Lecture Notes 4: Dynamic Programming*

Ivan Rudik

ECON 509: Computational Methods

February 12, 2016

1 The Building Blocks of a Dynamic Economic Model

In this section (and all future sections), it's assumed that you have some background in dynamics. Dynamic models are a central focus of numerical methods. By accounting for intertemporal behavior, simple problems can become analytically intractable. First we want to outline the key features of all dynamic models:

1. **Controls:** What variables are we optimizing over? What decision is the economic agent making? In an industrial organization example, a control may be the level of investment into research and development or entry into a new market. We wish to understand how the optimal policy changes as we move forward in time, as the state of the world changes, or how it is a function of model parameters or initial conditions.

2. **States:** What are the variables that change over time and interact with the agents decision-making in each period? If our control is research and development investment, state variables may include the current level of technology, and the level of technology of rival firms. If the effectiveness of R&D investment is endogenous (e.g. learning-by-doing), then we will also want to keep track of this effectiveness as a state. The state variables in a period completely characterize the economic problem. We will find that the number states completely determines the *dimensionality* of the problem, which governs its tractability. In both analytics

*This lecture note borrows from those of Derek Lemoine.

and numerics we will want to keep the number of states to a minimum while capturing all critical features.

3. **Payoff:** What is the single-period payoff function? This function tells us our immediate reward for a given decision, while the *continuation value* (defined in more detail later), tells us the value of the optimal policy trajectory conditional on that decision. The continuation value is composed of a sum of future one period pay off functions, evaluated at their optimal value. In our R&D example, the immediate payoff to investment may be just the cost, while the continuation value captures how this investment alters our future costs or revenues of selling goods.

4. **Equations of motion / transition equations:** How do the state variables evolve over time? Virtually all problems have the controls affecting how states transition, so that the current control can also affect future states and the continuation value (future payoffs). Investing in R&D does not just purely incur a cost, it alters our probability of having a large breakthrough in future product development.

5. **Planning horizon:** When does our problem terminate? Do we make an infinite sequence of decisions or is there only a finite number of periods? If it terminates, then the problem can be solved via backwards induction: find the optimal policy in the final period (as a function of our state variables). Then use the payoff associated with this optimal policy as our continuation value in our second-to-last period and find the optimal policy. Continue the process until you've reached the first period. We will be focusing on *autonomous problems* in which time does not directly enter payoffs or transitions, but only affects the problem through discounting.

1.1 Two Classes of Solutions

Dynamic problems can be solved in two ways. In one way we can treat the model as a sequence of static problems that are solved by simultaneous optimization. This is just one large non-linear optimization problem so we can just send it to our solvers for the solution, called an *open loop solution.* The big drawback is that we can only generate solutions that are a function of time, and not of the state variables. This causes problems if we want to include uncertainty, stochastic shocks, or strategic behavior between agents. The second way to solve dynamic problems is with a *feedback solution.* In the feedback case, we represent the model as a single-period optimization involving the immediate payoff and the continuation value. This yields a solution that is a function of the state so that if we have an unexpected transition from a stochastic shock, we can still find optimal policies in the next period. The optimization of the feedback problem is not difficult because we only account for one period's worth of controls, but we still must find a way to recover the continuation value. This is the hard part.

2 Markov Chains

A stochastic process is a sequence of random vectors. States in the feedback representation of a dynamic programming problem follow stochastic processes. A critical feature of dynamic programming problems is that they are *Markovian*,

Markov Property: A stochastic process $\{x_t\}$ is said to have the *Markov property* if for all $k \geq 1$ and all t,

$$Prob(x_{t+1}|x_t, x_{t-1}, ..., x_{t-k}) = Prob(x_{t+1}|x_t).$$

The distribution of the next vector in the sequence (i.e. the distribution of next period's state) is a function of only the current vector (state). The Markov property is necessary for the feedback representation since we are only keeping track of the current state, and next period's state when solving the problem in this manner. We characterize our stochastic state transition process with a *Markov chain*. A Markov chain is characterized by an n-dimensional state space with vectors e_i, $i = 1, ..., n$, where e_i is an $n \times 1$ unit vector whose ith entry is 1 and all others are 0; an $n \times n$ *transition matrix* P which captures the probability of transitioning from one point of the state space to another point of the state space next period; and an $n \times 1$ vector π_0 whose ith value is the probability of being in state i at time 0: $\pi_{0i} = \text{Prob}(x_0 = e_i)$. The elements of the transition matrix P are,

$$P_{ij} = \text{Prob}(x_{t+1} = e_j | x_t = e_i).$$

We have one assumption that must be satisfied for this representation to be valid:

Assumption 1. *For $i = 1, .., n$, $\sum_{j=1}^{n} P_{ij} = 1$ and π_0 satisfies: $\sum_{i=1}^{n} \pi_{0i} = 1$.*

We can use our transition matrix to determine the probability of moving to another state in *two* periods by P^2 since,

$$\begin{aligned}
&\text{Prob}(x_{t+2} = e_j | x_t = e_i) \\
&= \sum_{h=1}^{n} \text{Prob}(x_{t+2} = e_j | x_{t+1} = e_h)\text{Prob}(x_{t+1} = e_h | x_t = e_i) \\
&= \sum_{h=1}^{n} P_{ih} P_{hj} = P_{ij}^2
\end{aligned}$$

We can iterate on this to show that,

$$\text{Prob}(x_{t+k} = e_j | x_t = e_i) = P_{ij}^k.$$

All of this holds for continuous states, except we will have transition probability distributions instead of transition probability matrices.

3 Dynamic Programming

Now we will set up the basic, recursive (feedback representation) optimization problem. We begin by characterizing a general sequential problem. Let $\beta \in (0, 1)$. The economic agent selects a sequence of controls, $\{u_t\}_{t=0}^{\infty}$ to maximize,

$$\sum_{t=0}^{\infty} \beta^t r(x_t, u_t), \tag{1}$$

subject to $x_{t+1} = g(x_t, u_t)$[1] with the initial condition x_0 given. Much like static problems, assume that the payoff function is concave and that the state space is convex and compact. We wish to recover a *policy function* h which maps the current state x_t into the current control u_t, such that the sequence $\{u_s\}_{s=0}^{\infty}$ generated by iterating,

$$\begin{aligned}
u_t &= h(x_t) \\
x_{t+1} &= g(x_t, u_t),
\end{aligned}$$

starting from x_0, solves our original optimization problem in equation (1). This type of problem is called *recursive*. To solve our recursive problem, we require knowing a second function, $V(x)$, our *continuation value function*, or just value function for short. The value function yields the optimal

[1] I.e., the problem is Markovian.

value of our original problem in equation (1) given that we are starting at state vector x. The value function is defined as,

$$V(x_0) = \max_{\{u_s\}_{s=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t r(x_t, u_t) \tag{2}$$

subject to the transition equation: $x_{t+1} = g(x_t, u_t)$. How can we possibly know what $V(x)$ is? We actually won't until we've solved the problem. For now, suppose we do know $V(x)$. Then we can compute the policy function h by solving for each $x \in X$,

$$\max_u r(x, u) + \beta V(x') \tag{3}$$

where $x' = g(x, u)$ and primes on state variables indicate next period. So conditional on having the value function $V(x)$, we can solve our dynamic programming problem. We have exchanged our original problem of solving for an infinite sequence of control that maximizes equation (1) for a problem where we must find the $V(x)$ and h that solves the *continuum* of maximization problems like equation (3), where there is a unique maximization problem for each x. This is often an *easier* problem to solve than solving for an infinite sequence of controls. And importantly, once we have obtained the value function, we can recover the optimal policy path conditional on being at any point in our state space.

How do we solve for $V(x)$ and $h(x)$? We must first define the *Bellman equation*,

$$V(x) = \max_u r(x, u) + \beta V[g(x, u)] \tag{4}$$

Our policy function $h(x)$ is what maximizes the right hand side of the Bellman. The policy function therefore satisfies,

$$V(x) = r[x, h(x)] + \beta V\{g[x, h(x)]\}$$

Equation (4) is called a functional equation since it is a function that maps functions into scalar values. Solving the problem yields a solution that is a function, $V(x)$. Moreover, it's a *recursive functional* since it maps itself into a scalar value. Under a set of standard economic assumptions,[2] we can assure that,

1. The solution to the Bellman equation is strictly concave

2. The solution is approached in the limit as $j \to \infty$ by iterations on: $V_{j+1}(x) = \max_u r(x, u) +$

[2]See Stokey and Lucas (1989) for more details.

$\beta V_j(x')$, given any bounded and continuous V_0 and our transition equation

3. There exists a unique and time-invariant optimal policy function $u_t = h(x_t)$ where h maximizes the right hand side of the Bellman

4. The value function $V(x)$ is differentiable

3.1 Euler Equations

The analog to first-order conditions for dynamic problems is the *Euler equation.* The Euler equation equalizes the marginal effects of an optimal policy over time. Often we are setting the current marginal benefit, perhaps from using energy via burning fossil fuels, with the future marginal costs, global warming. Euler equations characterize the optimal inter-temporal policy path.

Let's begin with a simple example of obtaining an Euler equation (this will be necessary for some numerical solution methods or for checking the error in other numerical methods). Consider a setting where we have a stock of capital, and can invest to increase our future capital. Our per-period payoff is the gain in utility (increase and convex) from consuming output that is perfectly linear in our current capital, $u(K)$, minus the cost (increasing and convex) of investing in future capital, $c(I)$. Our control variable is investment, our state is capital, and we have an infinite horizon. We can write our Bellman equation as,

$$V(K_t) = \max_{I_t} \{u(K_t) - c(I_t) + \beta V(K_{t+1})\}$$
$$\text{subject to:} \quad K_{t+1} = \delta K_t + \gamma I$$

where $\beta \in (0,1)$ is our per-period discount factor, $\delta \in (0,1)$ determines the depreciation of capital, and γ is the effectiveness of investment. We obtain the first-order condition by differentiating the right hand side of the Bellman equation with respect to I,

$$c'(I_t) = \beta\,\gamma\,V_K(K_{t+1}),$$

where subscripts indicate derivatives with respect to that variable. $V_K(K_{t+1})$ is the shadow value of capital: the additional value obtained from marginally increasing how much capital we have. The shadow value can be interchangeably called the co-state variable. We can then apply the envelope theorem to obtain the co-state equation:[3]

$$V_K(K_t) = u'(K_t) + \beta\,\delta\,V_K(K_{t+1}(I_t^*)).$$

[3]Recall that at the optimum, the envelope theorem must hold

Since the first-order conditions and co-state equation must always hold, regardless of which period we are in, we can advance both of them by one period,

$$c'(I_{t+1}) = \beta\,\gamma\,V_K(K_{t+2})$$
$$V_K(K_{t+1}) = u'(K_{t+1}) + \beta\,\delta\,V_K(K_{t+2}(I^*_{t+1}))$$

Now, if we substitute our time t and time $t+1$ first-order conditions into our time $t+1$ co-state equation, we will have,

$$\frac{c'(I_t)}{\beta\,\gamma} = u'(K_{t+1}) + \beta\,\delta\frac{c'(I_{t+1})}{\beta\,\gamma}$$
$$\Rightarrow c'(I_t) = \beta\left[\gamma\,u'(K_{t+1}) + \delta\,c'(I_{t+1})\right]$$

The left hand side is the marginal payoff (investment cost) today, the right hand side gives the marginal change in the discounted future payoff. Along an optimal trajectory, these must be equal. Alternatively we can think of this as a no-arbitrage condition. Suppose we're on the optimal capital path. Deviating by decreasing today's investment yields a marginal benefit today of saving us some investment cost. However there are two costs associated with it: (1) we will have lower utility tomorrow because we will have a smaller capital stock and (2) we will need to invest more tomorrow to return to the optimal capital trajectory. If this deviation (or deviating by investing more today) were profitable, we would do. Therefore the optimal policy must have zero additional profit opportunities. This is what the Euler equation defines.

Consider a second, simpler example. Suppose you are given a cake of size W. At each point in time, $t = 1, 2, ..., T$, you eat some of the cake and what you do not eat is left over for future periods. Your consumption in period t is c_t and consumption yields you utility $u(c_t)$ where u is strictly increasing, strictly concave and continuously differentiable. Also assume that $\lim_{c\to 0} u'(c) \to \infty$. You discount future periods with some discount factor $\beta \in (0, 1)$. Your lifetime utility is given by,

$$\max_{\{c_t\}_{t=1}^T} \sum_{t=1}^{T} \beta^{t-1}\,u(c_t)$$
$$\text{subject to: } W_{t+1} = W_t - c_t, W_{t+1} \geq 0$$

In Bellman form, the problem is,

$$V(W_t) = \max_{c_t} u(c_t) + \beta\,V(W_{t+1}),$$

Subject to the transition equation. The first-order condition is then

$$u'(c_t) + \beta\, V_W(W_{t+1})\frac{\partial W_{t+1}}{\partial c_t} = 0$$
$$\Rightarrow u'(c_t) = \beta\, V_W(W_{t+1})$$

The envelope theorem guarantees that our co-state equation holds, $V_W(W_t) = \beta V_W(W_{t+1})$. Using the trick of advancing both equations forward one period, we obtain the Euler equation,

$$u'(c_t) = \beta\, u'(c_{t+1}) \tag{5}$$

Marginal utility today must be equal to our discounted marginal utility tomorrow. If we instead decided to eat more today than the optimal level c_t^*, we will have lower utility tomorrow since we will need to eat less to return to the optimal trajectory. Note that the loss of this utility tomorrow is larger than the gain today since marginal utility is declining, so this decision makes us worse off.

3.2 Basic Theory

Building off the refreshed intuition from the previous examples, we continue onto more formal theory behind dynamic programming and Bellman equations. We consider only a deterministic system but, by and large, the intuition passes through to stochastic problems. Consider an infinite horizon problem,[4] for an economic agent who gains payoff $u(s_t, c_t)$ in some period t as a function of the state vector s_t and control vector c_t. States and controls are linked over time by the transition equations, $s_{t+1} = g(s_t, c_t)$, such that we can determine tomorrow's state conditional on knowing today's state and controls. The current state vector *completely* summarizes all the information of the past and is all the information the agent needs to make a forward-looking decision. I.e. our problem has the Markov property. Assume that $c \in C$ and $s \in S$, and that the payoff is bounded: $u(s_t, c_t)$.

The two final properties that are typical of dynamic economic problems are stationarity and discounting. Stationary implies the problem does not explicitly depend on time, the utility function is the same no matter the period. The rest of the problem is a function of the current state vector s_t, but t could take on any arbitrary value. This allows us to get rid of the time subscripts like we did previously and only indicate the current period and the next period. We also assume that the future is discounted at some factor $\beta \in (0, 1)$. This along with the assumption of the bounded flow payoff typically ensures that the total welfare of the problem is bounded as well. We can then

[4] Infinite horizon problems typically have more analytical tractability.

represent this payoff as,

$$\sum_{t=0}^{\infty} \beta^t u(s_t, c_t)$$

Suppose we wish to maximize this stream of payoffs. Define the value of the maximized discounted stream of payoffs as,

$$V(s_0) = \max_{c_0 \in C(s_0)} u(s_t, c_t) + \beta \left[\max_{\{c_t\}_{t=1}^{\infty}} \sum_{t=t}^{\infty} \beta^t u(s_t, c_t) \right]$$
$$\text{subject to: } s_{t+1} = g(s_t, c_t)$$

This formulation appears uglier than the previous, but it will actually allow us to reduce the problem into the familiar Bellman. Notice that the terms inside the square brackets is the maximized discounted stream of payoffs beginning at state s_1. Therefore the problem can be written recursively as,

$$V(s_0) = \max_{c_0 \in C(s_0)} u(s_t, c_t) + \beta V(s_1)$$
$$\text{subject to: } s_{t+1} = g(s_t, c_t),$$

which is our Bellman.[5] Note that we can re-formulate the problem as,

$$V(s) = \max_{s' \in \Gamma(s)} u(s, s') + \beta V(s'), \ \ \forall s \in S \tag{6}$$

Where $\Gamma(s)$ is our set of feasible states next period. There exists a solution to the Bellman under a set of sufficient conditions,[6]

Theorem 2. *Assume $u(s_t, c_t)$ is real-valued, continuous and bounded, $\beta \in (0, 1)$ and the feasible set of states for next period is non-empty, compact, and continuous. Then there exists a unique value function $V(s)$ that solves the Bellman equation.*

[5]This is just using Bellman's Principle of Optimality: our optimal policy must be optimal for each subproblem, where a subproblem is a subset of the future time periods.

[6]There exist many other sets, see Stokey and Lucas (1989) for others.

Here is an intuitive sketch of why this holds. We begin by defining an operator T as,

$$T(W)(s) = \max_{s' \in \Gamma(s)} u(s,s') + \beta\, W(s'), \;\; \forall s \in S \tag{7}$$

This operator takes some value function $W(s)$, maximizes it, and returns another $T(W)(s)$. It is easy to see that any $V(s)$ that satisfies $V(s) = T(V)(s) \forall s \in S$ solves the Bellman equation. Therefore we simply search for the fixed point of $T(W)$ to solve our dynamic problem.

How do we find the fixed point? First we must show that a way exists by showing that $T(W)$ is a contraction: as we iterate using the T operator, we will get closer and closer to the fixed point. Blackwell's sufficient conditions for a contraction are (1) monotonicity: if $W(s) \geq Q(s) \forall s \in S$, then $T(W)(s) \geq T(Q)(s) \forall s \in S$. This holds under our maximization. And (2) discounting: there exists a $\beta \in (0,1)$ such that $T(W+k)(s) \leq T(W)(s) + \beta\, k$. This just reflects that we must be discounting the future. If these two conditions hold then we have a contraction with modulus β.

Why do we care this is a contraction? Because we can take advantage of the contraction mapping theorem which states that T has a unique fixed point, $T(V^*) = V^*$, and we can start from any arbitrary initial function W, iterate using the operator and reach the fixed point. This fact lays the foundation for one of the most widely used algorithms to solve dynamic problems: value function iteration.

4 Laying Out the Model Framework

How do you actually write out a dynamic model for a reader?[7] Dynamic models are complex with many nuances and moving parts. Precisely describing all aspects of the model is paramount.

References

Judd, Kenneth L. (1998) *Numerical Methods in Economics*, Cambridge, MA: MIT Press.

Ljungqvist, Lars and Thomas J Sargent (2004) *Recursive Macroeconomic Theory*: MIT press.

Miranda, Mario J. and Paul L. Fackler (2002) *Applied Computational Economics and Finance*, Cambridge, MA: MIT Press.

Stokey, Nancy and Robert Lucas (1989) *Recursive Methods in Economic Dynamics*: Harvard University Press.

[7]This follows the notes of Leigh Tesfatsion.