Lecture 2
Thursday, January 18, 2007
Scribe: Omar Bakr

To illustrate the control problem for finite horizon, fully observable systems, consider the *trellis*:

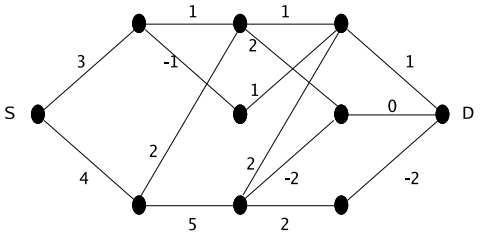

Figure 1: Minimum weight path example (Trellis)

We define a trellis to be a directed acyclic graph whose nodes are partitioned into disjoint nonempty subsets called *levels*, numbered 0 through N. There is unique node at level 0, called the *source*, and a unique node at level N, called the *destination*. Every edge starting from a node at level k terminates at a node at level $k+1$. Further, every node must lie on some path from the source to the destination. The trellis in Figure 1 has $N = 4$, source node S and destination node D.

The number on each edge is called its *weight*. Our goal is to find the minimum weight path from the source node to the destination node, where the weight of a path is defined as the sum of weights of the edges along the path. The direct (brute force) way is to enumerate all paths, evaluate their weights, and then choose the path with the minimum weight. We will now discuss a more intelligent approach to solve this problem, whose underlying principles will apply to a broader range of problems.

The minimum weight path problem on a trellis is a discrete time control problem (finite horizon, fully observed) of the general kind introduced in Lecture 1. The underlying dynamical system is:

$$x_{k+1} = f_k(x_k, u_k, w_k) \quad 0 \leq k \leq N-1$$

Here $w_k = 0$ for all k, i.e. the system is deterministic. x_k denotes the node at level k in the path, u_k specifies the outgoing edge at state x_k, its choice then determines the node on the path at level $k+1$. The aim to minimize a cost, which fits into our general form:

$$E[\sum_{k=0}^{N-1} g_k(X_k, U_k, w_k) + g_N(X_N)] \leftarrow \text{ informal expression}$$

where $g_k(x_k, u_k, w_k)$ is the weight of the edge corresponding to choosing the control u_k at the node x_k at level k and $g_N(x_N)$ may be taken to be 0 (note that D is only possibility for x_N). Since $w_k = 0$ for all k, the one step costs are also deterministic.

A simpler approach than brute force to solve the minimum weight path problem becomes apparent if one recognizes "Bellman's principle of optimality" (i.e. if a strategy is optimal at time 0, then for each

$k \geq 0$, the portion of the strategy from time k onwards, conditioned on the knowledge of the controller at time k, must be optimal). So in the minimum weight path problem, we can apply this principle. For each node, we compute the minimum *cost-to-go* recursively from the end point, as shown in Figure 2.

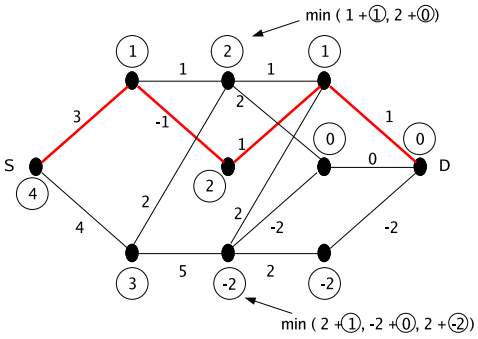

Figure 2: Computing the minimum weight path recursively from the end for a weighted trellis

The cost-to-go at any state (node) at time k is expressed as a minimum, over control choices at that state, of the sum of the one step cost and the remaining cost-to-go after taking that step. Any control that minimizes in this calculation could be chosen as the control choice at that state in defining an optimum strategy (going forward).

Note that in the minimum weight path problem any optimal control strategy can be thought of as being defined in open-loop (i.e. without reference to the states). This is because the underlying dynamical system is deterministic. In the example above, for instance, an optimal strategy would be to simply specify, in open loop, the control sequence (A,B,A,A), see Figure 3.

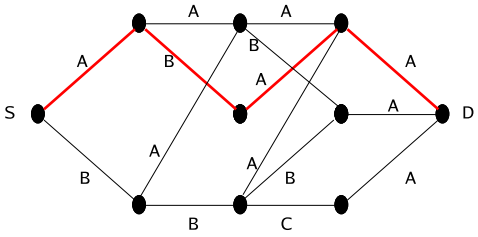

Figure 3: Viewing the minimum weight path as a control problem: A, B, C are names of controls. The red line is the (unique) minimum weight path in this example

This is not true in the stochastic case (i.e. open loop controls can be strictly worse than closed loop controls. Let's see this in a stochastic version of the minimum weight path problem (Figure 4).

In Figure 4 both controls A and B lead from x_{12} to D, but with different costs (see Table 1). Similarly, both controls A and B lead from x_{22} to D, but with different costs (see Table 1). Assume that $g(S, A) = 0$ and that action A leads from S to x_{11}. Assume that $g(S, B) = 0$ and that action B leads from S to x_{21}.

The cost-to-go at state x_{12} is $\min\{1, 50\} = 1$, and the best control in state x_{12} is A. The cost-to-go at state x_{22} is $\min\{50, 2\} = 2$, and the best control in state x_{22} is B. For convenience these values have been written on the respective edges from x_{12} to D and x_{22} to D; don't confuse this with edge weights.

The expected cost-to-go at state x_{11} (see Table 2) is thus:

$$\min\{\underbrace{3 + (\frac{1}{2}(1) + \frac{1}{2}(2))}_{\text{for control A}}, \underbrace{-2 + (\frac{2}{3}(1) + \frac{1}{3}(2))}_{\text{for control B}}\} = -\frac{2}{3} .$$

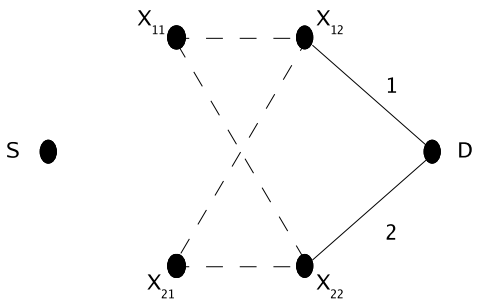

Figure 4: Stochastic version of the trellis problem: Two control actions A and B are possible in each state. Both transitions are possible from each of x_{11} and x_{21} under control A and under control B with some probabilities (say $\frac{1}{2}, \frac{1}{2}$ for A, $\frac{2}{3}, \frac{1}{3}$ for B).

The best control choice in state x_{11} is B.

The expected cost-to-go at state x_{21} (see Table 2) is:

$$\min\{1 + (\frac{1}{2}(1) + \frac{1}{2}(2)), -4 + (\frac{2}{3}(1) + \frac{1}{3}(2))\} = -\frac{8}{3} \ .$$

The best control choice in state x_{21} is B.

Finally, the expected cost-to-go in state S is $\min\{0 - \frac{2}{3}, 0 - \frac{8}{3}\} = -\frac{8}{3}$. The best choice of control in state S is B.

State	Control	$g(x,u)$
X_{12}	A	1
	B	50
X_{22}	A	50
	B	2

Table 1: Stochastic trellis at time 2. Cost $g(x,u)$ is assumed to be deterministic.

State	Control	$g(x,u)$
X_{11}	A	3
	B	-2
X_{21}	A	1
	B	-4

Table 2: Stochastic trellis at time 1. Cost $g(x,u)$ is assumed to be deterministic.

Note that the optimal strategy has to be closed loop. This is because we cannot decide the last (time 2) control action in open loop, since whether we want to make it A or B depends on where the randomness leads us when we use the control action at time 1 (which in this example is B in either of the states x_{11} and x_{21} at time 1). The outcome of this randomness cannot be seen until time 2, so the optimal control has to be decided based on this feedback.

For another example of this important conceptual point (relevance of feedback in stochastic control is deeper than in deterministic control), consider the scalar linear system control problem (fully observed, finite horizon)

$$x_{k+1} = x_k + u_k + w_k$$

$$w_k \backsim N(0, \sigma^2) \text{ and independent}$$

starting at x_0, we want to minimize $E[\sum_{k=0}^{N} x_k^2]$. Consider the closed loop strategy:

$$u_k = -x_k$$

The state sequence is then $x_0 = 0, x_1 = w_1, x_2 = w_2, \ldots, x_N = w_N$. The realized mean total cost is $E[\sum_{k=1}^{N} w_k^2] = N\sigma^2$.

However, for any open loop strategy $u_0, \ldots, u_{N-1}$:

$$x_0 = 0, x_1 = u_0 + w_1, x_2 = u_0 + u_1 + w_1 + w_2, \ldots, x_k = \sum_{l=0}^{k} u_{l-1} + \sum_{l=0}^{k} w_l$$

$$E[\sum_{k=0}^{N} x_k^2] \geq \sigma^2 \sum_{k=1}^{N} k = \sigma^2 \frac{N(N+1)}{2}$$

This lower bound on what can be achieved by open loop control strategies is, for large N, dramatically worse than what the closed loop control strategy can achieve.

Looking ahead to the next lecture, let us introduce some notation. Consider the dynamical system:

$$x_{k+1} = f_k(x_k, u_k, w_k)$$

over a finite horizon N and consider the fully observed control problem (in informal notation) of minimizing:

$$E[\sum_{k=0}^{N-1} g_k(X_k, U_k, w_k) + g_N(X_N)]$$

To be precise about what the problem actually is, we define a *control strategy* $\Psi = (\nu_0, \nu_1, \ldots, \nu_{N-1})$, via a collection of functions, where $\nu_k(x_1, \ldots, x_k)$ for $k = 0, \ldots, N-1$ is a map from the available observations (states) up to time k, taking values in the space of possible controls at state x_k (at time k). Our optimization problem is really one of minimizing the cost over all such Ψ (where the states and controls at each time in the expression for the expected cost to be minimized are now random variables that depend not only on the underlying noise, but also on the choice of control strategy Ψ).

A special kind of control strategies is the so called *Markov strategies*:

$$\pi = (\mu_0, \mu_1, \ldots, \mu_{N-1})$$

defined through functions $\mu_k(x_k)$ taking values in the possible control actions at state x_k at time k. Thus, in Markov strategies at each time only the most recent state is used in choosing the control.

To solve the control problem, let us recursively (from the end) define the following functions:

$$J_N(x_N) = g_N(x_N)$$

$$J_k(x_k) = \inf_{u \text{ in the set of possible controls at state } x_k} E[g_k(x_k, u, w_k) + J_{k+1}(f_k(x_k, u, w_k))]$$

These equations do not refer to any strategy as such.

The main result in the theory, informally, will be that the optimal control strategies are Markov and given by the minimizers in these equations (as a function of x_k for each k).