u(x,t) = -\text{sign}(\psi_2(t)) = -\text{sign}(\sin(t-T))$$

As an example consider $T = 2\pi$. Then, the optimal control is

$$u = -1, \quad 0 < t < \pi$$

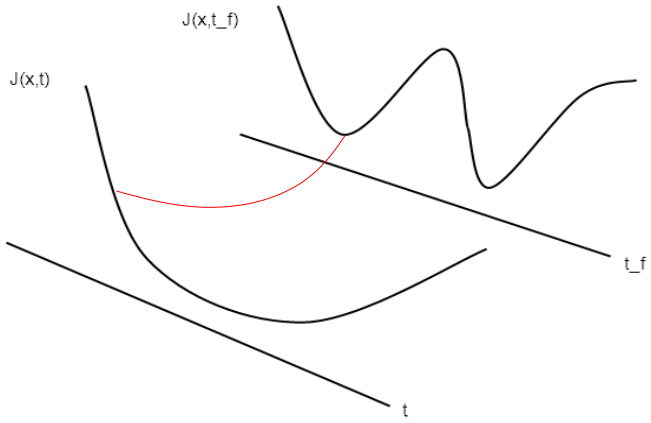
$$u = 1, \quad \pi < t < 2\pi$$



Pontryagin minimum principle

The HJB equation is a PDE with boundary condition at future time. The PDE is solved using discretization of space and time.

The solution is an optimal cost-to-go for all x and t. From this we compute the optimal trajectory and optimal control.



An alternative approach is a variational approach that directly finds the optimal trajectory and optimal control.

Pontryagin minimum principle

We can write the optimal control problem as a constrained optimization problem with independent variables $u(0 \to T)$ and $x(0 \to T)$

$$\min_{u(0\to T),x(0\to T)} \phi(x(T)) + \int_0^T dt R(x(t), u(t), t)$$

subject to the constraint

$$\dot{x} = f(x, u, t)$$

and boundary condition $x(0) = x_0$.

Introduce the Lagrange multiplier function $\lambda(t)$:

$$\begin{aligned} C &= \phi(x(T)) + \int_0^T dt \left[R(t, x(t), u(t)) - \lambda(t)(f(t, x(t), u(t)) - \dot{x}(t))\right] \\ &= \phi(x(T)) + \int_0^T dt[-H(t, x(t), u(t), \lambda(t)) + \lambda(t)\dot{x}(t))] \\ -H(t, x, u, \lambda) &= R(t, x, u) - \lambda f(t, x, u) \end{aligned}$$

Derivation PMP

The solution is found by extremizing C. This gives a necessary but not sufficient condition for a solution.

If we vary the action wrt to the trajectory x, the control u and the Lagrange multiplier λ, we get:

$$\begin{aligned}
\delta C &= \phi_x(x(T))\delta x(T) \\
&+ \int_0^T dt[-H_x\delta x(t) - H_u\delta u(t) + (-H_\lambda + \dot{x}(t))\delta\lambda(t) + \lambda(t)\delta\dot{x}(t)] \\
&= (\phi_x(x(T)) + \lambda(T))\,\delta x(T) \\
&+ \int_0^T dt\left[(-H_x - \dot{\lambda}(t))\delta x(t) - H_u\delta u(t) + (-H_\lambda + \dot{x}(t))\delta\lambda(t)\right]
\end{aligned}$$

For instance, $H_x = \frac{\partial H(t,x(t),u(t),\lambda(t))}{\partial x(t)}$.

We can solve $H_u(t, x, u, \lambda) = 0$ for u and denote the solution as

$$u^*(t, x, \lambda)$$

Assumes H convex in u.

The remaining equations are

$$\begin{aligned} \dot{x} &= H_\lambda(t, x, u^*(t, x, \lambda), \lambda) \\ \dot{\lambda} &= -H_x(t, x, u^*(t, x, \lambda), \lambda) \end{aligned}$$

with boundary conditions

$$x(0) = x_0 \qquad \lambda(T) = -\phi_x(x(T))$$

Mixed boundary value problem.

Again mass on a spring

Problem

$$\begin{aligned} \dot{x_1} &= x_2, & \dot{x_2} &= -x_1 + u \\ R(x,u,t) &= 0 & \phi(x) &= -x_1 \end{aligned}$$

Hamiltonian

$$\begin{aligned} H(t,x,u,\lambda) &= -R(t,x,u) + \lambda' f(t,x,u) = \lambda_1 x_2 + \lambda_2(-x_1 + u) \\ H^*(t,x,\lambda) &= \lambda_1 x_2 - \lambda_2 x_1 - |\lambda_2| \qquad u^* = -\text{sign}(\lambda_2) \end{aligned}$$

The Hamilton equations

$$\begin{aligned} \dot{x} = \frac{\partial H^*}{\partial \lambda} \quad &\Rightarrow \quad \dot{x_1} = x_2, \quad \dot{x_2} = -x_1 - \text{sign}(\lambda_2) \\ \dot{\lambda} = -\frac{\partial H^*}{\partial x} \quad &\Rightarrow \quad \dot{\lambda_1} = \lambda_2, \quad \dot{\lambda_2} = -\lambda_1 \end{aligned}$$

with $x(t=0) = x_0$ and $\lambda(t=T) = (1,0)$.

Example

Consider the control problem:

$$\begin{aligned} dx &= udt \\ C &= \frac{\alpha}{2}x(T)^2 + \int_{t_0}^{'} dt\frac{1}{2}u(t)^2 \end{aligned}$$

with initial condition $x(t_0)$.

Solve the control problem using the PMP formalism.

Solution

The PMP recipe is

1. Construct the Hamiltonian

$$H(t,x,u,\lambda) = -R(t,x,u) + \lambda f(t,u,x) = -\frac{1}{2}u^2 + \lambda u$$

2. Construct the optimized Hamiltonian

$$H^*(t,x,\lambda) = H(t,x,u^*,\lambda) = \frac{1}{2}\lambda^2 \qquad u^* = \lambda$$

3. Solve the Hamilton equations of motion

$$\begin{aligned} \frac{dx}{dt} &= \frac{\partial H^*}{\partial \lambda} = \lambda \\ \frac{d\lambda}{dt} &= -\frac{\partial H^*}{\partial x} = 0 \end{aligned}$$

with boundary conditions $x(t_0)$ and $\lambda(t=T) = -\alpha x(T)$[6]. The solution for λ is constant $\lambda(t) = \lambda = -\alpha x(T)$. The solution for $x(t)$ is

$$x(t) = x(t_0) + \lambda(t - t_0)$$

[6]Note, that $\phi(x) = \frac{\alpha}{2}x^2$ so that $\phi_x = \alpha x$.

Combining these two results, we get $\lambda = -\alpha x(T) = -\alpha(x(t_0) + \lambda(T - t_0))$, or

$$\lambda = \frac{-\alpha x(t_0)}{1 + \alpha(T - t_0)}$$

Since $u^* = \lambda$, this is the optimal control law.

Relation to classical mechanics

The equations look like classical mechanics

$$\begin{aligned} \dot{x} &= H_\lambda(t, x, u^*(t, x, \lambda), \lambda) \qquad x(0) = x_0 \\ \dot{\lambda} &= -H_x(t, x, u^*(t, x, \lambda), \lambda) \qquad \lambda(T) = -\phi_x(x(T)) \end{aligned}$$

In classical mechanics H is called the Hamiltonian. Consider the time evolution of H:

$$\begin{aligned} \dot{H} &= H_t + H_u \dot{u} + H_x \dot{x} + H_\lambda \dot{\lambda} = H_t \\ H(t, x, u, \lambda) &= -R(t, x, u) + \lambda f(t, u, x) \end{aligned}$$

So, for problems where R, f do not explicitly depend on time, H is a constant of the motion.

Example

Consider the control problem:

$$\begin{aligned} dx &= udt \\ C &= \int_{t_0}^{'} dt \frac{1}{2} u(t)^2 + V(x(t)) \end{aligned}$$

with initial condition $x(t_0)$.

1. $H(x, u, \lambda) = -\frac{1}{2}u^2 - V(x) + \lambda u$
2. $u^* = \lambda$, $H^*(x, \lambda) = \frac{1}{2}\lambda^2 - V(x)$
3.

$$\dot{x} = \frac{\partial H^*}{\partial \lambda} = \lambda \qquad \dot{\lambda} = -\frac{\partial H^*}{\partial \lambda} = \frac{\partial V(x)}{\partial x}$$

Control cost V play role of *minus* potential energy.
Control solution has constant *difference* of kinetic energy and state cost

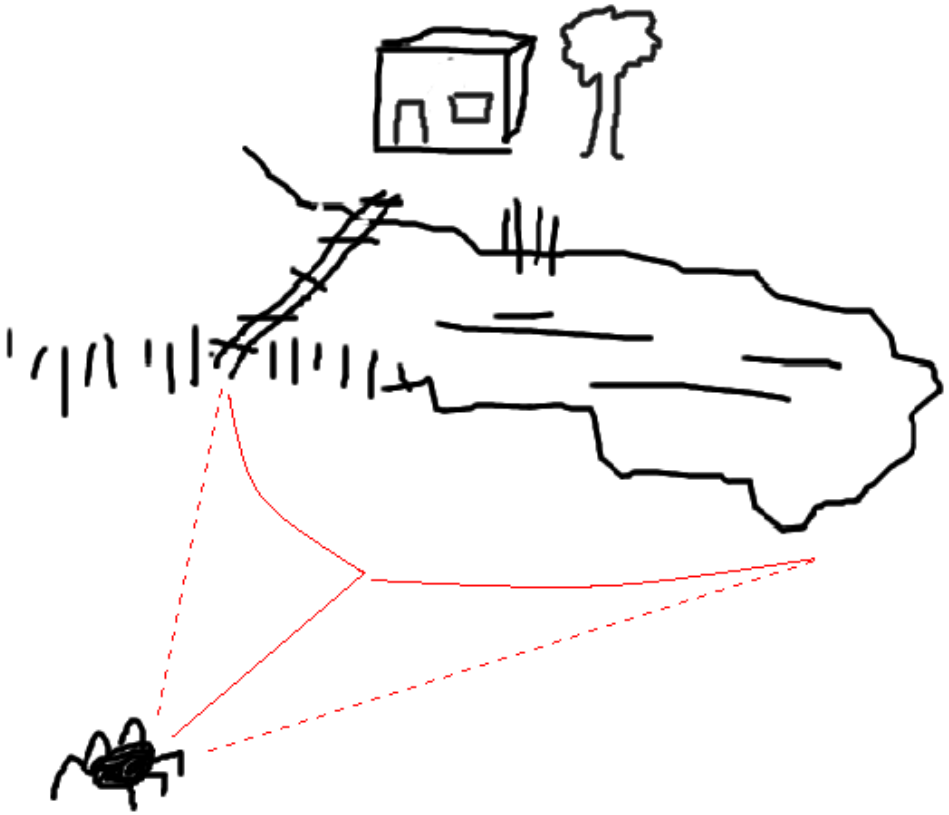
Comments

The solution of the HJB PDE is expensive.

The PMP method is computationally less complicated than the HJB method because it does not require discretisation of the state space.

HJB generalizes to the stochastic case, PMP does not (at least not easy).

Stochastic control



Stochastic differential equations

Consider the random walk on the line:

$$X_{t+1} = X_t + \xi_t \qquad \xi_t = \pm 1$$

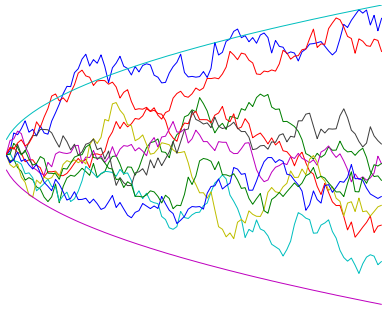
with $x_0 = 0$. We can compute

$$X_t = \sum_{i=1}^{'} \xi_i$$

Since x_t is a sum of random variables, x_t becomes Gaussian distributed with

$$\mathbb{E}x_t = \sum_{i=1}^{'} \mathbb{E}\xi_i = 0$$

$$\mathbb{V}x_t = \sum_{i,j=1}^{'} \mathbb{V}\xi_i = t$$



Note, that the fluctuations $\propto \sqrt{t}$.

Stochastic differential equations

In the continuous time limit we define

$$dX_t \quad = \quad X_{t+dt} - X_t = dW_t$$

with dW_t an infinitesimal mean zero Gaussian variable: $\mathbb{E}dW_t = 0, \mathbb{V}dW_t = \nu dt$.

Then with initial condition x_1 at t_1

$$X_t = x_1 + \int_{t_1}^{t} dW_s \qquad \mathbb{E}X_t = x_0 \qquad \mathbb{V}X_t = \nu t$$

is called a Wiener process or Brownian motion.

Since the increments are independent, X_t is Gaussian distributed

$$p(x_2, t_2|x_1, t_1) \quad = \quad \frac{1}{\sqrt{2\pi\nu(t_2 - t_1)}} \exp\left(-\frac{(x_2 - x_1)^2}{2\nu(t_2 - t_1)}\right)$$

Stochastic differential equations

Consider the stochastic differential equation

$$dX_t = f(X_t, t)dt + dW_t$$

W_t is a Wiener process.

In this case $\rho(x_2, t_2|x_1, t_1)$ may be very complex and is generally not known.

Define $\rho(x, t) = p(x, t|x_0, 0)$. Then (Fokker-Planck forward equation)

$$\partial_t \rho(x,t) \quad = \quad -\nabla(f(x,t)\rho(x,t)) + \frac{1}{2}\nu\nabla^2\rho(x,t), \qquad \rho(x,0) = \delta(x - x_0)$$

Define $\psi(x, t) = p(z, T|x, t)$. Then (Kolmogorov backward equation)

$$-\partial_t \psi(x,t) \quad = \quad f(x,t)\nabla\psi(x,t) + \frac{1}{2}\nu\nabla^2\psi(x,t) \qquad \psi(x,T) = \delta(z - x)$$

Example: Brownian motion

$$
\begin{aligned}
X_t &= x_0 + \int_0^{'} dW_s \\
\rho(x,t) &= p(x,t|x_0,0) = \frac{1}{\sqrt{2\pi\nu t}} \exp\left(-\frac{(x-x_0)^2}{2\nu t}\right) \\
\psi(x,t) &= p(z,T|x,t) = \frac{1}{\sqrt{2\pi\nu(T-t)}} \exp\left(-\frac{(x-z)^2}{2\nu(T-t)}\right)
\end{aligned}
$$

Stochastic optimal control

Consider a stochastic dynamical system

$$dX_t = f(t, X_t, u)dt + dW_t$$

W_t is a Wiener process with $\mathbb{E}dW_t^2 = \nu(t, x, u)dt$. [7]

The cost becomes an expectation:

$$C(t, x, u) \quad = \quad \mathbb{E}\left(\phi(X_T) + \int_t^T d\tau R(t, X_t, u(X_t, t))\right)$$

over all stochastic trajectories starting at x with control function $u(\cdot, t)$.

Optimize with respect to the set of functions $u(\cdot, t)$.

[7] Our notation is for one dimensional X, but the theory generalizes trivially to higher dimension.

Stochastic optimal control

We obtain the Bellman recursion

$$\begin{aligned} J(t, x_t) &= \min_{u_t} R(t, x_t, u_t)dt + \mathbb{E}J(t + dt, X_{t+dt}) \\ J(t + dt, x_t + dX_t) &= J(t, x_t) + dt\partial_t J(t, x_t) + dX_t\partial_x J(t, x_t) + \frac{1}{2}dX_t^2\partial_x^2 J(t, x_t) \\ \mathbb{E}J(t + dt, x_t + dX_t) &= J(t, x_t) + dt\partial_t J(t, x_t) + fdt\partial_x J(t, x_t) + \frac{1}{2}\nu dt\partial_x^2 J(t, x_t) \end{aligned}$$

because $\mathbb{E}dX_t = fdt$ and $\mathbb{E}dX_t^2 = \nu dt + (fdt)^2 = \nu dt + O(dt^2)$.

Thus (Stochastic Hamilton-Jacobi-Bellman equation)

$$-\partial_t J(t, x) = \min_u \left(R(t, x, u) + f(x, u, t)\partial_x J(x, t) + \frac{1}{2}\nu(t, x, u)\partial_x^2 J(x, t) \right)$$

with boundary condition $J(x, T) = \phi(x)$.

Linear Quadratic control

The dynamics is linear

$$dX_t = [A(t)X_t + B(t)u_t + b(t)]dt + \sum_{j=1}^{m}(C_j(t)X_t + D_j(t)u_t + \sigma_j(t))dW_j, \quad \left\langle dW_j dW_{j'} \right\rangle = \delta_{jj'}dt$$

The cost function is quadratic

$$\begin{aligned} \phi(x) &= \frac{1}{2}x'Gx \\ R(x,u,t) &= \frac{1}{2}x'Q(t)x + u'S(t)x + \frac{1}{2}u'R(t)u \end{aligned}$$

In this case the optimal cost-to-go is quadratic in x:

$$\begin{aligned} J(t,x) &= \frac{1}{2}x'P(t)x + \alpha'(t)x + \beta(t) \\ u_t &= -\Psi(t)x_t - \psi(t) \end{aligned}$$

Substitution in the HJB equation yields ODEs for P, α, β:

$$
\begin{aligned}
-\dot{P} &= PA + A'P + \sum_{j=1}^{m} C_j' P C_j + Q - \hat{S}'\hat{R}^{-1}\hat{S} \\
-\dot{\alpha} &= [A - B\hat{R}^{-1}\hat{S}]'\alpha + \sum_{j=1}^{m} [C_j - D_j\hat{R}^{-1}\hat{S}]' P\sigma_j + Pb \\
\dot{\beta} &= \frac{1}{2}\left|\sqrt{\hat{R}}\psi\right|^2 - \alpha' b - \frac{1}{2}\sum_{j=1}^{m} \sigma_j' P \sigma_j \\
\hat{R} &= R + \sum_{j=1}^{m} D_j' P D_j \\
\hat{S} &= B'P + S + \sum_{j=1}^{m} D_j' P C_j \\
\Psi &= \hat{R}^{-1}\hat{S} \\
\psi &= \hat{R}^{-1}(B'\alpha + \sum_{j=1}^{m} D_j' P \sigma_j)
\end{aligned}
$$

with $P(t_f) = G$ and $\alpha(t_f) = \beta(t_f) = 0$.

Example

Find the optimal control for the dynamics

$$\begin{aligned} dX_t &= udt + dW_t, \qquad \left\langle dW_t^2 \right\rangle = \nu dt \\ C &= \left\langle \frac{1}{2} G x(T)^2 + \int_0^{'} dt \frac{1}{2} u(x,t)^2 \right\rangle \end{aligned}$$

with end cost $\phi(x) = \frac{1}{2} G x^2$ and path cost $R(x,u) = \frac{1}{2} u^2$.
$(A = 0, B = 1, b = 0, C = D = 0, \sigma_j = \sqrt{\nu}, m = 1, \hat{R} = 1, \hat{S} = P, \Psi = P, \psi = \alpha)$

The Ricatti equations reduce to

$$\begin{aligned} \dot{P} &= P^2 \qquad P(T) = G \\ \dot{\alpha} &= P\alpha \qquad \alpha(T) = 0 \\ \dot{\beta} &= \frac{1}{2}\alpha^2 - \frac{1}{2}\nu P \end{aligned}$$

The solution is $\alpha(t) = 0$ and

$$P(t) = \frac{1}{c-t} \qquad \frac{1}{c-T} = G$$

and β not relevant.

$$u(x,t) \quad = \quad -P(t)x - \alpha(t) = -\frac{Gx}{1+G(T-t)}$$

Compare with deterministic case considered earlier, is identical due to certainty equivalence.

When $G \to \infty$ we obtain the Brownian bridge The control law and dynamics becomes

$$\begin{aligned} dx &= udt + d\xi \\ u &= \frac{-x(t_0)}{T - t_0} \end{aligned}$$

$x(T) \to 0$ w.p. 1.

Example

Find the optimal control for the dynamics

$$dX_t = udt + dW_t, \qquad \left\langle dW_t^2 \right\rangle = \nu dt$$

with end cost $\phi(x) = 0$ and path cost $R(x,u) = \frac{1}{2}(Qx^2 + Ru^2)$.

The Ricatti equations reduce to

$$\begin{aligned} -\dot{P} &= Q - R^{-1}P^2 \\ -\dot{\alpha} &= -R^{-1}P\alpha = 0 \\ \dot{\beta} &= -\frac{1}{2}\nu P \end{aligned}$$

with $P(T) = \alpha(T) = \beta(T) = 0$ and

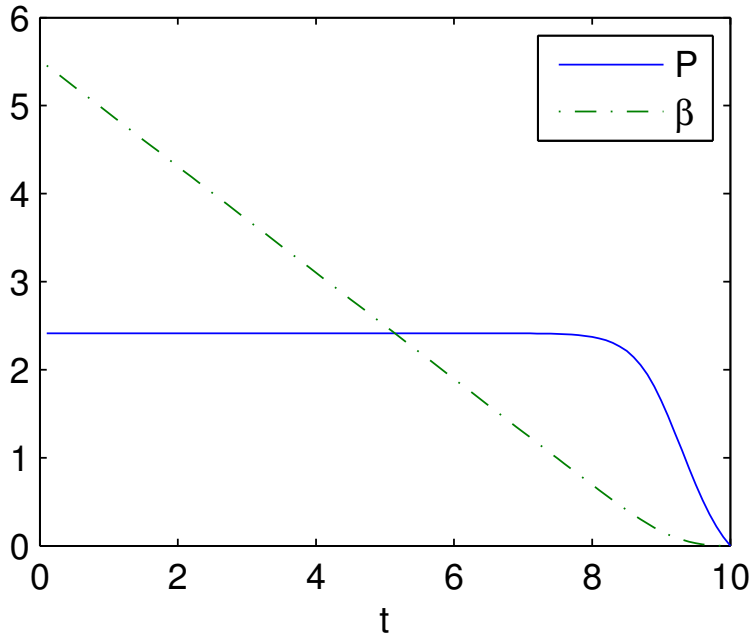
$$u(x,t) = -R^{-1}P(t)x$$

The solution is

$$
\begin{aligned}
P(t) &= \sqrt{RQ}\tanh\left(\sqrt{\frac{Q}{R}}(T-t)\right) \\
\alpha(t) &= 0 \\
\beta(t) &= \frac{1}{2}\nu R \log\cosh\left(\sqrt{\frac{Q}{R}}(T-t)\right) \\
\Psi(t) &= R^{-1}P(t) \qquad \psi(t) = 0
\end{aligned}
$$

The control is given by Eq. **??**:

$$u(x,t) = -R^{-1}P(t)x \tag{2}$$


6
5
4
3
2
1
0
0
2
4
6
8
10
t
P
β

Comments

Note, that in the last example the optimal control is independent of ν, i.e. optimal stochastic control equals optimal deterministic control.

In general:

- If $C_j = D_j = 0$ (only 'additive noise') $\dot{P}, \dot{\alpha}$ independent of noise σ, $\dot{\beta}$ depends on σ, but control independent of β. Thus control independent of σ (certainty equivalence)
- If $C_j \neq 0$ or $D_j \neq 0$, control depends on C_j, D_j, σ_j (no certainty equivalence)

Example: Portfolio selection

[8] Consider a market with p stocks and one bond. The bond price process is subject ot the following deterministic ordinary differential equation:

$$dP_0(t) = r(t)P_0(t)dt, \quad P_0(0) = p_0 > 0 \tag{3}$$

The other assets have price processes $P_i(t), i = 1, \ldots, p$ satisfying stochastic differential equations

$$dP_i(t) = P_i(t)\left(b_i(t)dt + \sum_{j=1}^{m} \sigma_{ij}(t)d\xi_j(t)\right), \quad P_i(0) = p_i > 0 \tag{4}$$

Consider an investor whose total wealth at time t is denoted by $x(t)$

$$x(t) = \sum_{i=0}^{p} N_i(t)P_i(t) \tag{5}$$

with N_i the number of stocks/bond of type i. For given $N_i(t)$,

$$dx(t) = \sum_{i=0}^{p} N_i(t)dP_i(t) = \left(r(t)x(t) + \sum_{i=1}^{p}(b_i(t) - r(t))u_i(t)\right)dt + \sum_{i=1}^{p}\sum_{j=1}^{m} \sigma_{ij}(t)u_i(t)d\xi_j(t) \tag{6}$$

with $u_i(t) = N_i(t)P_i(t), i = 1, \ldots, p$ the rescaled control variable.

[8] This section is from [Yong and Zhou, 1999] section 6.8 (pg. 335).

The objective of the investor is to maximize the mean terminal wealth $\left\langle x(t_f) \right\rangle$ and minimize at the same time the variance

$$\Sigma^2 = \left\langle x(t_f)^2 \right\rangle - \left\langle x(t_f) \right\rangle^2$$

This is a multi-objective optimization problem with an efficient frontier of optimal solutions: for each given mean there is a minimial variance.

These pairs can be found by minimizing the single objective criterion

$$\mu\Sigma^2 - \left\langle x(t_f) \right\rangle \tag{7}$$

for different values of the weighting factor μ.

This objective, however, is not an expectation value of some stochastic quantity due to the $\langle\cdot\rangle^2$ term. Consider a slightly different problem, minimizing the objective

$$\left\langle \mu x(t_f)^2 - \lambda x(t_f) \right\rangle \tag{8}$$

which is of the standard stochastic optimization form. One can show that one can construct a solution of Problem 7 by solving problem 8 for suitable $\lambda(\mu)$. [9]

Our goal is thus to minimize eq. 8 subject to the stochastic dynamics eq. 6.

[9] and finding λ from

$$\lambda = 1 + 2\mu \left\langle x(t_f) \right\rangle (\lambda, \mu)$$

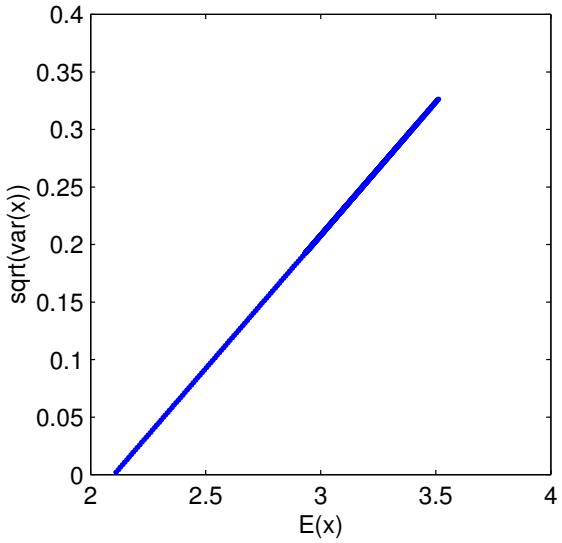
([Yong and Zhou, 1999] Theorem 8.2 pg. 338)

This is an LQ problem. The solution is computed from the Ricatti equations

$$u_i(x,t) = \psi_i(t)x + \phi_i(t)$$

As an example we consider the simplest possible case: $p = m = 1$ and r, b, σ independent of time.

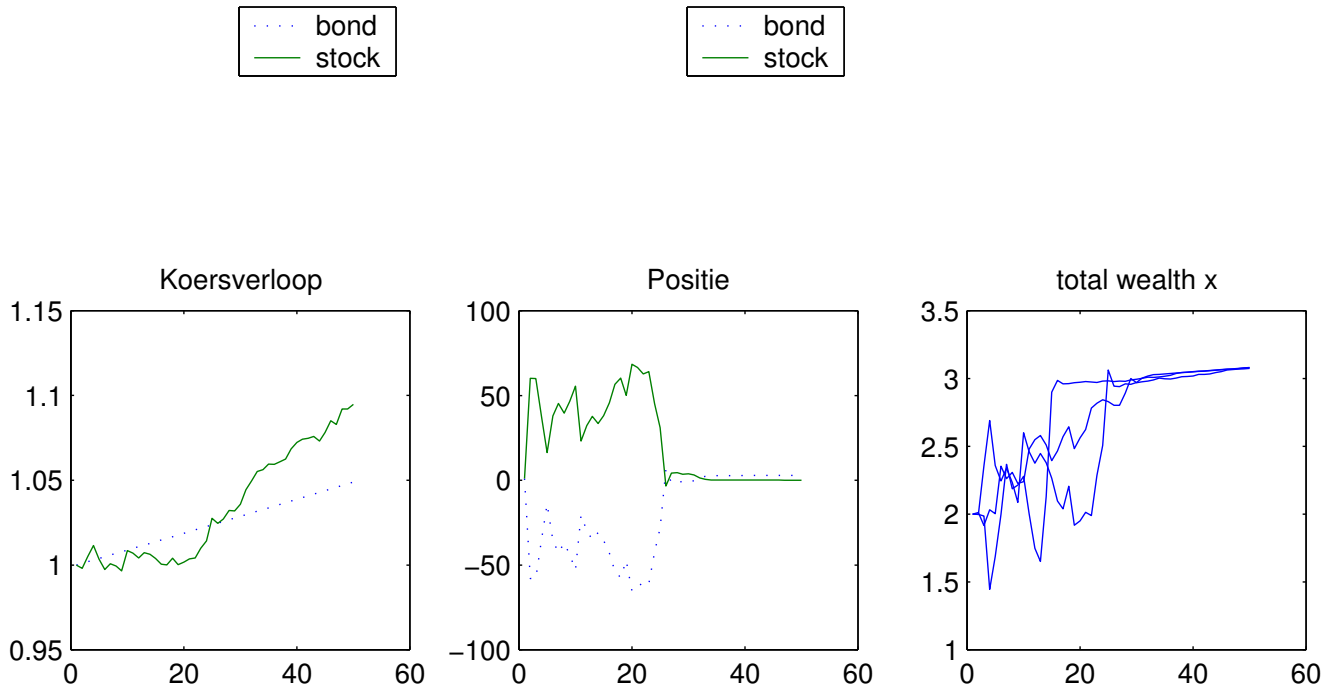
Efficient boundary



Parameter values are: $p = m = 1$. Trading period is one year weekly. annual bond rate 5 % ($r = 0.0009758$), annual expected stock rate is 10 % ($b = 0.0019$), volatility $\sigma = 2b$. $x_0 = 2$. Shows var x versus $\langle x \rangle$ scatter plot for various values of μ. Small μ corresponds to risky investments with high expected return and large fluctuation. $\mu \to \infty$ corresponds to riskless investment in bond only and a return of 5 %.

$\mu = 10$ corresponds to $\langle x \rangle = 3$ and $\sqrt{var} = 0.2$.

Making money



Simulation of optimal control with $\mu = 10$, The optimal strategy is to borrow many stocks and sell them as soon as the objective is achieved.

Indeed, $\langle x \rangle = 3$ as expected. The strategy to get at this 50 % increase in wealth is to buy many stocks and hope they will give the expected wealth increase. As soon as this occurs, all stocks are sold and the money is put in the bank. [10]

[10] When? Say borrow 50, find t such that

$$(2 + 50)(1 + bt) - 50(1 + rt) = 3 \qquad 50(b - r)t \approx 1$$

Path integral control

The n-dimensional path integral control problem is defined as

$$\begin{aligned} dX_t &= f(X_t,t)dt + g(x,t)(u(X_t,t)dt + dW_t) \\ C(t,x,u) &= \mathbb{E}\left(\phi(X_T) + \int_t^T dsV(X_s,s) + \frac{1}{2}u'(X_s,s)Ru(X_s,s)\right) \end{aligned}$$

with $\mathbb{E}dW_t dW_t' = \nu dt$. g is $n \times m$ matrix, ν is $m \times m$ matrix and u, dW_t are m dimensional.

The cost is an expectation over all stochastic trajectories starting at x with control function $u(x,t)$.

The stochastic HJB equation becomes

$$-\partial_t J = \min_u \left(\frac{1}{2}u'Ru + V + (\nabla J)'(f + gu) + \frac{1}{2}\mathrm{Tr}\left(g\nu g'\nabla^2 J\right)\right)$$

which we need to solve with end boundary condition $J(x,t_f) = \phi(x)$ for all x.