EE365 Stochastic Control / MS&E251 Stochastic Decision Models
Profs. S. Lall, S. Boyd
June 5–6 or June 6–7, 2013

Final exam solutions

This is a 24 hour take-home final. Please turn it in to one of the TAs, at Bytes Cafe in the Packard building, 24 hours after you pick it up.

You may use any books, notes, or computer programs (*e.g.*, Matlab), but you may not discuss the exam with anyone until June 9, after everyone has taken the exam. The only exception is that you can ask us for clarification, via the course staff email address. We've tried pretty hard to make the exam unambiguous and clear, so we're unlikely to say much.

Please make a copy of your exam before handing it in.

Please attach the cover page to the front of your exam. Assemble your solutions in order (problem 1, problem 2, problem 3, . . .), starting a new page for each problem. Put everything associated with each problem (*e.g.*, text, code, plots) together; do not attach code or plots at the end of the final.

We will deduct points from long needlessly complex solutions, even if they are correct. Our solutions are not long, so if you find that your solution to a problem goes on and on for many pages, you should try to figure out a simpler one. We expect neat, legible exams from everyone, including those enrolled Cr/N.

When a problem involves computation you must give all of the following: a clear discussion and justification of exactly what you did, the Matlab (or other) source code that produces the result, and the final numerical results or plots. To download Matlab files containing problem data, you'll have to type the whole URL given in the problem into your browser; there are no links on the course web page pointing to these files. To get a file called `filename.m`, for example, you would retrieve

    http://www.stanford.edu/class/ee365/data_for_final/filename.m

with your browser.

All problems have equal weight.

Be sure to check your email often during the exam, just in case we need to send out an important announcement.

1. *Optimal investment in a startup.* Let v_t denote the valuation of a start-up company at time t, $t = 0, 1, \ldots$ (say, in months). If $v_t = 0$, then the company goes bankrupt, and stops operating; if $v_t = v^{\max}$, then the company is acquired by a larger company, you receive a payout $v^{\max}$, and the company stops operating. In each time period that the company operates, you incur an operating cost c_o, and you decide whether to invest more money in the company, depending on its current value. If you decide to invest, then you invest a fixed amount c_i. We model the valuation as a Markov decision process: the states $v_t = 0$ and $v_t = v_{\max}$ are absorbing; if $0 < v_t < v^{\max}$ and you invest at time t, then

$$v_{t+1} = \begin{cases} v_t + \delta & \text{with probability } p_1, \\ v_t - \delta & \text{with probability } 1 - p_1; \end{cases}$$

if $0 < v_t < v^{\max}$ and you do not invest at time t, then

$$v_{t+1} = \begin{cases} v_t + \delta & \text{with probability } p_0, \\ v_t - \delta & \text{with probability } 1 - p_0. \end{cases}$$

Here $\delta > 0$ is a given parameter. The initial valuation v_0 is an integer multiple of δ, as is $v^{\max}$, so all v_t are also integer multiples of δ. With this model, you will eventually either go bankrupt or be acquired, whether you make investments or not.

(a) Explain how to find an investment policy that maximizes your expected profit. (Profit is the payout, when and if the company is acquired, minus the total operating cost, minus the total of any investments made.) And yes, we mean over infinite time, although any given realization will terminate in bankruptcy or acquisition in a finite number of periods.

(b) Consider the instance of the problem with

$$v_0 = \$10\text{M}, \quad v^{\max} = \$100\text{M}, \quad c_o = \$10\text{K}, \quad c_i = \$400\text{K},$$
$$p_1 = 0.60, \quad p_0 = 0.50, \quad \delta = \$2\text{M}.$$

What is the optimal investment policy? Report the expected profit, the probability that the startup goes bankrupt, and the expected time until the startup goes bankrupt or is acquired, all under the optimal policy. Use Monte Carlo simulation with the optimal policy to give a histogram of the profit. Give 10 trajectories of valuation on the same plot.

2. *Opportunistic wireless transmission.* A wireless transmission link consists of a queue that stores data to be transmitted (sent), and a radio transmitter that transmits (sends) data to the receiver. We will measure data in (integer) units of some standard (fixed size) packet. In each time period (called a *time slot*), we start with $q_t \geq 0$ packets in the queue. Then, $a_t \geq 0$ new packets arrive, so we have $q_t + a_t$ packets. After the new packets arrive, we transmit s_t packets to the receiver, where $0 \leq s_t \leq q_t + a_t$, Thus, there are $q_{t+1} = q_t + a_t - s_t$ packets in the queue at the beginning of the next time period. We also require that $s_t \geq q_t + a_t - Q$, where $Q > 0$ is the queue capacity; this ensures that $q_{t+1} \leq Q$, so we never exceed the queue capacity. We model the packet arrivals, a_t, as IID random variables with a known distribution.

We use the queue length, q_t, as a measure of how well the wireless link performs, with smaller values being better than larger values. (One justification for using this metric is that the average queue length is related to the average queuing delay for a packet.) In particular, we assess a queue storage cost $c_t = \alpha q_t + \beta q_t^2$, where α and β are nonnegative parameters.

In each period we can choose s_t, the number of packets to send. Sending s_t packets requires a transmitter power

$$p_t = \eta n_t (e^{s_t/\gamma} - 1),$$

where η and γ are known positive constants, and $n_t > 0$ (which can be a real number, not just an integer) is the wireless channel noise (plus interference) power during time slot t. (This formula is derived from the capacity of the wireless channel, which is proportional to $\log(1 + \eta p_t / n_t)$, but you don't need to know this to solve the problem.) The noise power n_t is modeled as a sequence of IID random variables with a known distribution. The number of packets to transmit, s_t, is chosen *after* the channel noise power, n_t, and the arrivals, a_t, are revealed. Thus, the number of packets to transmit is chosen as a function of the queue level, arrivals, and the channel noise power: $s_t = \mu(q_t, a_t, n_t)$. This is called the *transmission policy.*

The goal is to choose the transmission policy to minimize the sum of the average transmitter power p_t and the average queue cost c_t.

(a) Explain how to find the optimal transmission policy. You can assume that no pathologies occur in the DP iteration.

(b) Find the optimal transmission policy for the problem with data

$$\alpha = 0.05, \quad \beta = 0.01, \quad \gamma = 100, \quad \eta = 500, \quad Q = 20.$$

Assume n takes the values $(0.1, 1.0, 2.0, 3.0)$ with probabilities $(0.1, 0.4, 0.4, 0.1)$, and a takes values $(0, 1, 2, 3, 4, 5)$ with probabilities $(0.2, 0.3, 0.2, 0.1, 0.1, 0.1)$. Report the optimal average power and the optimal average queue cost. Give a time trace of a sample realization showing the channel noise, n_t, the number of transmitted packets, s_t, and the queue level q_t. (Your trace should start after the closed-loop system has reached statistical equilibrium.)

3. *Appliance scheduling with fluctuating real-time prices.* An appliance has C cycles, $c = 1, \ldots, C$, that must be run, in order, in $T \geq C$ time periods, $t = 0, \ldots, T - 1$. A schedule consists of a sequence $0 \leq t_1 < \cdots < t_C \leq T - 1$, where t_c is the time period in which cycle c is run. Each cycle c uses a (known) amount of energy $e_c > 0$, $c = 1, \ldots, C$, and, in each period t, there is an energy price p_t. The total energy cost is then $J = \sum_{c=1}^{C} e_c p_{t_c}$. In the lecture on deterministic finite-state control, we considered an example of this type of problem, where the prices are known ahead of time. Here, however, we assume that the prices are independent log-normal random variables, with known means, $\bar{p}_t$, and variances, σ_t^2, $t = 0, \ldots, T - 1$. You can think of $\bar{p}_t$ as the predicted energy price (say, from historical data), and p_t as the actual realized real-time energy price.

The following questions pertain to the specific problem instance defined in `appliance_sched_data.m`.

(a) *Minimum mean cost schedule.* Find the schedule that minimizes $\mathbf{E}\, J$. Give the optimal value of $\mathbf{E}\, J$, and show a histogram of J (using Monte Carlo simulation). Here you do not know the real-time prices; you only know their distributions.

(b) *Optimal policy with real-time prices.* Now suppose that right before each time period t, you are told the real-time price p_t, and then you can choose whether or not to run the next cycle in time period t. (If you have already run all cycles, there is nothing you can do.) Find the optimal policy, $\mu^\star$. Find the optimal value of $\mathbf{E}\, J$, and compare it to the value found in part (a). Give a histogram of J.

You may use Monte Carlo (or simple numerical integration) to evaluate any integrals that appear in your calculations. For simulations, the following facts will be helpful: If $z \sim \mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2)$, then $w = \exp z$ is log-normal with mean μ and variance σ^2 given by

$$\mu = e^{\tilde{\mu} + \tilde{\sigma}^2/2}, \qquad \sigma^2 = \left(e^{\tilde{\sigma}^2} - 1 \right) e^{2\tilde{\mu} + \tilde{\sigma}^2}.$$

We can solve these equations for

$$\tilde{\mu} = \log\left(\frac{\mu^2}{\sqrt{\mu^2 + \sigma^2}} \right), \qquad \tilde{\sigma}^2 = \log(1 + \sigma^2/\mu^2).$$

4. *Linear quadratic regulator with random actuator availability.* Consider the discrete-time linear dynamical system

$$x_{t+1} = Ax_t + Bu_t + w_t, \quad t = 0, 1, \ldots,$$

where $x_t \in \mathbf{R}^n$ and $u_t \in \mathbf{R}^m$. We assume that the $w_t \in \mathbf{R}^n$ are IID with $\mathbf{E}\,w_t = 0$ and $\mathbf{E}\,w_t w_t^T = W$. The stage cost is $(1/2)(x^T Q x + u^T R u)$, where $Q \geq 0$ and $R > 0$. The twist in this problem is that, in each period, you are told if the actuator is available for use. The actuator being unavailable for use in period t is equivalent to requiring that $u_t = 0$; if the actuator is available for use in period t, then u_t is unconstrained. The actuator availability is random, and modeled as follows. Let $a_t \in \{0, 1\}$ be IID random variables with $\mathbf{Prob}(a_t = 1) = p$. Additionally, assume that the a_t are independent of the w_t. When $a_t = 1$, the actuator is available for use; $a_t = 0$ means it is not.

The information pattern is this: In each period t, you know the state x_t, and you know a_t (*i.e.*, whether or not you can use the actuator), but you do not know w_t. When $a_t = 1$, you can choose $u_t = \mu^{\mathrm{av}}(x_t)$, where $\mu^{\mathrm{av}} : \mathbf{R}^n \to \mathbf{R}^m$ is the policy when the actuator is available. When $a_t = 0$, we have $u_t = 0$.

The goal is to find a μ^{av} that minimizes the average stage cost. You may invoke the ITAP assumption; that is, you can assume that no pathologies occur. (You may not, however, assume that any miracles occur.)

(a) Explain how to find μ^{av}. Give its (parametric) form, and explain how to find its parameters (possibly in the limit of an iteration).

The information pattern is *not* one of the ones we have seen in the lectures, so you will have to come up with your own variation on the traditional DP iteration. You don't have to prove that your DP method leads to an optimal policy, or even derive it; it is enough to clearly describe it.

(b) Carry out your method on the problem instance with data

$$A = \begin{bmatrix} 0.3 & 0.6 & 0.1 \\ 0.6 & 0.2 & 0.2 \\ 0.5 & 0.5 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 0.6 \\ 0.2 \\ 0.1 \end{bmatrix}, \qquad W = I,$$

$Q = I$, $R = 1$, and $p = 0.2$ (so $n = 3$ and $m = 1$). Give the optimal μ^{av}, and the optimal average stage cost. Perform a Monte Carlo simulation of the closed-loop system. Plot a_t, $\|x_t\|_2$, and u_t versus t, for some range of t after the closed-loop system has come to statistical equilibrium, and estimate the average stage cost from your simulation. (You might start with $x_0 = 0$, simulate for 100 time steps, then plot the next 100 time steps. To estimate the average stage cost, you can compute the average cost over 10000 steps.)

5. *Absorbing Markov chains.* This problem concerns the specific Markov chain $x_0, x_1, \ldots$ with transition matrix

$$
P = \begin{bmatrix}
0.1 & 0 & 0.2 & 0.7 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.5 & 0 & 0 & 0.4 & 0 & 0 & 0 & 0.1 & 0 \\
0.3 & 0 & 0.3 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.6 & 0 & 0.1 & 0.3 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.4 & 0 & 0 & 0.1 & 0 & 0 & 0 & 0.5 & 0 \\
0.2 & 0.2 & 0 & 0 & 0 & 0.2 & 0.2 & 0.2 & 0 & 0 \\
0 & 0 & 0.1 & 0.1 & 0.1 & 0.2 & 0.2 & 0.1 & 0.1 & 0.1 \\
0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\
0 & 0.3 & 0 & 0 & 0.3 & 0 & 0 & 0 & 0.4 & 0 \\
0.4 & 0 & 0 & 0 & 0 & 0.6 & 0 & 0 & 0 & 0
\end{bmatrix}.
$$

The matrix P is defined in `absorbing_markov_data.m`.

(a) Find the communicating classes. For each class, give a list of the states in the class, and say whether the class is transient or closed.

(b) Find $\lim_{t \to \infty} P^t$. Use the symbol '?' to denote an entry that does not converge.

(c) Find $\sum_{t=0}^{\infty} P^t$.

(d) Suppose the initial state is $x_0 = 1$. Find the steady state distribution $\lim_{t \to \infty} \pi_t$, where π_t is the distribution of x_t.

(e) Let A be the closed class containing state 1, and let B be the closed class containing state 2. The state is eventually absorbed in one of these classes. For each state i, find the probability that the state is absorbed in class A if $x_0 = i$.

(f) Suppose we are charged a cost of 10 if the state is absorbed in class A, and a cost of 20 if the state is absorbed in class B. For each state i, find the expected cost at absorption if $x_0 = i$.