Part I. Dynamic Programming

The first six lectures have covered deterministic and stochastic dynamic programming over both finite and infinite horizons. The work has been entirely in discrete time, and mostly in the state-structured Markov case. (Continuous time problems are addressed in the third part of the course.) Good books for this part of the course are Ross, *Introduction to Stochastic Dynamic Programming*, Chapters I–V, and Bertsimas, *Dynamic Programming and Optimal Control*, Volume I, Chapters 1 and 7 and Volume II, Chapters 1, 3 and 4.

As a result of studying this material you should be able to:

- explain the meaning of the key terms used in lectures and listed overleaf.
- make use of the standard notation summarised overleaf.
- recall illustrative examples of various problem types (e.g., optimal gambling, asset selling, parking, queueing control, etc.)
- know the following results and conditions under which they hold:
  - Theorem 1.1 validity of the finite-horizon optimality equation.
  - Theorem 3.1 validity of the infinite-horizon optimality equation in D, P and N cases.
  - Theorem 4.1 sufficient condition for a policy to be optimal in D and P cases.
  - Section 4.4 efficacity of value iteration.
  - Theorem 5.1 sufficient condition for a policy to be optimal in D and N cases.
  - Theorems 5.2, 5.4 optimality of one-step-look-ahead rules for stopping problems.
  - Theorems 6.1, 6.2 sufficient condition for a policy to be optimal in the average-cost case.
  - Section 6.4 efficacity of policy improvement.
- write down optimality equations for dynamic programming problems that are similar to the examples in lectures and on Examples Sheet 1, and solve for optimal policies by methods of backward recursion, interchange arguments, applications of Theorems 4.1, 5.1, 6.2, and policy improvement.
- construct proofs based upon the following ideas:
  - Definition, (e.g., $F(x) \leq F(\pi, x)$),
  - Dominance, (e.g., $\mathbf{C}_h(x_h) = 0$ and case N, then $F_s(\pi, x) \leq F_{s+1}(\pi, x)$ and $F_s(x) \leq F(x)$),
  - Limits, (e.g., $\mathbf{C}_h(x_h) = 0$, then $F(\pi, x) = \lim_{s\to\infty} F_s(\pi, x)$ exists, by monotone convergence in cases P, N, and by $|F_s(\pi, x) - F_t(\pi, x)| \leq \beta^t B/(1-\beta)$, $s \geq t$, in case D),
  - Repeated Substitution, of an optimality equation for $F(x)$, or a recursion for $F(\pi, x)$, into itself.

Key terms in Lectures 1–6

Notation in Lectures 1–6

x	state variable (x_t denotes its value at time t)
u	control variable (u_t denotes its value at time t)
t	time, ($t = 0, 1, \ldots$)
h	horizon (cost is over $t = 0, \ldots, h$)
s	time to go, $s = h - t$, (but also simply used as alternate time variable)
X_t	state history $(x_0, \ldots, x_t)$
U_{t-1}	control history $(u_0, \ldots, u_{t-1})$
$W_t = (X_t, U_{t-1})$	history available at the point u_t is chosen
β	discount factor
$c(x_t, u_t, t)$	one step cost
$c(x_t, u_t)$	time-homogeneous one step cost
$r(x_t, u_t)$	time-homogeneous one step reward
$C_h(x_h)$	terminal cost
D	discounted case: $\lvert c(x, u)\rvert < B$ and $0 < \beta < 1$
N	negative case: $c(x, u) \geq 0$ and $0 \leq \beta$, usually $\beta = 1$
P	positive case: $c(x, u) \leq 0$ and $0 \leq \beta$, usually $\beta = 1$
π	policy
$E_\pi[\cdots]$	expectation of $[\cdots]$ when system evolves under policy π
$F(\pi, x, t)$	cost over $t, \ldots, h$ under π, i.e., $E_\pi[\sum_{s=t}^{h-1} \beta^{s-t} c(x_s, u_s, s) + C_h(x_h) \mid x_t = x]$
$F(x, t)$	infimal cost over $t, \ldots, h$, i.e., $\inf_\pi F(\pi, x, t)$
$F_s(\pi, x)$	cost under π over s remaining steps (time-homogeneous case)
$F_s(x)$	infimal cost over s remaining steps, i.e., $\inf_\pi F(\pi, x)$
$F_\infty(x)$	$\lim_{s\to\infty} F_s(x)$ (assuming this exists)
$F(\pi, x)$	cost under π over infinite horizon, i.e., $\lim_{s\to\infty} E_\pi[\sum_{t=0}^{s-1} \beta^t c(x_t, u_t) \mid x_0 = x]$
$F(x)$	infimal cost over infinite horizon (n.b., $F(x) = F_\infty(x)$ if value interation works)
$\pi = (f_0, f_1, \ldots)$	a Markov policy, i.e., $u_t = f_t(x_t)$
$\pi = f^\infty$	a stationary Markov policy, i.e., $\pi = (f, f, \ldots)$