ECON-GA 1025 Macroeconomic Theory I

Lecture 10

John Stachurski

Fall Semester 2018

Today's Lecture

- Quadratic optimization
- Linear quadratic optimal control
- Discrete dynamic programs

Notes

1. Congratulations to Paul Romer & NYU

   When I learned mathematical economics, a different equilibrium prevailed. Not universally, but much more so than today, when economic theorists used math to explore abstractions, it was a point of pride to do so with clarity, precision, and rigor. – Romer, AER P&P, 2015

2. Guest lecture by Tom Sargent tomorrow

Preliminary Discussion: Quadratic Optimization

A function $f : \mathbb{R}^n \to \mathbb{R}$ is called a **quadratic form** if it is a polynomial where each element has order 2

Examples.

- $f(x, y) = x^2 - y^2$ is a quadratic form
- $f(x, y) = x^2 - xy - y^2$ is a quadratic form
- $f(x, y) = x^2 - x$ is **not** a quadratic form

Fact. $f\colon \mathbb{R}^n \to \mathbb{R}$ is a quadratic form if and only if there exists a symmetric $n \times n$ matrix Q such that

$$f(x) = x'Qx \quad \text{for all} \quad x \in \mathbb{R}^n$$

Fact. If $f(x) = x'Qx$ for symmetric $n \times n$ matrix Q, then

- f is convex $\iff$ Q is positive semidefinite
- f is concave $\iff$ Q is negative semidefinite
- f is strictly convex $\iff$ Q is positive definite
- f is strictly concave $\iff$ Q is negative definite

Suppose we wish to solve

$$v(x) = \min_{u \in \mathbb{R}^m} \{u'Qu + (Ax + Bu)'P(Ax + Bu)\}$$

where

- P is symmetric, positive semidefinite and $n \times n$
- Q is symmetric, positive definite and $m \times m$
- A is $n \times n$ and B is $n \times m$

Ex. Show that $Q + B'PB$ is nonsingular

Lemma. The minimizer of v is

$$u^* := -(Q + B'PB)^{-1}B'PAx$$

and the minimized value v satisfies

$$v(x) = x'Mx$$

where

$$M := A'PA - A'PB(Q + B'PB)^{-1}B'PA$$

Ex. Confirm these claims using matrix algebra and the following two facts from matrix calculus:

$$\frac{\mathrm{d}}{\mathrm{d}u}a'u = a \quad \text{and} \quad \frac{\mathrm{d}}{\mathrm{d}u}u'Hu = (H + H')u$$

Linear Control Systems

Linear quadratic dynamic programming problems are those where

- the law of motion is linear
  - in state, control and shocks
- rewards are sums of quadratic forms
  - in state and controls

Also called

- LQ control problems
- linear regulator problems

Costs: Assumptions are restrictive

Benefits: Tractable even in very high dimesions

Examples.

- Optimal fiscal policy
- monetary policy
- energy policy
- operations research

Refs:

- "Recursive methods of dynamic linear economies." Hansen and Sargent, Princeton UP, 2013

Dynamics:

$$x_{t+1} = Ax_t + Bu_t + C\xi_{t+1}$$

Here

- x_0 given
- $\{x_t\}$ takes values in $\mathbb{R}^n$
- $\{u_t\}$ takes values in $\mathbb{R}^m$
- A and B are $n \times n$ and $n \times m$ respectively
- C is $n \times j$ and $\{\xi_t\}$ is IID with $\mathbb{E}\xi_t = 0$ and $\mathbb{E}\xi_t\xi_t' = I$

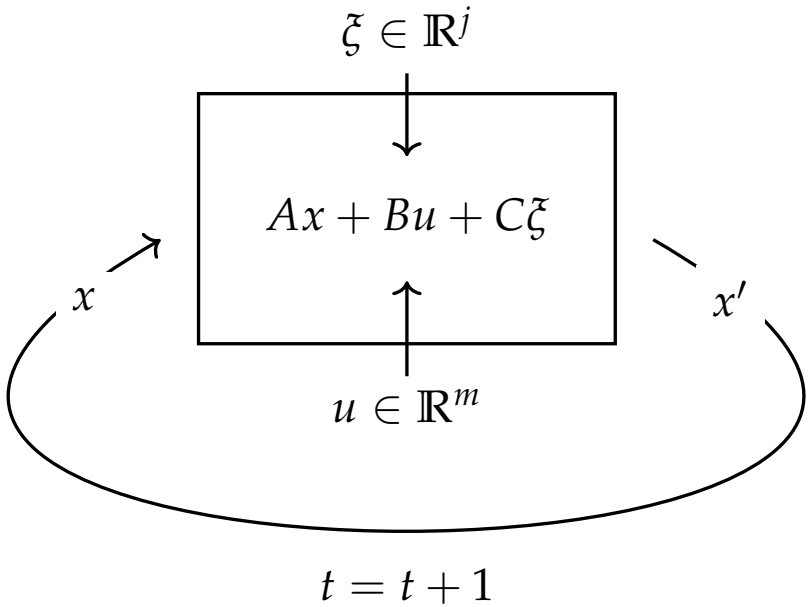

Figure: State dynamics for LQ control

Example. Consider the law of motion for wealth

$$w_{t+1} = (1+r)(w_t - c_t) + y_{t+1}$$

Assume that

$$y_t = \mu + \sigma\xi_t \text{ where } \{\xi_t\} \overset{\text{IID}}{\sim} N(0,1)$$

Can we express this as

$$x_{t+1} = Ax_t + Bu_t + C\xi_{t+1}?$$

One problem: y_t is IID but not zero mean

Step 1: Let $u_t := c_t - \bar{c}$ where $\bar{c} =$ "ideal" level of consumption

Then

$$w_{t+1} = (1+r)(w_t - u_t - \bar{c}) + \mu + \sigma\xi_{t+1}$$

is equal to the first row of

$$\begin{pmatrix} w_{t+1} \\ 1 \end{pmatrix} = \begin{pmatrix} 1+r & -(1+r)\bar{c} + \mu \\ 0 & 1 \end{pmatrix} \begin{pmatrix} w_t \\ 1 \end{pmatrix}$$

$$+ \begin{pmatrix} -(1+r) \\ 0 \end{pmatrix} u_t + \begin{pmatrix} \sigma \\ 0 \end{pmatrix} \xi_{t+1}$$

The linear specification is now complete

Set

$$x_t := \begin{pmatrix} w_t \\ 1 \end{pmatrix}, \qquad A := \begin{pmatrix} 1+r & -(1+r)\bar{c}+\mu \\ 0 & 1 \end{pmatrix},$$

$$B := \begin{pmatrix} -(1+r) \\ 0 \end{pmatrix} \quad \text{and} \quad C := \begin{pmatrix} \sigma \\ 0 \end{pmatrix}$$

Then the first row of

$$x_{t+1} = Ax_t + Bu_t + C\xi_{t+1}$$

is

$$w_{t+1} = (1+r)(w_t - u_t - \bar{c}) + \mu + \sigma\xi_{t+1}$$

In the LQ model we will aim to **minimize** a flow of **losses**

Current loss given by

$$\ell(x_t, u_t) := x_t' R x_t + u_t' Q u_t$$

Here

- R is $n \times n$, symmetric and positive semidefinite
- Q is $m \times m$, symmetric and positive definite

Example. Consider the household with

$$\text{state} = x_t = \begin{pmatrix} w_t \\ 1 \end{pmatrix}, \qquad \text{control} = u_t = c_t - \bar{c}$$

A typical choice of R and Q would be

$$Q := 1 \quad \text{and} \quad R := \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Then

$$x_t' R x_t + u_t' Q u_t = u_t^2 = (c_t - \bar{c})^2$$

- current loss for household = squared deviation of consumption from ideal

Finite Horizon Optimality

Infinite horizon problems

- technically challenging — cannot use backward induction
- but often neat because decisions are time invariant

Time matters little because agents always face an infinite future

But in some settings we specifically wish to inject time

Common **finite horizon problems** include

- life cycle savings and consumption
- retirement planning

Finite Horizon Theory

Problem: choose $u_0, \ldots, u_{T-1}$ to minimize

$$\mathbb{E}\left\{\sum_{t=0}^{T-1} \beta^t (x_t' R x_t + u_t' Q u_t) + \beta^T x_T' R_f x_T\right\}$$

subject to

$$x_{t+1} = Ax_t + Bu_t + C\xi_{t+1} \quad \text{at each } t$$

with $\beta \in (0, 1]$

- R_f is $n \times n$ and positive semidefinite
- Note $\beta = 1$ is permitted

To solve the finite horizon problem we use backwards induction

Let

$$J_T(x) := x' R_f x$$

Controller at $T-1$ facing state x_{T-1}

- takes x_{T-1} as given
- solves

$$\min_u \{x'_{T-1} R x_{T-1} + u' Q u + \beta \, \mathbb{E} \, J_T(A x_{T-1} + B u + C \xi_T)\}$$

Now let

$$J_{T-1}(x) := \min_u \{x' R x + u' Q u + \beta \, \mathbb{E} J_T(A x + B u + C \xi_T)\}$$

Consider the decision problem at $T-2$

- $J_{T-1}(x)$ gives us **minimum cost-to-go** from state x

Controller chooses u to solve

$$\min_u \{x'_{T-2} R x_{T-2} + u' Q u + \beta \, \mathbb{E} J_{T-1}(A x_{T-2} + B u + C \xi_{T-1})\}$$

Let $J_{T-2}(x)$ be the minimum cost-to-go from state x:

$$J_{T-2}(x) = \min_u \{x' R x + u' Q u + \beta \, \mathbb{E} J_{T-1}(A x + B u + C \xi_{T-1})\}$$

The pattern for backwards induction is now clear...

Calculate the **cost-to-go functions** $\{J_t\}$ recursively via

$$J_{t-1}(x) = \min_u \{x'Rx + u'Qu + \beta \mathbb{E} J_t(Ax + Bu + C\xi_t)\}$$

and

$$J_T(x) = x'R_f x$$

- a version of the **Bellman equation**
- $J_t(x)$ represents total cost-to-go from time t and state x when the controller behaves optimally

Minimizers at each stage are the **optimal controls**

Questions: Given the structure of our model,

- is there a parsimonious way to represent J_t at each t?
- is there a parsimonious way to represent the optimal choices?

Proposition. Each J_t has the form $J_t(x) = x' P_t x + d_t$, where

- The sequence $\{P_t\}$ is defined recursively by $P_T := R_f$ and

$$P_{t-1} = R - \beta^2 A' P_t B (Q + \beta B' P_t B)^{-1} B' P_t A + \beta A' P_t A$$

- The scalar sequence $\{d_t\}$ is defined recursively by $d_T = 0$ and

$$d_{t-1} = \beta(d_t + \text{trace}(C' P_t C))$$

- The optimal controls are given by

$$u_{t-1} = -F_t x_t \quad \text{where} \quad F_t := (Q + \beta B' P_t B)^{-1} \beta B' P_t A$$

Proof is by induction

The claim is true at $t = T$ with $P_T = R_f$ and $d_T = 0$

Suppose now that it holds at some $t \leqslant T$

We then have, for arbitrary $x \in \mathbb{R}^n$,

$$J_{t-1}(x) = \min_u$$

$$\{x'Rx + u'Qu + \beta\,\mathbb{E}(Ax + Bu + C\xi_t)'P_t(Ax + Bu + C\xi_t) + \beta d_t\}$$

Ex. Show that the minimizer is

$$u_{t-1} = -(Q + \beta B'P_tB)^{-1}\beta B'P_tAx$$

Ex. Show that

$$J_{t-1}(x) = x' P_{t-1} x + d_{t-1}$$

where

$$P_{t-1} = R - \beta^2 A' P_t B (Q + \beta B' P_t B)^{-1} B' P_t A + \beta A' P_t A$$

and

$$d_{t-1} = \beta (d_t + \text{trace}(C' P_t C))$$

Algorithm 1: Computing the cost-to-go in finite horizon LQ

$t \leftarrow T$;

$P_t \leftarrow R_f$;

$d_t \leftarrow 0$;

while $t > 0$ **do**

$P_{t-1} \leftarrow R - \beta^2 A' P_t B (Q + \beta B' P_t B)^{-1} B' P_t A + \beta A' P_t A$;

$d_{t-1} \leftarrow \beta (d_t + \text{trace}(C' P_t C))$;

$t \leftarrow t - 1$

end

return $\{P_t, d_t\}_{t=0}^{T}$

With

$$F_t := (Q + \beta B' P_{t+1} B)^{-1} \beta B' P_{t+1} A$$

we can simulate as follows

Algorithm 2: Simulate states and controls in finite horizon LQ

$t \leftarrow 0$;
$x_t \leftarrow$ initial condition x_0 ;
while $t < T$ **do**
    $u_t \leftarrow -F_t x_t$;
    $x_{t+1} \leftarrow Ax_t + Bu_t + C\xi_{t+1}$;
    $t \leftarrow t + 1$
end
return $\{x_t, u_t\}_{t=0}^{T-1} \cup \{x_T\}$

Example: Consumption Smoothing

Early Keynesian models assumed that households have a constant marginal propensity to consume from current income

Data contradicts this

- "Why is Consumption So Smooth?" Campbell and Deaton, REStud (1989)

Milton Friedman, Franco Modigliani and others built models based on preference for smooth consumption stream

Let's investigate an LQ version

Example. Recall the wealth dynamics

$$w_{t+1} = (1+r)(w_t - c_t) + \mu + \sigma\xi_{t+1}$$

expressed as

$$x_{t+1} = Ax_t + Bu_t + C\xi_{t+1}$$

where

$$x_t := \begin{pmatrix} w_t \\ 1 \end{pmatrix}, \qquad A := \begin{pmatrix} 1+r & -(1+r)\bar{c} + \mu \\ 0 & 1 \end{pmatrix},$$

$$u_t := c_t - \bar{c}, \quad B := \begin{pmatrix} -(1+r) \\ 0 \end{pmatrix} \quad \text{and} \quad C := \begin{pmatrix} \sigma \\ 0 \end{pmatrix}$$

The finite horizon objective is

$$\mathbb{E}\left\{\sum_{t=0}^{T-1}\beta^t(c_t-\bar{c})^2+\beta^T q w_T^2\right\}$$

where q is a large positive constant

- Why do we need it?

Ex. Pick R, Q and R_f to express this as

$$\mathbb{E}\left\{\sum_{t=0}^{T-1}\beta^t(x_t'Rx_t+u_t'Qu_t)+\beta^T x_T'R_f x_T\right\}$$

Set

- $r = 0.05$ and $\beta = 1/(1+r)$
- $\bar{c} = 2$, $\mu = 1$, $\sigma = 0.25$ and $q = 10^6$
- $T = 45$

Assume $\{\xi_t\} \overset{\text{IID}}{\sim} N(0,1)$

Ex. Complete the following tasks by computer

1. Construct the correspond matrices A, B, C, R, Q, R_f
2. Insert into the preceding algorithms
3. Solve, simulate, plot income, consumption, wealth

Figure should be similar to the next slides

Figure: Consumption and income in the life cycle problem

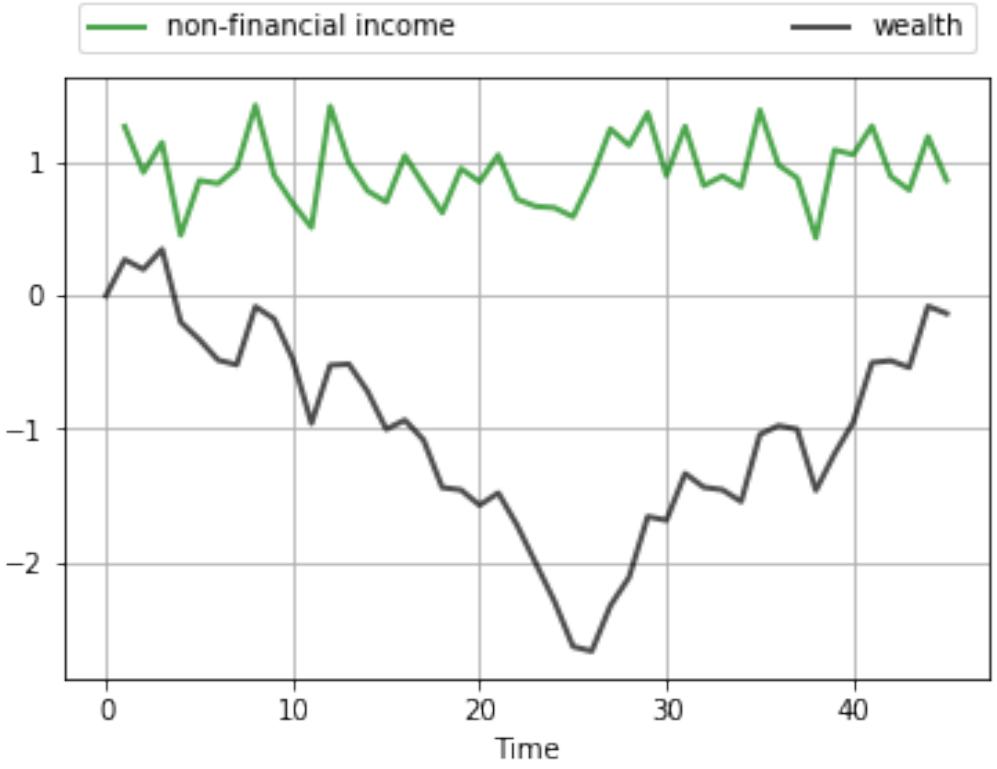

Figure: Consumption and wealth in the life cycle problem

Infinite Horizons

Unchanged dynamics, objective function

$$\mathbb{E}\left\{\sum_{t=0}^{\infty}\beta^t(x_t'Rx_t + u_t'Qu_t)\right\}$$

Time dependence in $\{P_t\}$, $\{d_t\}$ and $\{F_t\}$ are replaced by

$$P = R - (\beta B'PA)'(Q + \beta B'PB)^{-1}(\beta B'PA) + \beta A'PA$$

$$F = (Q + \beta B'PB)^{-1}(\beta B'PA)$$

and

$$d := \text{trace}(C'PC)\frac{\beta}{1-\beta}$$

The expression

$$P = R - (\beta B'PA)'(Q + \beta B'PB)^{-1}(\beta B'PA) + \beta A'PA$$

is called a **discrete time algebraic Riccati equation**

- is there a solution?
- is it unique?
- how can we compute it?

Depends on "controllability" and "observability" (see course notes)

Let

- $\mathscr{R}$ be the self-mapping on $\mathcal{M}(n \times n)$ defined by

$$\mathscr{R}(P) := R - (\beta B'PA)'(Q + \beta B'PB)^{-1}(\beta B'PA) + \beta A'PA$$

- $\mathcal{M}_P$ be the set of positive definite matricies in $\mathcal{M}(n \times n)$

Theorem. If (A, B) is controllable and (A, R) is observable, then

1. $(\mathcal{M}_P, \mathscr{R})$ is globally stable
2. If P^* is the unique fixed point of $\mathscr{R}$ in $\mathcal{M}_P$, then

$$u = -F^*x \quad \text{where} \quad F^* := (Q + \beta B'P^*B)^{-1}(\beta B'P^*A)$$

is the unique optimal policy for the LQ model (β, A, B, C, Q, R)

Example: Profit Maximization with Adjustment Costs

A monopolist faces **inverse demand function**

$$p_t := p(q_t, z_t) = a_0 - a_1 q_t + z_t$$

where

- q_t is output
- p_t is price
- the demand shock z_t follows

$$z_{t+1} = \rho z_t + \sigma \eta_{t+1}, \qquad \{\eta_t\} \overset{\text{IID}}{\sim} N(0,1)$$

The monopolist chooses $\{q_t\}$ to maximize PDV of profits:

$$\mathbb{E} \sum_{t=0}^{\infty} \beta^t \pi_t$$

Current profits are given by

$$\pi_t := (p_t - c)q_t - \gamma(q_{t+1} - q_t)^2$$

Here

- $\gamma(q_{t+1} - q_t)^2 =$ adjustment costs
- $\gamma \geqslant 0$ is a parameter
- $c > 0$ is unit cost of current production
- $0 < \beta < 1$

What would happen if $\gamma = 0$?

Monopolist should choose output to maximize current profit, implying

$$q_t = \bar{q}_t := \frac{a_0 - c + z_t}{2a_1}$$

For other γ, we might expect that

- if $\gamma \approx 0$, then q_t will track $\bar{q}_t$ closely
- if γ is larger, then q_t will be smoother than $\bar{q}_t$, as the monopolist seeks to avoid adjustment costs

Let's see if this intuition is correct

Step 1: Formulate as a dynamic programming problem

- State is $(q, z) \in \mathbb{R}^2$
- Control is $q' =$ next period output

Bellman equation is

$$v(q,z) = \max_{q'} \left\{ (p(q,z) - c)q - \gamma(q' - q)^2 + \beta \, \mathbb{E}_z v(q', z') \right\}$$

There's an easy way to solve this...

...**if** we can rephrase as an LQ model

First we modify rewards of the firm to

$$\mathbb{E}\sum_{t=0}^{\infty}\beta^t(\pi_t - a_1\bar{q}_t^2) \quad \text{where} \quad \bar{q}_t := \frac{a_0 - c + z_t}{2a_1}$$

Changes lifetime value but

$$\mathbb{E}\sum_{t=0}^{\infty}\beta^t(\pi_t - a_1\bar{q}_t^2) = \mathbb{E}\sum_{t=0}^{\infty}\beta^t\pi_t - a_1\mathbb{E}\sum_{t=0}^{\infty}\beta^t\bar{q}_t^2$$

Hence optimal production sequence $\{q_t\}$ will be identical

- Why?

Next set

$$u_t := q_{t+1} - q_t$$

Ex. Show that

$$\pi_t - a_1 \bar{q}_t^2 = -a_1 (q_t - \bar{q}_t)^2 - \gamma u_t^2$$

- Note this is quadratic in $(q_t, \bar{q}_t, u_t)$

Switching to a minimization problem, current loss is

$$\ell_t := a_1 (q_t - \bar{q}_t)^2 + \gamma u_t^2$$

Next we set up dynamics as linear in state and control

With $m_0 := (a_0 - c)/2a_1$ and $m_1 := 1/2a_1$, we have

$$\bar{q}_t = m_0 + m_1 z_t$$

Ex. Show that

$$\bar{q}_{t+1} = m_0(1 - \rho) + \rho \bar{q}_t + m_1 \sigma \xi_{t+1}$$

By our definition of u_t, the dynamics of q_t are

$$q_{t+1} = q_t + u_t$$

With these observations we can write the dynamic component of the LQ system as

$$x_{t+1} = Ax_t + Bu_t + C\xi_{t+1}$$

when

$$x_t := \begin{pmatrix} \bar{q}_t \\ q_t \\ 1 \end{pmatrix}$$

and

$$A = \begin{pmatrix} \rho & 0 & m_0(1-\rho) \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} m_1\sigma \\ 0 \\ 0 \end{pmatrix}$$

Ex. Check it

Recall our intuition:

if $\gamma = 0$, then monopolist should set

$$q_t = \bar{q}_t \text{ for all } t$$

For other γ, we expect that

- if γ close to zero $\implies$ q_t will track $\bar{q}_t$ closely
- if γ is larger, then q_t will be smoother

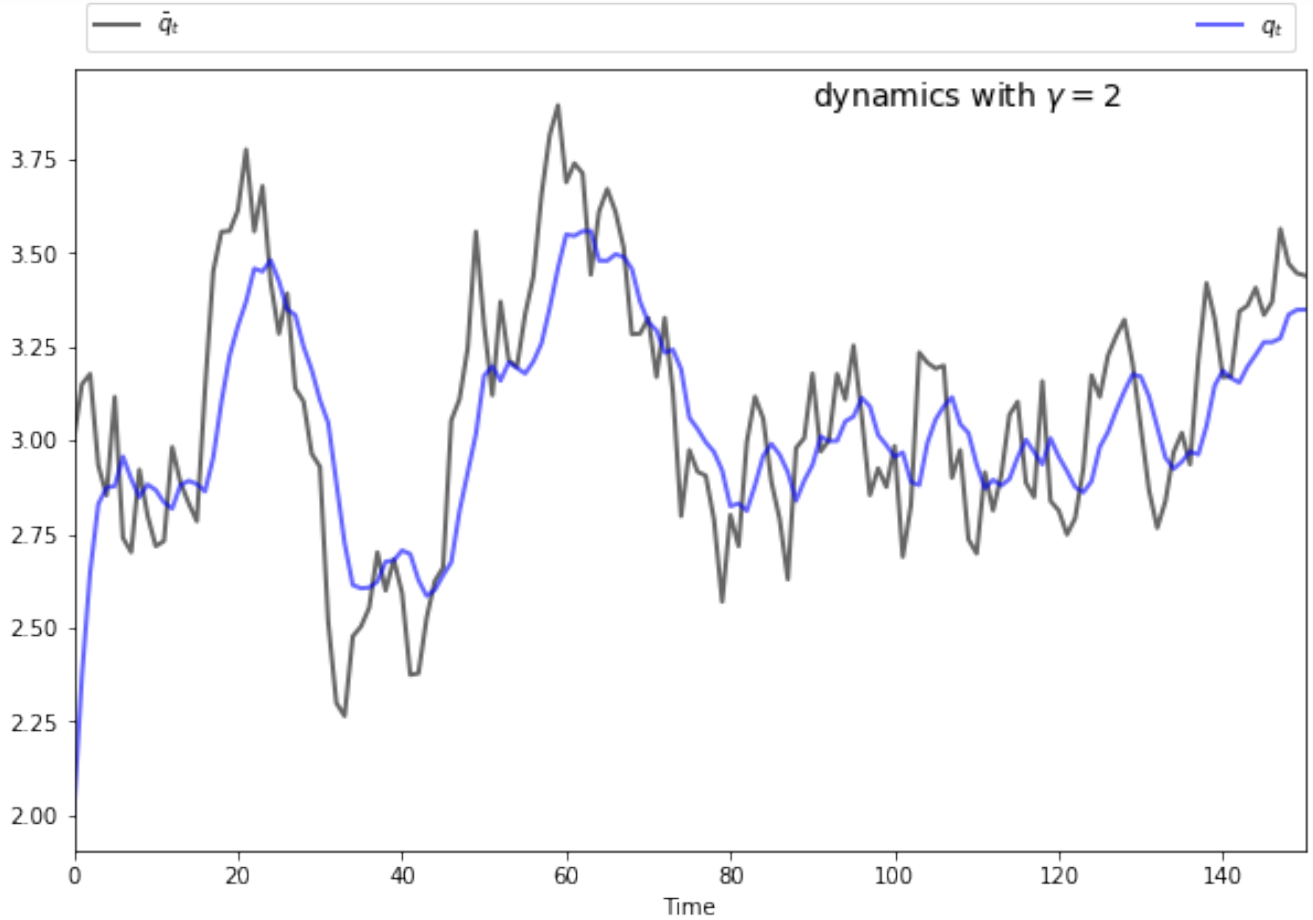

Figure: Output with adjustment costs when $\gamma = 2$

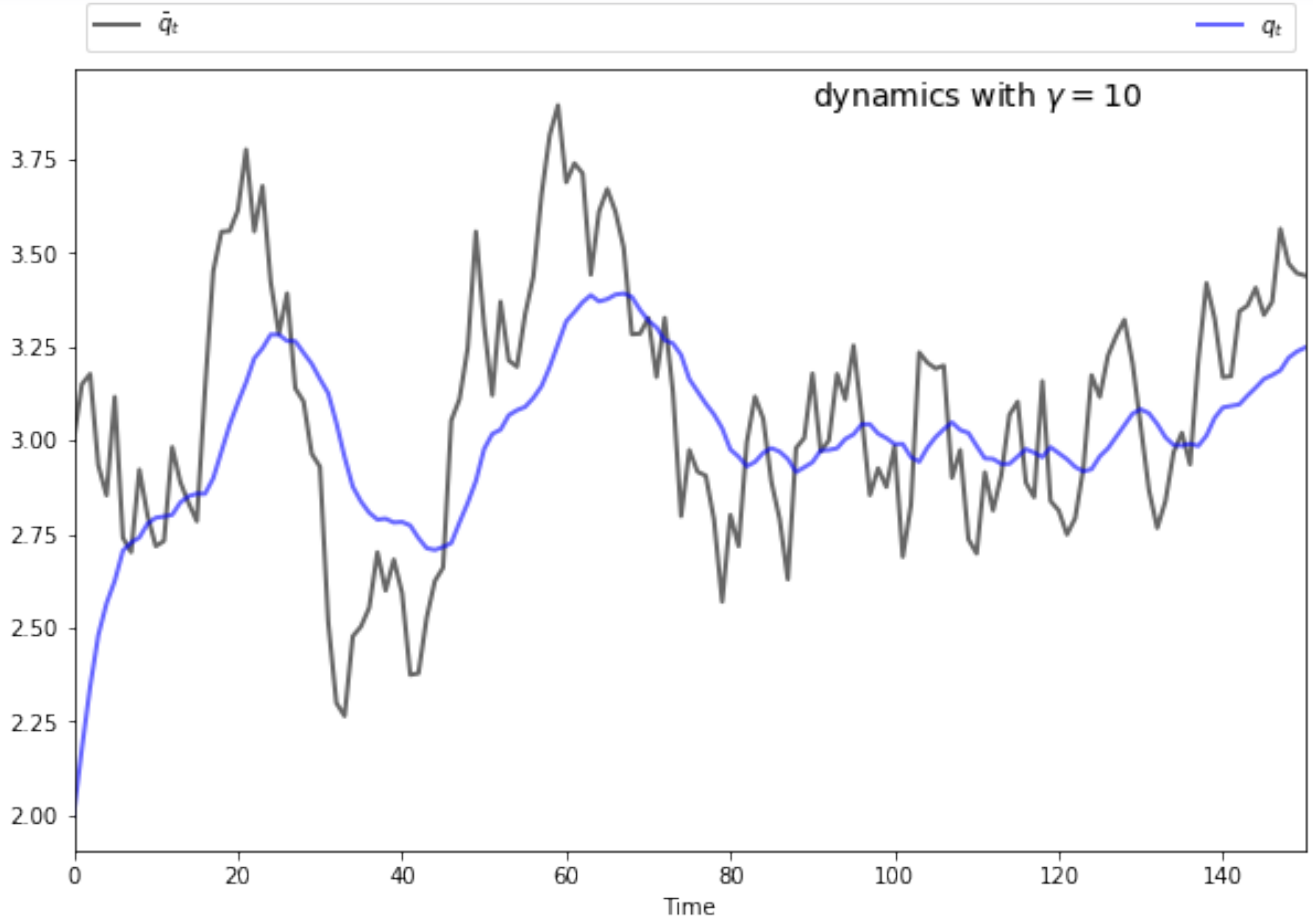

Figure: Output with adjustment costs when $\gamma = 10$

Finite State Markov Decision Processes

Let X and A be finite, $\Gamma(x) \subset$ A at each $x \in$ X

A **stochastic kernel** from

$$\mathbb{G} := \{(x, a) \in \mathsf{X} \times \mathsf{A} : a \in \Gamma(x)\}$$

to X is a family of distributions $\Pi(x, a, \cdot)$ over X, one for each (x, a) in $\mathbb{G}$

Example.

- X = inventory levels for a firm, A = {order stock, don't order}
- $\Pi(x, a, \cdot)$ is a distribution for next period inventory given current state and action

A **finite state Markov decision process (MDP)** consists of

1. a nonempty finite set X called the **state space**,
2. a nonempty finite set A called the **action space**,
3. a **feasible correspondence** Γ from $\mathsf{X} \to \mathsf{A}$,
4. a **reward function** $r\colon \mathbb{G} \to \mathbb{R}$, where

$$\mathbb{G} := \{(x, a) \in \mathsf{X} \times \mathsf{A} : a \in \Gamma(x)\}$$

5. a **discount factor** $\beta \in (0, 1)$ and
6. a **stochastic kernel** Π from $\mathbb{G}$ to X

The set $\mathbb{G}$ is called the set of **feasible state-action pairs**

An informal algorithm illustrating dynamics and reward flow:

---

set $t \leftarrow 0$ and take input x_0 ;
while $t < \infty$ **do**
    controller observes x_t ;
    chooses action a_t ;
    receives $r(x_t, a_t)$;
    nature draws x_{t+1} from $\Pi(x_t, a_t, \cdot)$;
    $t \leftarrow t + 1$;
end

Objective: choose a **state-contingent** action path $\{a_t\}$ that maximizes

$$\mathbb{E} \sum_{t \geqslant 0} \beta^t r(x_t, a_t)$$

- State contingency means that a_t is a function of x_t

The set of **feasible policies** is

$$\Sigma := \{\sigma \in \mathsf{A}^{\mathsf{X}} : \sigma(x) \in \Gamma(x) \text{ for all } x \in \mathsf{X}\}$$

Interpretation: Choosing σ from Σ means

- respond to state x_t with action $a_t := \sigma(x_t)$ at every t

If we commit to σ in Σ then x_{t+1} is drawn from $\Pi(x_t, \sigma(x_t), \cdot)$ at every t

Given $x_0 = x$, this is an (x, Π_σ)-chain for Π_σ defined by

$$\Pi_\sigma(x, y) := \Pi(x, \sigma(x), y) \qquad (x, y \in \mathsf{X})$$

Rewards at each point in time are $r(x_t, a_t) = r(x_t, \sigma(x_t))$

Let

$$r_\sigma(x) := r(x, \sigma(x))$$

Now

$$\mathbb{E}[r(x_t, a_t) \,|\, x_0 = x] = \mathbb{E}[r_\sigma(x_t) \,|\, x_0 = x] = \Pi_\sigma^t r_\sigma(x)$$

The lifetime value of following σ starting from state x can now be written as

$$\begin{aligned} v_\sigma(x) &= \mathbb{E}\left[\sum_{t\geqslant 0} \beta^t r(x_t, \sigma(x_t)) \,|\, x_0 = x\right] \\ &= \sum_{t\geqslant 0} \beta^t \mathbb{E}\left[r(x_t, \sigma(x_t)) \,|\, x_0 = x\right] \\ &= \sum_{t\geqslant 0} \beta^t (\Pi_\sigma r_\sigma)(x) \end{aligned}$$

In vector notation with v_σ and r_σ viewed as column vectors, this is

$$v_\sigma = \sum_{t\geqslant 0} \beta^t \Pi_\sigma^t r_\sigma$$

Optimality

The **value function** is defined as

$$v^*(x) := \sup_{\sigma \in \Sigma} v_\sigma(x) \qquad (x \in \mathsf{X})$$

- The maximal lifetime value we can extract from each state
- consistent with previous usage of the term "value function"

A policy $\sigma \in \Sigma$ is called **optimal** if $v_\sigma = v^*$

- Attains the supremum at all states

Theorem. v^* satisfies the Bellman equation

$$v^*(x) = \max_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \sum_{y \in \mathsf{X}} v^*(y) \Pi(x,a,y) \right\}$$

at every $x \in \mathsf{X}$

Moreover, $\sigma \in \Sigma$ is optimal if and only

$$\sigma(x) \in \operatorname*{argmax}_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \sum_{y \in \mathsf{X}} v^*(y) \Pi(x,a,y) \right\}$$

at every $x \in \mathsf{X}$ and at least one such policy exists

Proof: Coming soon

The statement that a feasible policy σ is optimal if and only

$$\sigma(x) \in \operatorname*{argmax}_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \sum_{y \in \mathsf{X}} v^*(y) \Pi(x,a,y) \right\}$$

at every $x \in \mathsf{X}$ is called **Bellman's principle of optimality**

Given arbitrary $v \in \mathbb{R}^{\mathsf{X}}$, we say that σ is **v-greedy** if

$$\sigma(x) \in \operatorname*{argmax}_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \sum_{y \in \mathsf{X}} v(y) \Pi(x,a,y) \right\} \qquad \forall\, x \in \mathsf{X}$$

For $\sigma \in \Sigma$, Bellman's principle of optimality becomes:

$$\sigma \text{ is optimal} \iff \sigma \text{ is } v^*\text{-greedy}$$

If we can compute v^* then the rest is easy

To do so we use the Bellman operator

$$Tv(x) = \max_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \sum_{y \in \mathsf{X}} v(y) \Pi(x,a,y) \right\}$$

Proposition. Under the stated assumptions,

1. T is a contraction of modulus β on $(\mathbb{R}^{\mathsf{X}}, d_\infty)$
2. The unique fixed point of T in $\mathbb{R}^{\mathsf{X}}$ is v^*

To prove that T is a contraction we can use the following

Lemma. If E is any set and $f, g \in bE$, then

$$|\sup_{a\in E} f(a) - \sup_{a\in E} g(a)| \leqslant \sup_{a\in E} |f(a) - g(a)|$$

Proof: If f and g have the stated properties, then

$$f = f - g + g \leqslant |f - g| + g$$

$$\therefore \quad \sup f \leqslant \sup |f - g| + \sup g$$

$$\therefore \quad \sup f - \sup g \leqslant \sup |f - g|$$

Reversing the roles of f and g completes the proof

Now we show that T is a contraction of modulus β on $\mathbb{R}^{\mathsf{X}}$

For any v, w in $\mathbb{R}^{\mathsf{X}}$ and $x \in \mathsf{X}$ we have

$$
\begin{aligned}
|Tv(x)| - Tw(x)| &\leqslant \beta \max_{a \in \Gamma(x)} \left| \sum_y \Pi(x,a,y)[v(y) - w(y)] \right| \\
&\leqslant \sum_y \Pi(x,a,y) \beta \ |v(y) - w(y)| \\
&\leqslant \beta \|v - w\|_\infty
\end{aligned}
$$

Taking the supremum over all $x \in \mathsf{X}$ yields the desired result

Algorithm 3: Value function iteration (finite state space)

input $v_0 \in \mathbb{R}^{\mathsf{X}}$, an initial guess of v^* ;
input τ, a tolerance level for error ;
$\epsilon \leftarrow \tau + 1$;
$n \leftarrow 0$;
while $\epsilon > \tau$ **do**
    for $x \in \mathsf{X}$ **do**
        $v_{n+1}(x) \leftarrow Tv_n(x)$;
    end
    $\epsilon \leftarrow \|v_n - v_{n+1}\|_\infty$;
    $n \leftarrow n + 1$;
end
return v_n

An alternative algorithm:

Algorithm 4: Howard's policy iteration algorithm

input $\sigma_0 \in \Sigma$, an initial guess of σ^* ;
$n \leftarrow 0$;
$\epsilon \leftarrow 1$;
while $\epsilon > 0$ **do**
    $v_n \leftarrow$ the σ_n-value function $\sum_{t \geqslant 0} \beta^t \Pi^t_{\sigma_n} r_{\sigma_n}$;
    $\sigma_{n+1} \leftarrow$ the v_n greedy policy ;
    $\epsilon \leftarrow \|\sigma_n - \sigma_{n+1}\|_\infty$;
    $n \leftarrow n + 1$;
end
return σ_n

Fact. When X is finite, $\{\sigma_n\}$ converges to the exact optimal policy in a finite number of steps

Proofs can be found in

- "Markov Decision Processes." Puterman (Wiley, 2005)
- "EDTC." Stachurski (MIT Press, 2009)

Intuition:

- $v_{n+1}(x) - v_n(x) > 0$ at some $x \in \mathsf{X}$ when σ_n is not optimal
- Hence $\{\sigma_n\}$ does not cycle
- Since Σ is finite, eventual convergence is guaranteed

One step in Howard's PI algorithm is computing v_σ given σ

We could do this by truncating: With T large,

$$v_\sigma \approx \sum_{t=0}^{T} \beta^t \Pi_\sigma^t r_\sigma$$

Another way to compute v_σ is by making use of the operator T_σ defined at $v \in \mathbb{R}^{\mathsf{X}}$ by

$$T_\sigma v(x) = r(x, \sigma(x)) + \beta \sum_{y \in \mathsf{X}} v(y) \Pi(x, \sigma(x), y)$$

or, in vector notation,

$$T_\sigma v = r_\sigma + \beta \Pi_\sigma v$$

Lemma. For any given σ in Σ,

1. the σ-value function v_σ is the unique fixed point of T_σ in $\mathbb{R}^{\mathsf{X}}$
2. Moreover, $T_\sigma^n v \to v_\sigma$ as $n \to \infty$ for all $v \in \mathbb{R}^{\mathsf{X}}$

Proof: For fixed σ in Σ we have

$$T_\sigma v_\sigma = r_\sigma + \beta \Pi_\sigma \left(\sum_{t \geqslant 0} \beta^t \Pi_\sigma^t r_\sigma \right)$$

$$= r_\sigma + \left(\sum_{t \geqslant 1} \beta^t \Pi_\sigma^t r_\sigma \right) = \sum_{t \geqslant 0} \beta^t \Pi_\sigma^t r_\sigma$$

In other words, $T_\sigma v_\sigma = v_\sigma$

Moreover, T_σ is a contraction of modulus β on $\mathbb{R}^X$

Indeed, for any v, w in $\mathbb{R}^X$ we have

$$
\begin{aligned}
|T_\sigma v(x)| - T_\sigma w(x)| &= \beta \left| \sum_y \Pi(x, \sigma(x), y)[v(y) - w(y)] \right| \\
&\leqslant \sum_y \Pi(x, \sigma(x), y) \beta \ |v(y) - w(y)| \\
&\leqslant \beta \|v - w\|_\infty
\end{aligned}
$$

Taking the supremum over all $x \in X$ yields the desired result

Alternatively, we can use the Neumann series theorem

In particular, the linear system

$$v = r_\sigma + \beta \, \Pi_\sigma v$$

has the unique solution

$$v_\sigma = \sum_{t \geqslant 0} \beta^t \Pi_\sigma^t r_\sigma = (I - \beta \, \Pi_\sigma)^{-1} r_\sigma$$

whenever the spectral radius of $\beta \, \Pi_\sigma$ is less than one

- This is always true (why?)
- matrix inversion approach which works well when X is not large

An Inventory Problem

Previously we studied a firm whose inventory behavior followed (s, S) dynamics

- large, infrequent orders

Can we replicate this in an optimizing model?

Inventory for the firm obeys

$$i_{t+1} = (i_t - D_{t+1})_+ + Sa_t$$

Here

- $\{D_t\}$ is a demand shock and $t_+ := \max\{t, 0\}$
- action $a_t \in \{0, 1\}$

State is $x = i =$ level of inventory

The firm can stock at most kS items at one time, so

$$\Gamma(x) = \begin{cases} \{0,1\} & \text{if } x \leqslant (k-1)S \\ \{0\} & \text{otherwise} \end{cases}$$

- feasible choices for a_t when current state is x

Assume IID demand shocks with common PMF φ on $\{0,1,\ldots\}$

The stochastic kernel Π is given by

$$\begin{aligned} \Pi(x,a,y) &= \mathbb{P}\{(x - D_{t+1})_+ + Sa = y\} \\ &= \sum_{d \geqslant 0} \mathbb{1}\{(x-d)_+ + Sa = y\}\varphi(d) \end{aligned}$$

Assuming a unit markup, profits are

$$\mathbb{E}\sum_{t\geqslant 0}\beta^t\pi_t \qquad \text{where } \pi_t := i_t \wedge D_{t+1} - ca_t$$

- c is a fixed cost of ordering inventory
- orders in excess of inventory are lost rather than backfilled

Bellman equation:

$$v(x) = \max_{a\in\Gamma(x)}\left\{\sum_d (x\wedge d)\varphi(d) - ca + \beta\sum_d v((x-d)_+ + Sa)\varphi(d)\right\}$$

Here x in $\mathsf{X} := \{0,1,\ldots kS\}$

Finite state MDP theory implies that v^* satisfies the Bellman equation

$$v^*(x) = \max_{a \in \Gamma(x)} \left\{ \sum_d (x \wedge d)\varphi(d) - ca + \beta \sum_d v^*((x-d)_+ + Sa)\varphi(d) \right\}$$

at every $x \in \mathsf{X}$

Moreover, a feasible policy σ is optimal if and only

$$\sigma(x) \in \operatorname*{argmax}_{a \in \Gamma(x)} \left\{ \sum_d (x \wedge d)\varphi(d) - ca + \beta \sum_d v^*((x-d)_+ + Sa)\varphi(d) \right\}$$

at every $x \in \mathsf{X}$ and at least one such policy exists

We can solve this problem by value function iteration

The Bellman operator in this context is

$$Tv(x) = \max_{a \in \Gamma(x)} \left\{ \sum_d (x \wedge d)\varphi(d) - ca + \beta \sum_d v((x-d)_+ + Sa)\varphi(d) \right\}$$

From the theory of finite state MDPs we know that

- T is a contraction of modulus β on $(\mathbb{R}^X, d_\infty)$
- Its unique fixed point in $\mathbb{R}^X$ is v^*

Ex. Using the bound

$$|\sup_{a\in D} f(a) - \sup_{a\in D} g(a)| \leqslant \sup_{a\in D} |f(a) - g(a)|$$

show directly that, for any v, w in $\mathbb{R}^X$,

$$|Tv(x)| - Tw(x)| \leqslant$$

$$\beta \max_{a\in\Gamma(x)} \left| \sum_d v((x-d)_+ + Sa)\varphi(d) - \sum_d w((x-d)_+ + Sa)\varphi(d) \right|$$

Ex. Use the last bound to show that

$$\|Tv - Tw\|_\infty \leqslant \beta \|v - w\|_\infty$$

Implementation, experiments:

- See the notebook `inventory_dp.ipynb`