ORF 544
Stochastic Optimization and Learning

Takehome final exam
Spring, 2019

---

a) This is an open book, take-home midterm. You have four days to finish the exam. Turn the exam in to Kim Lupinacci in Room 120 in Sherrerd Hall by 4pm.

b) Under no circumstances are you to discuss the midterm with *anyone*.

c) The questions below take you on a tour through the course textbook *Stochastic Optimization and Learning*. Each question is numbered xx.yy, where xx indicates the chapter from which the question is derived.

8.1 (5 points) What is the distinguishing characteristic of a state-dependent *problem*, as opposed to the state-independent problems we considered in chapters 5 and 7? Why do we make the distinction between the two problem classes, since both can still be modeled as dynamic programs?

8.2 (20 points) Below is a series of variants of our familiar newsvendor (or inventory) problem. In each, describe the pre- and post-decision states, decision and exogenous information in the form:

$$(S_0, x_0, S_0^x, W_1, S_1, x_1, S_1^x, W_2, \ldots)$$

Specify S_t, S_t^x, x_t and W_t in terms of the variables of the problem.

a) (5 points) The basic newsvendor problem where we wish to find x that solves

$$\max_x \mathbb{E}\{p \min(x, \hat{D}) - cx\} \tag{22.1}$$

where the distribution of $\hat{D}$ is unknown.

b) (3 points) The same as (a), but now we are given a price p_t at time t and asked to solve (22.1) using this information. Note that p_t is unrelated to any prior history or decisions.

c) (3 points) Repeat (b), but now $p_{t+1} = p_t + \hat{p}_{t+1}$.

d) (3 points) Repeat (c), but now leftover inventory is held to the next time period.

e) (3 points) Of the problems above, which (if any) are *not* dynamic programs? Explain.

f) (3 points) Of the problems above, which would be classified as solving state-dependent vs. state-independent functions.

9.1 (5 points) What are the five elements of a sequential decision problem?

9.2 (5 points) Two definitions are given of a state variable. Explain the difference in the two settings.

9.3 (5 points) Explain the statement *Every properly modeled problem is Markovian.* I will give 25 points to anyone who can show a counterexample (remember: you cannot simply leave information out of the state variable, since this would be an example of a problem that is not being properly modeled).

9.4 Consider the problem of controlling the amount of cash a mutual fund keeps on hand. Let R_t be the cash on hand at time t. Let $\hat{R}_{t+1}$ be the net deposits (if $\hat{R}_{t+1} > 0$) or withdrawals (if $\hat{R}_{t+1} < 0$), where we assume that $\hat{R}_{t+1}$ is independent of $\hat{R}_t$. Let M_t be the stock market index at time t, where the evolution of the stock market is given by $M_{t+1} = M_t + \hat{M}_{t+1}$ where $\hat{M}_{t+1}$ is independent of M_t. Let x_t be the amount of money moved from the stock market into cash ($x_t > 0$) or from cash into the stock market ($x_t < 0$).

a) (10 points) Give a complete model of the problem, including both pre-decision and post-decision state variables.

b) (5 points) Suggest a simple parametric policy function approximation, and give the objective function as an online learning problem.

9.5 (20 points) In this exercise you are going to model an energy storage problem, which is a problem class that arises in many settings (how much cash to keep on hand, how much inventory on a store shelf, how many units of blood to hold, how many milligrams of a drug to keep in a pharmacy, ...). We will begin by describing the problem in English with a smattering of notation. Your job will be to develop it into a formal dynamic model.

Our problem is to decide how much energy to purchase from the electric power grid at a price p_t. Let x_t^{gs} be the amount of power we buy (if $x^{gs} > 0$) or sell (if $x^{gs} < 0$). We then have to decide how much energy to move from storage to meet the demand D_t in a commercial building, where $x_t^{sb} \geq 0$ is the amount we move to the building to meet the demand D_t. Unsatisfied demand is penalized at a price c per unit of energy.

Assume that prices evolve according to a time-series model given by

$$p_{t+1} = \theta_0 p_t + \theta_1 p_{t-1} + \theta_2 p_{t-2} + \varepsilon_{t+1}, \tag{22.2}$$

where ε_{t+1} is a random variable with mean 0 that is independent of the price process. We do not know the coefficients θ_i for $i = 0, 1, 2$, so instead we use estimates $\bar{\theta}_{ti}$. As we observe p_{t+1}, we can update the vector $\bar{\theta}_t$ using the recursive formulas for updating linear models as described in chapter **??**, section 3.8 (you will need to review this section to answer parts of this question).

Every time period we are given a forecast $f_{tt'}^D$ of the demand $D_{t'}$ at time t' in the future, where $t' = t, t+1, \ldots, t+H$. We can think of $f_{tt}^D = D_t$ as the actual demand. We can also think of the forecasts $f_{t+1,t'}^D$ as the "new information" or define a "change in the forecast" $\hat{f}_{t+1,t'}^D$ in which case we would write

$$f_{t+1,t'}^D = f_{tt'}^D + \hat{f}_{t+1,t'}^D.$$

a) What are the elements of the state variable S_t (we suggest filling in the other elements of the model to help identify the information needed in S_t). Define both the pre- and post-decision states.

b) What are the elements of the decision variable x_t? What are the constraints (these are the equations that describe the limits on the decisions). Finally introduce a function $X^\pi(S_t)$ which will be our policy for making decisions to be designed later (but we need it in the objective function below).

c) What are the elements of the exogenous information variable W_{t+1} that become known at time $t+1$ but which were not known at time t.

d) Write out the transition function $S_{t+1} = S^M(S_t, x_t, W_{t+1})$, which is the equations that describe how each element of the state variable S_t evolves over time. There needs to be one equation for each state variable.

e) Write out the objective function by writing:

The contribution function $C(S_t, x_t)$.

The objective function where you maximize expected profits over some general set of policies (to be defined later - not in this exercise).

10.1 (3 points for each uncertainty class) Pick a sequential decision problem of your choosing. Provide a brief explanation, and then list all the types of uncertainty that might arise in this setting. You will get more points if you pick a richer problem.

11.1 Consider two policies:

$$X^{\pi^A}(S_t|\theta) = \arg\max_{x_t} \left(C(S_t, x_t) + \sum_{f\in\mathcal{F}} \theta_f \phi_f(S_t) \right), \tag{22.3}$$

and

$$X^{\pi^B}(S_t|\theta) = \arg\max_{x_t} \left(C(S_t, x_t) + \sum_{f\in\mathcal{F}} \theta_f \phi_f(S_t) \right). \tag{22.4}$$

In the case of the policy π^A in equation (22.3), we search for the parameter vector θ by solving

$$\max_{\theta} \mathbb{E} \sum_{t=0}^{T} C(S_t, X^{\pi^A}(S_t|\theta)). \tag{22.5}$$

In the case of policy π^B, we wish to find θ so that

$$\sum_{f\in\mathcal{F}} \theta_f \phi_f(S_t) \approx \mathbb{E} \sum_{t'=t}^{T} C(S_t, X^{\pi^B}(S_t|\theta)). \tag{22.6}$$

a) (5 points) Classify policies π^A and π^B among the four classes of policies.

b) (5 points) Which of the two policies π^A and π^B should work best? Explain.

11.2 Below is a list of problems with a proposed method for making decisions. Classify each method based on the four classes of policies (you may decide that a method is a hybrid of more than one class).

a) (3 points) You use Google maps to find the best path to your destination.

b) (3 points) You are managing a shuttle service between the mainland and a small resort island. You decide to dispatch the shuttle as soon as you reach a minimum number of people, or when the wait time of the first person to board exceeds a particular amount.

c) (3 points) An airline optimizes its schedule over a month using schedule slack to protect against potential delays.

d) (3 points) Upper confidence bounding policies for performing sequential learning (these were introduced in chapter 7).

e) (3 points) A computer program for playing chess using a point system to evaluate the value of each piece that has not yet been captured. Assume it chooses the move that leaves it with the highest number of points after one move.

f) (3 points) Imagine an improved computer program that enumerates all possible chess moves after three moves, and then applies its point system.

g) (3 points) Thompson sampling for sequential learning (also introduced in chapter 7).

12.1 (10 points) What is an affine policy? Give an example, and set up the general objective function for finding the best affine policy for a particular model.

12.2 (3 points) What is meant by a monotone policy?

12.3 (10 points) Consider a problem of managing water in a reservoir where the water level R_t evolves according to

$$R_{t+1} = \max\{0, R_t - x_t + \hat{R}_{t+1}\}$$

where $\hat{R}_{t+1}$ represents exogenous input (rainfall) between t and $t+1$. Assume your control is given by

$$X^\pi(S_t|\theta) = \theta_0 + \theta_1 R_t + \theta_2 R_t^2.$$

Analytically fill in as much of equations (12.15) and (12.16) as you can given this information. Your goal is to try to find the gradient of the objective function with respect to the policy parameter vector θ.

14.1 (5 points) Give the relationship between the one-step transition matrix and the transition function.

14.2 (5 points) Approximate value iteration (which includes Q-learning) is probably the most widely used strategy in approximate dynamic programming (this is the original form of reinforcement learning).

a) Write out the basic equations for performing approximate value iteration when using a lookup table architecture for the value function and a pure forward pass learning process.

b) A stepsize of $1/n$ is known to produce a convergent learning algorithm if we use appropriate exploration policies. What is the problem with a $1/n$ stepsize rule?

c) Write out the basic equation if we use a two-pass algorithm. What role is the value function approximation playing here?

d) How does the use of a two-pass learning process change your thoughts toward the choice of stepsize.

14.3 (5 points) Illustrate how each of the three curses of dimensionality can arise when computing a one-step transition matrix?

20.1 (5 points) What are the five types of approximations that are made when creating an approximate lookahead model?

20.2 (5 points) What approximations are being made when using two-stage stochastic programs as the basis of a policy? If we can solve the two-stage stochastic program optimally (this is hard for some large problems), is the resulting policy optimal? Briefly explain.

20.3 (5 points) What is meant by a "scenario tree." Sketch an example of a scenario tree. If we model a multistage problem with 10 time periods, and where there are 5 different outcomes in each time period, how many sample paths would we be representing in our scenario tree?