1 Introduction to Dynamic Programming

Dynamics: $x_{k+1}=f_k(x_k,u_k,w_k)$ with $k=0,1,\dots,N-1$. $x_k\in\mathcal{S}_k$ state space, $u_k\in\mathcal{U}_k(x_k)$ control space and w_k is the disturbance.

1.1 Open Loop and Closed Loop Control

Open loop: controls $\bar{u}_k$ are fixed at time $k=0$, used in deterministic problems. *Closed loop*: controls u_k are state dependent, used in stochastic problems. The *expected closed loop cost* is $J_\pi(x) := \underset{w_k}{\mathrm{E}}\left[g_N(x_N)+\sum_{k=0}^{N-1}g_k(x_k,\mu_k(x_k),w_k)\right]$. Let Π denote the set of all admissible policies. The *optimal cost* is $J^*(x_0):=J_{\pi^*}(x_0)$ where π^* is called an *optimal policy* if $J_{\pi^*}(x)\le J_\pi(x)$, $\forall\pi\in\Pi,\forall x\in\mathcal{S}_0$.

1.1.1 Computation

Consider a system with N_x states, N_u control inputs and N stages.The number of strategies for each control method is given by

Open loop	Closed loop	Brute force
N_u^N	$N_u^{N_x(N-1)+1}$	$N_u^{N_x N}$

2 The Dynamic Programming Algorithm (DPA)

Initialization $J_N(x_N)=g_N(x_N)$, $\forall x_N\in\mathcal{S}_N$

Recursion The cost-to-go at state $x\in\mathcal{S}_k$ is

$$J_k(x):=\min_{u_k\in\mathcal{U}_k(x)}\underset{w_k}{\mathrm{E}}[g_k(x_k,u_k,w_k)+J_{k+1}(f_k(x_k,u_k,w_k))],\ \forall x\in\mathcal{S}_k$$

If $u^*=:\mu_k^*(x)$ minimizes this recursion equation for each k, the policy $\pi^*=\{\mu_0^*(\cdot),\dots,\mu_{N-1}^*(\cdot)\}$ is optimal.

2.1 Converting non-standard problems to the standard form

2.1.1 Time Lags

Suppose the dynamics are $x_{k+1}=f_k(x_k,x_{k-1},u_k,u_{k-1},w_k)$.

Solution Define new states $y_k:=x_{k-1}$, $s_k:=u_{k-1}$, $\tilde{x}_k:=(x_k,y_k,s_k)$. Now the new dynamics are

$$\tilde{x}_{k+1}:=\begin{bmatrix}x_{k+1}\\y_{k+1}\\s_{k+1}\end{bmatrix}=\begin{bmatrix}f_k(x_k,y_k,u_k,s_k,w_k)\\x_k\\u_k\end{bmatrix}=:\tilde{f}_k(\tilde{x}_k,u_k,w_k)$$

Remark This works for an arbitrary number of lags.

2.1.2 Correlated Disturbances

Suppose the disturbance dynamics are $w_k=C_ky_{k+1}$ and $y_{k+1}=A_ky_k+\xi_k$, where A_k,C_k are given and ξ_k are independent RVs.

Solution Let the augmented state be $\tilde{x}_k=(x_k,y_k)$. Now the new dynamics are

$$\tilde{x}_{k+1}:=\begin{bmatrix}x_{k+1}\\y_{k+1}\end{bmatrix}=\begin{bmatrix}f_k(x_k,u_k,C_k(A_ky_k+\xi_k))\\A_ky_k+\xi_k\end{bmatrix}=:\tilde{f}_k(\tilde{x}_k,u_k,\xi_k).$$

2.1.3 Forecasts

Suppose we receive a forecast that $y_k=i$. We can generate w_k from the given distribution $p_{w_k|y_k}(\cdot|i)$. Suppose that the forecast has its own given prior $y_{k+1}=\xi_k$, where ξ_k are independent RVs taking the value $i\in\{1,\dots,m\}$ with probability $p_{\xi_k}(i)$.

Solution Let $\tilde{x}_k:=(x_k,y_k)$ and $\tilde{w}_k:=(w_k,\xi_k)$ where we specify $p(\tilde{w}_k|\tilde{x}_k,u_k)=p(w_k|y_k)p(\xi_k)$ by using the chain rule and eliminating the variables on which $\tilde{w}_k$ doesn't depend. Now the new dynamics are

$$\tilde{x}_{k+1}:=\begin{bmatrix}x_{k+1}\\y_{k+1}\end{bmatrix}=\begin{bmatrix}f_k(x_k,u_k,w_k)\\\xi_k\end{bmatrix}=:\tilde{f}_k(\tilde{x}_k,u_k,\tilde{w}_k)$$

The associated DPA is now: $J_N(\tilde{x})=J_N(x,y)=g_N(x)$, $x\in\mathcal{S}_N, y\in\{1,\dots,m\}$ and repeat:

$$J_k(\tilde{x})=\min_{u\in\mathcal{U}_k(x)}\underset{w_k}{\mathrm{E}}\left[g_k(x,u,w_k)+\underset{\xi_k}{\mathrm{E}}[J_{k+1}(f_k(x,u,w_k),\xi_k)]\right].$$

3 Infinite Horizon Problems (*i.e.* $N\to\infty$)

Bellman Equation (BE)

$$J(x)=\min_{u\in\mathcal{U}(x)}\underset{w}{\mathrm{E}}[g(x,u,w)+J(f(x,u,w))],\ \forall x\in\mathcal{S}$$

Assuming that the limit of $N\to\infty$ of the DPA converges, $J(x)$ is the optimal cost-to-go. Note that the BE has to be solved for all $x\in\mathcal{S}$ simultaneously.

3.1 The Stochastic Shortest Path (SSP) Problem

Suppose the dynamics are $x_{k+1}=w_k$, $x_k\in\mathcal{S}$ and we are given the *probability transition matrix* $P_{ij}(u)$, $u\in\mathcal{U}(i)$.

Assumption 1 There exists a cost-free termination state 0 such that $P_{00}(u)=1$ and $g(0,u,0)=0$, $\forall u\in\mathcal{U}(0)$.

Remarks A well defined infinite horizon problem satisfies $\sum_{j\in\mathcal{S}}P_{ij}(u)=1$, $\forall i\in\mathcal{S}$. The probability of leaving the termination state must be 0.

Notation $\mathcal{S}^+:=\mathcal{S}\setminus\{0\}$

3.2 Theorem: SSP and BE

Definiton A stationary policy μ is said to be *proper* if, when using this policy, the probability of reaching the termination state is >0.

Assumption 2 There exists at least one proper policy $\mu\in\Pi$. Furthermore, for every improper policy μ', the cost $J_{\mu'}(i)$ is ∞ for at least one state $i\in\mathcal{S}$.

Theorem Under assumptions 1 and 2, and for the SSP problem:

1) Given any initial conditions $V_0(i)$, the sequence $V_\ell(i)$ generated by the iteration $V_{\ell+1}(i)=\min_{u\in\mathcal{U}(i)}\left(q(i,u)+\sum_{j=1}^nP_{ij}V_\ell(j)\right),\forall i\in\mathcal{S}^+$, where $q(i,u):=\underset{w}{\mathrm{E}}[g(i,u,w)]$, converges to the optimal cost $J^*(i)$ for all $i\in\mathcal{S}^+$.

2) The optimal costs satisfy the BE $\forall i\in\mathcal{S}^+$.

3) The solution to the BE is unique.

4) The minimizing u for each $i\in\mathcal{S}^+$ of the BE gives an optimal policy, which is proper.

4 Solving the Bellman Equation

4.1 Value Iteration (VI)

$$V_{\ell+1}(i)=\min_{u\in\mathcal{U}(i)}\left(q(i,u)+\sum_{j=1}^nP_{ij}(u)V_\ell(j)\right),\ \forall i\in\mathcal{S}^+$$

Converges in an infinite number of steps.

4.2 Policy Iteration (PI)

Initialize with a proper policy $\mu^0\in\Pi$

Stage 1 Given a policy μ^h, solve for the corresponding cost J_{μ^h} by solving the linear system of equations

$$J_{\mu^h}(i)=q(i,\mu^h(i))+\sum_{j=1}^nP_{ij}(\mu^h(i))J_{\mu^h}(j),\ \forall i\in\mathcal{S}^+$$

Stage 2 Obtain new stationary policy satisfying

$$\mu^{h+1}(i)=\underset{u\in\mathcal{U}(i)}{\operatorname{argmin}}\left(q(i,u)+\sum_{j=1}^nP_{ij}(u)J_{\mu^h}(j)\right),\ \forall i\in\mathcal{S}^+$$

Repeat until $J_{\mu^{h+1}}(i)=J_{\mu^h}(i)\ \forall i\in\mathcal{S}^+$.

Theorem Under assumptions 1 and 2, PI converges to an optimal policy after a finite number of steps.

Remark In every iteration of PI, the cost either decreases or stays the same.

4.3 Analogy and Comparison between VI and PI

Let p denote the maximum size of $\mathcal{U}(i)$ for all $i\in\mathcal{S}^+$.

Complexity of PI S1: n linear equations with n unknowns: $\mathcal{O}(n^3)$. S2: n minimizations over p possible controls, and evaluating the sum takes n steps: $\mathcal{O}(n^2p)$. Total: $\mathcal{O}(n^2(n+p))$ at each iteration. Number of iterations in worst case: p^n.

Complexity of VI n minimizations over p possible controls, and evaluating the sum takes n steps: $\mathcal{O}(n^2p)$ at each iteration.

4.4 Linear Programming (LP)

Theorem The solution to the optimization problem $\max_V\sum_{i\in\mathcal{S}^+}V(i)$ subject to $V(i)\le\left(q(i,u)+\sum_{j=1}^nP_{ij}(u)V(j)\right)$, $\forall u\in\mathcal{U}(i)$, $\forall i\in\mathcal{S}^+$ also solves the BE to yield the optimal cost J^* for the SSP problem.

5 Discounted Problems

Class of infinite horizon problems where there is no assumption of a termination state. Discount factor $\alpha<1$. By introducing a « virtual termination state » we have the associated SSP:

1) $P_{ij}(u)\leftarrow\alpha P_{ij}(u)$, $u\in\mathcal{U}(i),\forall i,j\in\mathcal{S}^+$
2) $P_{i0}(u)\leftarrow1-\alpha$, $u\in\mathcal{U}(i),\forall i\in\mathcal{S}^+$
3) $P_{0j}(u)\leftarrow0$, $u=\mathtt{stay},\forall j\in\mathcal{S}^+$
4) $P_{00}(u)\leftarrow1$, $u=\mathtt{stay}$.

The new BE is given by

$$J^*(i)=\min_{u\in\mathcal{U}(i)}\left[q(i,u)+\alpha\sum_{j=1}^nP_{ij}(U)J^*(j)\right],\ i\in\mathcal{S}^+$$

where $q(i,u)=\sum_{j=0}^{N-1}P_{ij}(u)g(i,u,j)$. Note that we also have to consider the termination state in discounted problems.

Remark $I-P$ is invertible if the policy inducing P is proper.

6 Shortest Path Problems and Deterministic Finite State Systems

6.1 The Shortest Path (SP) Problem

Vertex space $\mathcal{V}$, weighted edge space $\mathcal{C}:=\{(i,j,c_{i,j})\in\mathcal{V}\times\mathcal{V}\times\mathbb{R}\cup\{\infty\}|i,j\in\mathcal{V}\}$, path $Q:=(i_1,\dots,i_q)\in\mathcal{V}^q$, set of all paths that start at S and end at T is $\mathbb{Q}_{S,T}$. Path length $J_Q=\sum_{h=1}^{q-1}c_{i_h,i_{h+1}}$, objective $Q^*=\underset{Q\in\mathbb{Q}_{S,T}}{\operatorname{argmin}}J_Q$.

Assumption 3 $c_{i,i}\ge0$, $\forall i\in\mathcal{V}$ (no negative cycles).

6.2 Deterministic Finite State (DFS) Problem

No feedback needed since deterministic (*i.e.* $w_k=0,\forall k$).

6.3 Equivalence of SP and DFS

6.3.1 DFS to SP

Every state $x_k\in\mathcal{S}_k$ at each stage k is represented by a node in the graph: $\mathcal{V}_k:=\{(k,x_k)|x_k\in\mathcal{S}_k\}$, $k=0,\dots,N$. A virtual termination node T is added such that the arc lengths to T are simply the terminal costs of the DFS.

6.3.2 SP to DFS

We are given $\mathcal{V}$ and $\mathcal{C}$ and need to find the SP from node S to node T. Assume $c_{i,i}=0$, $\forall i\in\mathcal{V}$. Set $N:=|\mathcal{V}|-1$. Then we have $\mathcal{S}_k:=\mathcal{V}\backslash\{T\}, k\in\{1,\dots,N-1\}$, $\mathcal{S}_N:=\{T\}$, $\mathcal{S}_0:=\{S\}$, $\mathcal{U}_k:=\mathcal{V}\backslash\{T\}, k\in\{0,\dots,N-2\}$, $\mathcal{U}_{N-1}:=\{T\}$.

Dynamics $x_{k+1}=u_k, u_k\in\mathcal{U}_k, k\in\{0,\dots,N-1\}$

Stage costs $g_k(x_k,u_k):=c_{x_k,u_k}, k\in\{0,\dots,N-1\}, g_N(T):=0$.

We can solve this DFS using DPA, where $J_k(i)$ is the optimal cost of getting from node i to node T in $N-k=|\mathcal{V}|-1-k$ moves.

Forward DP algorithm SP is symmetric. Set $c_{i,j}\leftarrow c_{j,i}$.

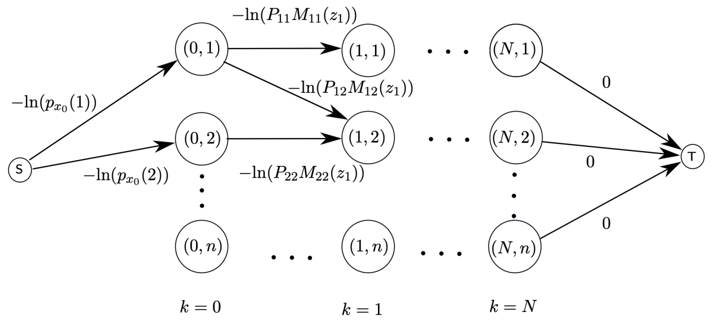
6.4 Hidden Markov Models and the Viterbi Algorithm

We want to convert an estimation problem to an SP problem. Consider the finite state, TI system $x_{k+1}=w_k$, $x_k\in\mathcal{S}$, $P_{ij}:=p_{w|x}(j|i)$, $\forall i,j\in\mathcal{S}$. The measurement model is $M_{ij}(z):=p_{z|x,w}(z|i,j), z\in\mathcal{Z}$ where $\mathcal{Z}$ is the measurement space and $p_{z|x,w}$ is the likelihood function. We assume independent observations, *i.e.* $z_k\perp(x_{n-1},z_n)|(x_{k-1},x_k), \forall n\leq k-1$.

Objective Let $Z_i:=(z_{i:N})$ and $X_i:=(x_{i:N})$. Given Z_1, we want to find most likely X_0, *i.e.* find a *maximum a posteriori (MAP)* estimate $\hat{X}_0=\operatorname{argmax}_{X_0}\ p(X_0|Z_1)$ or equivalently, find $\min_{X_0}\left(c_{S,(0,x_0)}+\sum_{k=1}^{N}c_{(k-1,x_{k-1}),(k,x_k)}\right)$ where

$$c_{S,(0,x_0)}=\begin{cases}-\ln p(x_0) & p(x_0)>0\\ \infty & p(x_0)=0\end{cases},\quad c_{(k-1,x_{k-1}),(k,x_k)}=\begin{cases}-\ln(\lambda) & \lambda>0\\ \infty & \lambda=0\end{cases}$$

and λ represents $P_{x_{k-1}x_k}M_{x_{k-1}x_k}(z_k)$.



7 Shortest Path Algorithms

7.1 Label correcting methods

To satisfy assumption 3, we assume additionally that $c_{i,j}\geq 0$, $\forall(i,j,c_{i,j})\in\mathcal{C}$.

0: Place S in OPEN, set $d_S=0, d_j=\infty\,\forall j\in\mathcal{V}\backslash\{S\}$.
1: Remove a node i from OPEN and execute step 2 for all children j of i.
2: If $d_i+c_{i,j}<\min\{d_j,d_T\}$, set $d_j=d_i+c_{i,j}$ and set i to be the parent of j. If $j\neq T$ place j in OPEN.
3: If OPEN is empty, we are done. Else go to step 1.

7.1.1 Methods to remove items from OPEN

- Depth-First Search: last in, first out
- Brendth-First Search: first in, first out
- Best-First Search (Dijkstra): remove best label, *i.e.* node i^* for which $d_{i^*}=\min_{i\in\text{OPEN}}d_i$

7.1.2 A* - Algorithm

Replace step 2 in the label correcting method by $d_i+c_{i,j}<\min\{d_j,d_T-h_j\}$, where h_j is some positive lower bound on the cost to go from j to T.

8 Deterministic Continuous Time Optimal Control and the HJB

Dynamics $\dot{x}(t)=f(x(t),u(t))$, state space $\mathcal{S}:=\mathbb{R}^n$, control constraint set $\mathcal{U}\subset\mathbb{R}^m$, feedback control law $u(t)=\mu(t,x)\in\mathcal{U}, \forall t\in[0,T], \forall x\in\mathcal{S}$, where $f\in C^1(\mathcal{S},\mathcal{U})$.

Assumption 4 For any admissible control law μ, initial time $t\in[0,T]$ and initial condition $x(t)\in\mathcal{S}$, there exists a unique state trajectory $x(\tau)$ that satisfies $\dot{x}(\tau)=f(x(\tau),u(\tau))$, $\forall\tau\in[t,T]$.

8.1 The HJB Equation

Assuming that $J^*(\cdot,\cdot)$ is differentiable w.r.t. t and x,

$$0=\min_{u\in\mathcal{U}}\left[g(x,u)+\frac{\partial J^*(t,x)}{\partial t}+\frac{\partial J^*(t,x)}{\partial x}f(x,u)\right],$$

$\forall x\in\mathcal{S}, \forall t\in[0,T]$ s.t. the terminal condition $J^*(T,x)=h(x), \forall x\in\mathcal{S}$.

8.1.1 Sufficiency of the HJB

Theorem Suppose $V(t,x)$ is a solution to the HJB equation and that $\mu(t,x)$ attains the minimum in the r.h.s of the HJB for all t and x. Then, under assumption 4, $V(t,x)$ is equal to the cost-to-go function, *i.e.* $V(t,x)=J^*(t,x), \forall x\in\mathcal{S}, t\in[0,T]$. Furthermore, the mapping μ is an optimal feedback law.

9 Pontryagin's Minimum Principle

Lemma Let $F(t,x,u)\in C^1(\mathbb{R},\mathbb{R}^n,\mathbb{R}^m)$ and let $\mathcal{U}\subseteq\mathbb{R}^m$ be a convex set. Assume $\mu^*(t,x):=\operatorname{argmin}_{u\in\mathcal{U}}F(t,x,u)$ exists and is continuously differentiable. Then for all t and x,

$$\frac{\partial\min_{u\in\mathcal{U}}F(t,x,u)}{\partial\lambda}=\frac{\partial F(t,x,u)}{\partial\lambda}\Big|_{u=\mu^*(t,x)}$$

where λ is either x or t.

9.1 The Minimum Principle

Cost: $h(x(T))+\int_0^T g(x(\tau),u(\tau))d\tau$.

Theorem For a given IC $x(0)=x_0\in\mathcal{S}$, let $u^*(t)$ be an optimal control trajectory with associated $x^*(t)$ for system $\dot{x}(t)=f(x(t),u(t))$. Then, we have

– State equation:

$$\dot{x}^*(t)=\frac{\partial H(x,u,p)}{\partial p}\Big|^{\top}_{x^*(t),u^*(t),p(t)},\ x^*(0)=x_0$$

– Adjoint (or co-state) equation:

$$\dot{p}(t)=-\frac{\partial H(x,u,p)}{\partial x}\Big|^{\top}_{x^*(t),u^*(t),p(t)},\ p(T)=\frac{\partial h(x)}{\partial x}\Big|^{\top}_{x^*(T)}$$

– Control input:

$$u^*(t)=\operatorname*{argmin}_{u\in\mathcal{U}}H(x^*(t),u,p(t))$$

– Hamiltonian:

$$H(x^*(t),u^*(t),p(t))=\text{const},\ \forall t\in[0,T]$$

where $H(x,u,p):=g(x,u)+p^{\top}f(x,u)$.

Remarks The minimum principle is a necessary condition for optimality. The HJB is a sufficient condition for optimality. If $f(x,u)$ is linear, $\mathcal{U}$ a convex set, h and g convex functions and the minimum principle is satisfied, then the solution is necessary and sufficient.

9.2 Fixed Terminal State (*i.e.* $x(T)=x_T$)

The ODE with $p(T)$ not valid anymore. The new ODEs are:

$$\dot{x}(t)=f(x(t),u(t)),\ x(0)=x_0,\ x(T)=x_T$$

$$\dot{p}(t)=-\frac{\partial H(x,u,p)}{\partial x}\Big|^{\top}_{x(t),u(t),p(t)}$$

9.3 Free initial state (*i.e.* $x(0)$ not fixed)

New total cost: $\ell(x(0))+h(x(T))+\int_0^T g(x(t),u(t))\,dt$. New boundary conditions: $p(0)=-\frac{\partial\ell(x)}{\partial x}\Big|^{\top}_{x(0)}$

9.4 Free Terminal Time (*i.e.* T not fixed)

$H(x(t),u(t),p(t))=0$, $\forall t\in[0,T]$ and the cost becomes $\int_0^T 1dt$ if no other cost is specified.

9.5 Time Varying System and Cost

Suppose $\dot{x}(t)=f(x(t),u(t),t)$ and the cost is $h(x(T))+\int_0^T g(x(\tau),u(\tau),\tau)\,d\tau$. Changes in Minimum Principle: H does not need to be constant along the optimal trajectory.

9.6 Singular Problems

In singular problems, $u(t)=\operatorname{argmin}_{u\in\mathcal{U}}H(x(t),u,p(t))$ is insufficient to determine $u(t)$ for all t, because H is independent of u over a time interval. The optimal trajectory is then divided into regular and singular arcs.

Hint For $t_2>t_1$ it holds: $p(t)=0\ \forall t\in[t_1,t_2]\implies\dot{p}(t)=0\ \forall t\in]t_1,t_2[$

A Notes on linear differential equations

Consider the ODE $\dot{x}(t)+x(t)=f(t)$. The homogeneous solution must satisfy $\dot{x}_H(t)+x_H(t)=0$ (*i.e.* the homogeneous equation). The particular solution $x_P(t)$ is a guess that should be of similar form to $f(t)$ and must satisfy the original ODE, *i.e.* $\dot{x}_P(t)+x_P(t)=f(t)$. The sum of both solutions $x(t)=x_H(t)+x_P(t)$ must satisfy the initial and terminal conditions.

A.1 Integrating factor

Consider the ODE $\dot{x}(t)+f(t)x(t)=h(t)$. Set the integrating factor to $I(t):=\exp\left[\int f(t)dt\right]$. A general solution is $x(t)=\frac{1}{I(t)}\left[\int I(t)h(t)dt+C\right]$.

B General notes

$E[f(X)]:=\sum_x f(x)p(X=x)$; $\text{var}(X):=E\left[(X-E[X])^2\right]=E[X^2]-E[X]^2$; $\text{var}(X+a)=\text{var}(X)$, $\text{var}(aX)=a^2\text{var}(X)$.

$\partial(Au)/\partial u=A$	$\partial(u^{\top}A)/\partial u=A^{\top}$
$\partial(x^{\top}Kx)/\partial x=x^{\top}(K+K^{\top})=2x^{\top}K$	

$$\begin{bmatrix}a & b\\ c & d\end{bmatrix}^{-1}=\frac{1}{\det(\cdot)}\begin{bmatrix}d & -b\\ -c & a\end{bmatrix}$$

Forcing Function	Trial Solution
ae^{rt}	Ae^{rt}
$a\sin(\omega t)$ or $a\cos(\omega t)$	$A\sin(\omega t)+B\cos(\omega t)$
at^n n a positive integer	$P(t)$ P a general polynomial of degree n
at^ne^{rt} n a positive integer	$P(t)e^{rt}$ P a general polynomial of degree n
$t^n[a\sin(\omega t)+b\cos(\omega t)]$ n a positive integer	$P(t)[A\sin(\omega t)+B\cos(\omega t)]$ P a general polynomial of degree n
$e^{rt}[a\sin(\omega t)+b\cos(\omega t)]$	$e^{rt}[A\sin(\omega t)+B\cos(\omega t)]$