Lecture 1 (*Scribe notes by Brian Gawalt*)

Stochastic control is the application of control theory to systems operating under uncertainty, and that uncertainty is modeled probabilistically. A dynamical system - input/output, state space, etc. - is observed and then acted upon in a closed loop. This course focuses on such systems, mostly in discrete time.

Below is a state space model for a stochastic, dynamical, discrete-time system.

$$\begin{aligned} x_{k+1} &= f_k(x_k, u_k, w_k) \\ x_k &:= \text{state at time } k \\ u_k &:= \text{input applied at time } k \\ w_k &:= \text{a random variable for time } k \\ f_k &:= \text{the nonrandom, time-dependent state-transition function at time } k \end{aligned}$$

This system has an initial condition of x_0. This value can be either deterministic or random.

This course discusses how to choose the controls to best influence the behavior of the above system. If all states are visible to the controller (making it a fully observed dynamical system), we could choose u_k as a function of x_0 through x_k. Often, however, the states are not completely visible. We assume in these conditions that we have a collection of observations of the system instead:

$$\begin{aligned} y_0 &= h_0(x_0, v_0) \\ y_k &= h_k(x_k, u_{k-1}, v_k) \end{aligned}$$

The observation at time k, y_k, is a time-dependent function of the current state, the previous input, and some observation noise (represented by the series of random variables v_k). We would then get to choose u_k as a function of $y_0, ..., y_k$ in this partially-observed case. It's worth noting here that unless otherwise noted, w_k, v_k, and x_0 are all assumed to be independent. However, problems with correlated noise can often be reduced to this case by augmenting the underlying state space of the model and treating the augmented state as partially observed, see section 1.4 of the text.

The control goal is viewed as the minimization of an expected cost. This cost can be finite horizon or infinite horizon - the situations differ in terms of how many iterations the system will be allowed (how big k can get). For a fully observed case, the finite horizon cost is:

$$E\left[\sum_{k=0}^{N-1} g_k(X_k^\Psi, \nu_k\left(X_0^\Psi, ..., X_k^\Psi\right), w_k) + g_N(X_N^\Psi)\right]$$

Here, N is the total number of iterations, and $g_k(x_k, u_k, w_k)$, $0 \leq k \leq N-1$, and $g_N(x_N)$ are prescribed cost functions. $\Psi = (\nu_0, ..., \nu_{N-1})$ is the control strategy, where $\nu_0, \ldots, \nu_{N-1}$ are N functions, one at each time $k = 0, \ldots, N-1$, defining the strategy. Each ν_k tells us how to choose the control at time k, u_k, in the space of allowed control actions at the state at time k, x_k, as a function of the available observations at time k, i.e. the states $x_0, \ldots, x_k$ (because the problem is a fully observed one). In the notation above, X_k^Ψ is a random variable giving the state at time k when the control strategy Ψ is used. Thus the control chosen at time k

would be $\nu_k\left(X_0^\Psi, ..., X_k^\Psi\right)$ and could also be written as U_k^Ψ. Also, X_0^Ψ equals X_0 the (possibly random) initial condition, irrespective of Ψ. We wish to minimize this expected cost over all possible strategies.

Example: Inventory Management

Let x_k be the amount of inventory in the warehouse on day k. u_k is the amount of new inventory ordered on day k. w_k, a random variable, represents the amount of inventory decreased by consumer demand on day k. Suppose we require that u_k be non-negative, but x_k itself can be any real number (with negative values meaning we've had to borrow inventory from somewhere else). This scenario has the following state transition function:

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

The cost of ordering new inventory is proportional to the amount ordered with proportionality constant c, and the cost of maintaining the warehouse's inventory is a function of the amount stored/borrowed $r(x_k)$. The terminal cost at the horizon is $R(x_N)$. See example 1.1.1 of the text. Our aim would be to minimize:

$$E\left[\sum_{k=0}^{N-1}\left(r\left(X_k\right) + cU_k\right) + R\left(X_N\right)\right]$$

The notation here is informal. Formally, we would write the goal as one of minimizing:

$$E\left[\sum_{k=0}^{N-1}\left(r\left(X_k^\Psi\right) + c\nu_k(X_0^\Psi, \ldots, X_k^\Psi)\right) + R\left(X_N^\Psi\right)\right]$$

over all control strategies Ψ.

Example: Queueing

Consider a finite buffer queue with packet capacity n. A random number of packets arrive between each time instant. At most, one packet is released at each instant. The buffer has the choice of fast, expensive service or slow, cheap service. Under fast service, the buffer is more likely to release a packet.

x_k denotes the number of packets in the buffer at time $k \in 0, 1, ..., n$. The states evolve as a controlled Markov chain. The function $\rho(j)$ is a probability distribution giving the odds of $j \in \mathcal{N}$ packet arrivals (where $\mathcal{N}$ denotes $\{0, 1, 2, \ldots\}$. The fast choice u_f and slow choice u_s each provide a different transition matrix, $p(u_f)$ and $p(u_s)$, because they have different packet release probabilities q_f and q_s. For instance:

$$\begin{array}{rcll} p_{0,j}(u_f) & = & \rho(j) & j = 0, 1, ..., n-1 \\ p_{0,j}(u_s) & = & \rho(j) & j = 0, 1, ..., n-1 \end{array}$$

$$p_{0,n}(u_f) = \sum_{j=n}^{\infty} \rho(j)$$

$$p_{0,n}(u_s) = \sum_{j=n}^{\infty} \rho(j)$$

These are the cases when no packets are in the buffer. If there is one or more packets already present, we can compute, for instance:

$$\begin{array}{rcl}
\text{For } i > 0, j < n-1: & & \\
p_{i,j}(u_f) & = & 0 \quad \text{if } j < i-1 \\
p_{i,i-1}(u_f) & = & q_f \rho(0) \\
p_{i,j}(u_f) & = & q_f \rho(j-i+1) + (1-q_f)\rho(j-1) \\
p_{i,j}(u_s) & = & 0 \quad \text{if } j < i-1 \\
p_{i,i-1}(u_s) & = & q_s \rho(0) \\
p_{i,j}(u_s) & = & q_s \rho(j-i+1) + (1-q_f)\rho(j-1)
\end{array}$$

Please see example 1.1.4 of the text for the cases where $j = n-1$ and $j = n$.

We informally pose the strategy optimization problem:

$$\text{minimize} \quad E\left[\sum_{k=0}^{N-1} (r(X_k) + c(U_k)) + R(X_N)\right]$$

$$c(u_f) > c(u_s)$$

representing the one step cost of using either of the strategies, and

$$\begin{array}{rcl}
r(i) & = & \text{cost of holding a buffer of size i during an arbitrary stage of operation} \\
R(i) & = & \text{terminal cost of a buffer of size i at end of operation}
\end{array}$$