Dynamic Programming

M403 Lecture Notes by Philip D. Loewen

Problem Statement. A "target set" S in $\mathbb{R}\times\mathbb{R}^n$ is given, along with a "domain of interest" $\Omega\subseteq\mathbb{R}\times\mathbb{R}^n$. We are interested in controlled trajectories $x(\cdot)$ with initial time τ and initial state ξ that can be steered to some point $(T,x(T))$ in S, without leaving the set Ω. These appear as triples in the following set:

$$\Sigma(\tau,\xi)=\left\{\begin{array}{lll}(T,u(\cdot),x(\cdot)) : & & \\ & T\ge\tau,\ u\in\mathrm{PWC}[0,T], & x\in\mathrm{PWS}[0,T] \\ & \dot{x}(t)=f(t,x(t),u(t)) & \text{a.e. } t\in[\tau,T] \\ & u(t)\in U & \text{a.e. } t\in[\tau,T] \\ & x(\tau)=\xi & \\ & (T,x(T))\in S & \\ & (t,x(t))\in\Omega, & \forall t\in(\tau,T).\end{array}\right\}$$

Note that $\Sigma(\tau,\xi)\ne\emptyset$ if and only if some triple actually satisfies the defining constraints; any such triple is called "feasible". Given any (T,u,x) in $\Sigma(\tau,\xi)$, restricting the time-varying control and state functions to a final subinterval of $[\tau,T]$ will show that $\Sigma(t,x(t))\ne\emptyset$ for each $t\in[\tau,T]$.

Given an endpoint cost $\ell\colon S\to\mathbb{R}$ and a running cost L, we let $P(\tau,\xi)$ denote the problem

$$P(\tau,\xi)\qquad\begin{cases}\text{minimize} & \ell(T,x(T))+\displaystyle\int_\tau^T L(r,x(r),u(r))\,dr \\ \text{over all} & (T,u(\cdot),x(\cdot))\in\Sigma(\tau,\xi).\end{cases}$$

Finally, we introduce the **value function** V to record the minimum value above as a function of the starting point: $V=V(\tau,\xi)$. In situations where $\Sigma(\tau,\xi)=\emptyset$, we declare $V(\tau,\xi)=+\infty$.

Proposition 1 (Principle of Optimality). *Fix any point $(\tau,\xi)\in\mathbb{R}\times\mathbb{R}^n$ where $\Sigma(\tau,\xi)\ne\emptyset$.*

(a) *For any $(T,u,x)\in\Sigma(\tau,\xi)$, the function below is nondecreasing on $[\tau,T]$:*

$$t\mapsto V(t,x(t))-\int_t^T L(r,x(r),u(r))\,dr.$$

(b) *if a triple $(\widehat{T},\widehat{u},\widehat{x})$ achieves the minimum in problem $P(\tau,\xi)$, then the associated function from (a) is constant, i.e.,*

$$V(t,\widehat{x}(t))-\int_t^{\widehat{T}} L(r,\widehat{x}(r),\widehat{u}(r))\,dr=\ell(\widehat{T},\widehat{x}(\widehat{T}))\qquad\forall t\in[\tau,\widehat{T}].$$

Proof. (a) Pick any subinterval $[r,s]$ of $[\tau,T]$. The minimization process that defines $V(r,x(r))$ allows for any feasible control at all defined on $[r,T]$. Restricting the choice to those controls that match the given function u for the segment $[r,s]$ allows less competition for the minimum, so we must have

$$V(r,x(r)) \le \int_r^s L(t,x(t),u(t))\,dt + V(s,x(s)).$$

Now

$$\int_r^s L(t,x(t),u(t))\,dt = \int_r^T L(t,x(t),u(t))\,dt - \int_s^T L(t,x(t),u(t))\,dt,$$

and this implies the desired result:

$$V(r,x(r)) - \int_r^T L(t,x(t),u(t))\,dt \le V(s,x(s)) - \int_s^T L(t,x(t),u(t))\,dt. \qquad (\dagger)$$

(b) Now suppose that for the triple in part (a), some subinterval $[r,s]$ of $[\tau,T]$ gives a strict inequality in $(\dagger)$. Then, by repeated application of part (a),

$$\begin{aligned} V(\tau,\xi) - \int_\tau^T L(t,x(t),u(t))\,dt &\le V(r,x(r)) - \int_r^T L(t,x(t),u(t))\,dt \\ &< V(s,x(s)) - \int_s^T L(t,x(t),u(t))\,dt \\ &\le V(T,x(T)) - 0 \\ &= \ell(T,x(T)). \end{aligned}$$

That is, $V(\tau,\xi) < \ell(T,x(T)) + \int_\tau^T L(t,x(t),u(t))\,dt$. In other words, (T,u,x) is not a true minimizer in problem $P(\tau,\xi)$. Take the contrapositive: any minimizing triple must give equality in $(\dagger)$ for every subinterval $[r,s]$ of $[\tau,T]$. ////

Corollary. *Suppose (τ,ξ) has a neighbourhood on which $\Sigma(\tau',\xi') \ne \emptyset$. If V is differentiable at (τ,ξ), then*

$$V_t(\tau,\xi) + \langle V_x(\tau,\xi)\,,\, f(\tau,\xi,w)\rangle + L(\tau,\xi,w) \ge 0 \qquad \forall w \in U. \qquad (1)$$

If, in addition, problem $P(\tau,\xi)$ has a true minimizer $(\widehat{T},\widehat{u},\widehat{x})$, then also

$$V_t(\tau,\xi) + \left\langle V_x(\tau,\xi)\,,\, f(\tau,\xi,\widehat{u}(\tau^+))\right\rangle + L(\tau,\xi,\widehat{u}(\tau^+)) = 0. \qquad (2)$$

Proof. The hypothesis of differentiability implies that V has a well-defined finite value at each point in some open set containing (τ,ξ). It follows that for any fixed $w \in U$, there is a small $h > 0$ for which applying the control $u(t) = w$ on the interval $[\tau,\tau+h]$ will drive the state to a point $x(\tau+h)$ where $V(\tau+h,x(\tau+h))$ is finite. Thus it is possible to find a triple (T, u, x) that satisfies the constraints in $\Sigma(\tau, \xi)$ and the condition $u(\tau^+) = w$. Along this triple, the monotonicity assertion of part (a) above implies

$$\begin{aligned} 0 &\le \frac{d}{dt}\left[V(t, x(t)) - \int_t^T L(r, x(r), u(r))\, dr\right]_{t=\tau^+} \\ &= V_t(\tau, \xi) + \left\langle V_x(\tau, \xi)\, ,\, \dot{x}(\tau^+)\right\rangle + L(\tau, \xi, u(\tau^+)) \\ &= V_t(\tau, \xi) + \langle V_x(\tau, \xi)\, ,\, f(\tau, \xi, w)\rangle + L(\tau, \xi, w). \end{aligned}$$

This gives inequality (1), and starting with part (b) above instead leads to (2) by a very similar process. ////

Hamiltonian Form. Recall the true (or "maximized") Hamiltonian

$$\mathcal{H}(t, x, p) = \max\left\{\langle p\, ,\, f(t, x, w)\rangle - L(t, x, w)\, :\, w \in U\right\}.$$

It will be convenient to have a name for the set of maximizers in the definition above:

$$\widehat{U}(t, x, p) = \arg\max\left\{\langle p\, ,\, f(t, x, w)\rangle - L(t, x, w)\, :\, w \in U\right\}.$$

Line (1) above says

$$\langle -V_x(\tau, \xi)\, ,\, f(\tau, \xi, w)\rangle - L(\tau, \xi, w) \le -V_t(\tau, \xi) \qquad \forall w \in U,$$

which is equivalent to

$$\mathcal{H}(\tau, \xi, -V_x(\tau, \xi)) \le V_t(\tau, \xi).$$

Line (2) says that equality holds, with the additional note that

$$\widehat{u}(\tau^+) \in \widehat{U}(\tau, \xi, -V_x(\tau, \xi)).$$

Summary (Best-Case). Suppose Ω is an open set, such that V is differentiable and problem P has a minimizer for each (τ, ξ) in Ω. Then V satisfies the Hamilton-Jacobi Equation with boundary condition below:

$$\begin{aligned} V_t(t, x) &= \mathcal{H}(t, x, -V_x(t, x)), && (t, x) \in \Omega, && (HJE) \\ V(t, x) &= \ell(t, x), && (t, x) \in S. && (BC) \end{aligned}$$

Verification Theory. Here comes our first **sufficient condition** for optimality. We try running the arguments describing the value function V in reverse, paying close attention to the properties required to make this succeed. In what follows, W is just some scalar-valued function; all of its properties are listed in the formal statement.

Proposition 2 (Verification). *Suppose $W\colon \Omega \to \mathbb{R}$ has the property that*

$$t \mapsto W(t, x(t)) - \int_t^T L(r, x(r), u(r))\, dr \qquad (*)$$

is nondecreasing along every triple $(T, x, u) \in \Sigma(\tau, \xi)$. Suppose also that

$$W(t, x) \le \ell(t, x) \qquad \forall (t, x) \in S.$$

Then $W(\tau, \xi) \ge V(\tau, \xi)$; indeed,

$$W(t, x(t)) \ge V(t, x(t)) \qquad \forall t \in [\tau, T].$$

If some special triple $(\widehat{T}, \widehat{u}, \widehat{x})$ makes the function in line $(*)$ a constant, and ends at a point of S where

$$W(\widehat{T}, \widehat{x}(\widehat{T})) = \ell(\widehat{T}, \widehat{x}(\widehat{T})),$$

then this triple is a true minimizer in $P(\tau, \xi)$, and $W(t, \widehat{x}(t)) = V(t, \widehat{x}(t))$ holds for each $t \in [\tau, \widehat{T}]$.

Proof. For each fixed triple (T, u, x) in $\Sigma(\tau, \xi)$, using $\xi = x(\tau)$ gives

$$W(\tau, \xi) - \int_\tau^T L(r, x(r), u(r))\, dr \le W(T, x(T)) \le \ell(T, x(T)), \qquad (\ddagger)$$

so

$$\ell(T, x(T)) + \int_\tau^T L(r, x(r), u(r))\, dr \ge W(\tau, \xi).$$

Taking the infimum over all triples in $\Sigma(\tau, \xi)$ gives $V(\tau, \xi) \ge W(\tau, \xi)$.

Given (τ, ξ) where the stated monotonicity and endpoint properties hold, pick any triple $(T, u, x) \in \Sigma(\tau, \xi)$. Then each point $(\tau', \xi') = (\tau', x(\tau'))$ with $\tau' \in [\tau, T]$ inherits the same monotonicity and endpoint properties, so the argument of the previous paragraph applies to it. Hence $V(t, x(t)) \le W(t, x(t))$ for each $t \in [\tau, T]$.

For the special triple $(\widehat{T}, \widehat{u}, \widehat{x})$, our hypotheses give equality in the two crucial points of $(\ddagger)$, leading to

$$\ell(\widehat{T}, \widehat{x}(\widehat{T})) + \int_\tau^{\widehat{T}} L(r, \widehat{x}(r), \widehat{u}(r))\, dr = W(\tau, \xi).$$

Combine this with the previous line to get the result. ////

Corollary A. *Suppose $W = W(t, x)$ is a differentiable function satisfying*

$$\begin{aligned} W_t(t, x) &\ge \mathcal{H}(t, x, -W_x(t, x)), && (t, x) \in \Omega, \\ W(t, x) &\le \ell(t, x), && (t, x) \in S. \end{aligned}$$

Then for any (τ,ξ), any triple $(\widehat{T},\widehat{u},\widehat{x})\in\Sigma(\tau,\xi)$ satisfying conditions (i)–(iii) below is a true minimizer for $P(\tau,\xi)$:

$$\begin{array}{llll}
\textit{(i)} & W_t(t,\widehat{x}(t)) & = \mathcal{H}(t,\widehat{x}(t),-W_x(t,\widehat{x}(t))), & t\in[\tau,\widehat{T}],\\
\textit{(ii)} & W(\widehat{T},\widehat{x}(\widehat{T})) & = \ell(\widehat{T},\widehat{x}(\widehat{T})), & (t,x)\in S,\\
\textit{(iii)} & \widehat{u}(t) & \in \widehat{U}(t,\widehat{x}(t),-W_x(t,\widehat{x}(t))), & t\in[\tau,\widehat{T}],
\end{array}$$

is a true minimizer in $P(\tau,\xi)$, and

$$W(t,\widehat{x}(t)) = V(t,\widehat{x}(t)) \qquad \forall t\in[\tau,\widehat{T}].$$

Proof. Consider any $(T,u,x)\in\Sigma(\tau,\xi)$. Then the general Hamiltonian inequality implies that

$$W_t(t,x(t)) \ge \langle -W_x(t,x(t))\,,\, f(t,x(t),u(t))\rangle - L(t,x(t),u(t)) \quad \text{a.e. } t\in[\tau,T],$$

i.e.,

$$\frac{d}{dt}\left[W(t,x(t)) - \int_t^T L(r,x(r),u(r))\,dr\right] \ge 0 \quad \text{a.e. } t\in[\tau,T].$$

This establishes the nondecreasing property required in the Proposition. The Hamiltonian equation involving $(\widehat{T},\widehat{u},\widehat{x})$ implies that equality holds along this particular triple, so the bracketed function is constant. The endpoint conditions here are the same as the ones in the Proposition, so the result follows. ////

The super-lovely case goes like this.

Corollary B. *Suppose W is a continuously differentiable function on Ω for which*

$$\begin{array}{lll}
W_t(t,x) = \mathcal{H}(t,x,-W_x(t,x)), & (t,x)\in\Omega, & (HJE)\\
W(t,x) = \ell(t,x), & (t,x)\in S. & (BC)
\end{array}$$

Then for each (τ,ξ), every triple (T,u,x) in $\Sigma(\tau,\xi)$ that solves the initial-value problem

$$\dot{x} = f(t,x,\widehat{U}(t,x,-W_x(t,x))), \qquad x(\tau)=\xi$$

is a true minimizer in $P(\tau,\xi)$, and $W(\tau,\xi)=V(\tau,\xi)$.

Proof. Each of (HJE) and (BC) confirms two lines in the setup of Corollary B. ////

Optimal Feedback Control. When a solution W to (HJE)/(BC) is available, it provides a recipe for generating the optimal input for any time and state. This is of huge practical importance: it's a "closed-loop" strategy instead of an "open-loop" one.

Example. Use a verification function of the form $W(t,x) = \phi(t)x + g(t)$ with Corollary A to prove that $\widehat{x}(t) = e^t$ is a true minimizer for

$$\int_0^1 \left(\tfrac{1}{2}\dot{x}(t)^2 + \tfrac{1}{2}x(t)^2\right)\,dt$$

subject to the endpoint constraints $x(0) = 1$, $x(1) = e$.

Proof. Set $u = \dot{x}$ to put this into a standard control form, with

$$f(t,x,u) = u, \qquad L(t,x,u) = \tfrac{1}{2}u^2 + \tfrac{1}{2}x^2, \qquad U = \mathbb{R}.$$

Observe that

$$\begin{aligned} H(t,x,p,u) &= pu - \tfrac{1}{2}u^2 - \tfrac{1}{2}x^2, \\ \widehat{U}(t,x,p) &= \{p\} \qquad \left[\text{from solving } H_u = 0\right], \\ \mathcal{H}(t,x,p) &= \tfrac{1}{2}p^2 - \tfrac{1}{2}x^2. \end{aligned}$$

Also, the proposed form for W entails

$$W_t(t,x) = \dot{\phi}(t)x + \dot{g}(t), \qquad W_x(t,x) = \phi(t).$$

Now consider $\widehat{u}(t) = \dot{\widehat{x}}(t) = e^t$. Condition (iii) of Cor. B requires

$$e^t = \widehat{u}(t) \in \{-W_x(t,\widehat{x}(t))\} = \{-\phi(t)\},$$

so $\phi(t) = -e^t$. Then condition (i) can only work if

$$\dot{\phi}(t)\widehat{x}(t) + \dot{g}(t) = \tfrac{1}{2}\phi(t)^2 - \tfrac{1}{2}\widehat{x}(t)^2, \quad \text{i.e.,} \quad -e^{2t} + \dot{g}(t) = \tfrac{1}{2}e^{2t} - \tfrac{1}{2}e^{2t} = 0, \quad \text{i.e.,} \quad \dot{g}(t) = e^{2t}.$$

We conclude that $g(t) = \frac{1}{2}e^{2t} + K$ for some constant K, so

$$W(t,x) = \tfrac{1}{2}e^{2t} - e^t x + K.$$

With this general form, we can evaluate

$$W_t(t,x) - \mathcal{H}(t,x,-W_x(t,x)) = e^{2t} - e^t x - \left(\tfrac{1}{2}(-e^t)^2 - \tfrac{1}{2}x^2\right) = \tfrac{1}{2}(e^t - x)^2 \geq 0.$$

Thus all the conditions of Cor. A except for the ones involving S and ℓ are in force. If we choose $S = \{(1,e)\}$ and $K = \frac{1}{2}e^{2t}$, then we have $W(1,e) = 0 = \ell(1,e)$ and the desired optimality of $\widehat{x}$ follows directly from Cor. A. ////

Example. Here's a classic from the Calculus of Variations:

$$\min\left\{\int_0^{\pi/2} \left(\tfrac{1}{2}u(t)^2 - \tfrac{1}{2}x(t)^2\right)\,dt \;:\; \dot{x} = u,\; x(0) = 0\right\}.$$

The fixed final time $T = \pi/2$ and free final state $x(\pi/2)$ reveal the target set $S = \{(\pi/2, x) : x \in \mathbb{R}\}$—a vertical line in the (t, x)-plane.

The given problem is case $P(0, 0)$ of the general form $P(\tau, \xi)$ we can produce by changing the initial condition to $x(\tau) = \xi$. Let's apply the PMP to find extremals. Here $f(t, x, u) = u$ and $L(t, x, u) = \frac{1}{2}u^2 - \frac{1}{2}x^2$, so

$$\begin{aligned} H(t, x, p, u) &= pu - \tfrac{1}{2}u^2 + \tfrac{1}{2}x^2, \\ \widehat{U}(t, x, p) &= p, \\ \mathcal{H}(t, x, p) &= \tfrac{1}{2}p^2 + \tfrac{1}{2}x^2. \end{aligned}$$

If $(\widehat{u}, x)$ is to be an extremal pair, then some costate arc p must obey

$$\dot{x} = \widehat{u} = p, \qquad -\dot{p} = H_x(t, x, p, \widehat{u}) = x.$$

Therefore $\ddot{x} + x = 0$, and $x(t) = A\cos t + B\sin t$ for some constants A, B. We deduce $p(t) = \dot{x} = B\cos t - A\sin t$. Transversality (note $\ell \equiv 0$) requires $-p(\pi/2) = 0$, so $A = 0$ and $x(t) = B\sin t$. The corresponding control is $\widehat{u}(t) = p(t) = B\cos t$. The cost of this pair is

$$\begin{aligned} \int_\tau^{\pi/2} \frac{1}{2}\left(B^2\cos^2 t - B^2\sin^2 t\right)\, dt &= \frac{B^2}{2}\left[\frac{\sin(2t)}{2}\right]_{t=\tau}^{\pi/2} \\ &= -\frac{B^2}{4}\sin(2\tau) \\ &= -\frac{B^2}{2}\sin(\tau)\cos(\tau). \end{aligned}$$

Using $x(\tau) = \xi$ determines $B = \xi/\sin(\tau)$, so in detail the extremal ingredients are

$$x(t) = \left(\frac{\sin t}{\sin\tau}\right)\xi, \qquad \widehat{u}(t) = p(t) = \left(\frac{\cos t}{\sin\tau}\right)\xi,$$

with cost $-\dfrac{\xi^2\cos\tau}{2\sin\tau}$. We suspect this of providing the minimum, so we investigate

$$W(t, x) := -\left(\frac{\cos t}{2\sin t}\right)x^2.$$

Here $W_x(t, x) = -x\cos t/\sin t$ and

$$W_t(t, x) = -\left(\frac{-\sin^2 t - \cos^2 t}{2\sin^2 t}\right)x^2 = \frac{x^2}{2\sin^2 t}.$$

This actually satisfies (HJE):

$$\mathcal{H}(t, x, -W_x(t, x)) = \tfrac{1}{2}\left(\frac{x\cos t}{\sin t}\right)^2 + \tfrac{1}{2}x^2 = \tfrac{1}{2}\frac{x^2\cos^2 t}{\sin^2 t} + \tfrac{1}{2}\frac{x^2\sin^2 t}{\sin^2 t} = \tfrac{1}{2}\frac{x^2}{\sin^2 t} = W_t(t, x).$$

Note also that $W(\pi/2, x) = 0 = \ell(\pi/2, x)$ for all x. All the nice results above are available, but we must be a little careful about the set Ω. The function W has trouble when $\tau = 0$, so we set $\Omega = (0, \pi/2) \times \mathbb{R}$. For each (τ, ξ) in Ω, Corollary B implies that the extremal arc identified above is a true minimizer; further, we have $V(\tau, \xi) = W(\tau, \xi)$. The optimal feedback law provides the flow-field in (t, x)-space: every trajectory for

$$\dot{x}(t) = \widehat{U}(t, x, -W_x(t, x)) = \left(\frac{\cos t}{\sin t} \right) x$$

on an interval of the form $[\tau, \pi/2]$ with $\tau > 0$ provides an optimal path to the target set for its particular initial point.

The original problem $P(0, 0)$ is interesting because $(0, 0) \notin \Omega$. However, any $(\pi/2, u, x)$ in $\Sigma(0, 0)$ must give a finite value to $\dot{x}(0^+)$, and therefore

$$\lim_{t \to 0^+} V(t, x(t)) = \lim_{t \to 0^+} W(t, x(t)) = -\tfrac{1}{2} \lim_{t \to 0^+} \left[\frac{x(t)}{t} \right] \left(\frac{t}{\sin t} \right) (x(t) \cos(t)) = 0.$$

Therefore $V(0, 0) = 0$ and there is a huge tie for optimality in $P(0, 0)$ between all arcs of the form $x(t) = B \sin t$, $B \in \mathbb{R}$.

Linear-Quadratic Regulator. Solve the Matrix Riccati Equation backwards in time.