Dynamic Programming

Chapter 10: Continuous Time

Thomas J. Sargent and John Stachurski

2024

Background

Earlier chapters treated dynamics in discrete time

Now we switch to continuous time

We restrict ourselves to finite state spaces (pure jump processes)

- permits a rigorous and self-contained treatment
- covers useful models
- lays foundations for a treatment of general state problems

Our first step is to review continuous time Markov chains

Recall: if $(X_t) = (X_0, X_1, \ldots)$ is P-Markov and $\psi_t \stackrel{d}{=} X_t$, then

$$\psi_{t+1} = \psi_t P \quad \text{for all } t$$

This rule is a linear difference equation in distribution space

How to shift to continuous time?

Answer: distributions follow a linear **differential equation** in distribution space

Hence we recall some facts about linear differential equations

- start with scalar case
- then shift to vector valued linear ODEs

The exponential function

The real-valued **exponential function** can be defined by

$$\mathrm{e}^x =: \exp(x) := \sum_{k \geqslant 0} \frac{x^k}{k!} \qquad (x \in \mathbb{R})$$

Properties: For $a, b \in \mathbb{R}$,

- $\mathrm{e}^{a+b} = \mathrm{e}^a \mathrm{e}^b$
- $t \mapsto \mathrm{e}^{ta}$ is differentiable and

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathrm{e}^{ta} = a\mathrm{e}^{ta}$$

Example. Let $u_t =$ balance of a savings account paying continuously compounded interest rate r

Then

$$u'_t := \frac{\mathrm{d}}{\mathrm{d}t} u_t = r u_t \quad \text{for all } \; t \geqslant 0, \quad u_0 \text{ given} \tag{1}$$

Ex. Show that $u_t := \mathrm{e}^{rt} u_0$ is the only solution to $u'_t = r u_t$

Proof: This function is a solution because

$$\frac{\mathrm{d}}{\mathrm{d}t} u_t = \frac{\mathrm{d}}{\mathrm{d}t} \mathrm{e}^{rt} u_0 = r\mathrm{e}^{rt} u_0 = r u_t$$

Why is it the only solution?

Suppose $t \mapsto y_t$ also satisfies $y_t' = r y_t$ and $y_0 = u_0$

Then

$$\frac{\mathrm{d}}{\mathrm{d}t} \left(y_t \, \mathrm{e}^{-rt} \right) = y_t' \, \mathrm{e}^{-rt} - r y_t \, \mathrm{e}^{-rt} = r y_t \, \mathrm{e}^{-rt} - r y_t \, \mathrm{e}^{-rt} = 0$$

Hence $y_t \, \mathrm{e}^{-rt}$ is constant in t on $\mathbb{R}_+$

In other words, $y_t = c \, \mathrm{e}^{rt}$ for some c

Setting $t = 0$ and using the initial condition gives $c = u_0$

$$\therefore \quad y_t = \mathrm{e}^{rt} u_0 = u_t$$

Complex exponentials

The exponential e^{λ} of $\lambda \in \mathbb{C}$ is defined analogously:

$$\mathrm{e}^{\lambda} =: \exp(\lambda) := \sum_{k \geqslant 0} \frac{\lambda^k}{k!}$$

From the identity $\mathrm{e}^{ib} = \cos(b) + i\sin(b)$

- i is the imaginary unit

Using this identity and $\lambda = a + ib$ gives

$$\mathrm{e}^{\lambda} = \mathrm{e}^{a+ib} = \mathrm{e}^{a}(\cos(b) + i\sin(b))$$

This equation will soon prove useful

Extension to matrices

The real exponential formula extends to the **matrix exponential** via

$$\mathrm{e}^A := I + A + \frac{A^2}{2!} + \cdots = \sum_{k \geqslant 0} \frac{A^k}{k!}$$

- A is any square matrix (or linear operator)
- the series always converges in norm

In the next slide, $\sigma(A) :=$ all eigenvalues (**spectrum**) of A

Lemma. Let A and B be square matrices

1. If A is diagonalizable with $A = PDP^{-1}$, then $\mathrm{e}^{A} = P\mathrm{e}^{D}P^{-1}$
2. If $AB = BA$, then $\mathrm{e}^{A+B} = \mathrm{e}^{A}\mathrm{e}^{B}$
3. $\mathrm{e}^{A^\top} = (\mathrm{e}^{A})^\top$ and $\mathrm{e}^{mA} = (\mathrm{e}^{A})^m$ for all $m \in \mathbb{N}$
4. $\lambda \in \sigma(A)$ iff $\mathrm{e}^{\lambda} \in \sigma(\mathrm{e}^{A})$
5. $t \mapsto \mathrm{e}^{tA}$ is differentiable and
$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{tA} = A\mathrm{e}^{tA} = \mathrm{e}^{tA}A$$
6. The fundamental theorem of calculus holds:
$$\mathrm{e}^{tA} = \mathrm{e}^{sA} + \int_s^t \mathrm{e}^{\tau A}A\,\mathrm{d}\tau \quad \text{for all } s \leqslant t$$

In the last slide, differentiation and integration are element-by-element

Example.

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} t^2 \\ \ln t \end{pmatrix} = \begin{pmatrix} (1/2)t \\ (1/t) \end{pmatrix}$$

and

$$\int \begin{pmatrix} f(t) & g(t) \\ u(t) & v(t) \end{pmatrix} \mathrm{d}t = \begin{pmatrix} \int f(t)\,\mathrm{d}t & \int g(t)\,\mathrm{d}t \\ \int u(t)\,\mathrm{d}t & \int v(t)\,\mathrm{d}t \end{pmatrix}$$

Ex. Confirm that $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{tA} = A\mathrm{e}^{tA}$

Proof: Observe that, for any $t \in \mathbb{R}$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{tA} = \lim_{h\to 0} \frac{\mathrm{e}^{tA+hA} - \mathrm{e}^{tA}}{h} = \mathrm{e}^{tA} \lim_{h\to 0} \frac{\mathrm{e}^{hA} - I}{h}$$

By definition,

$$\frac{\mathrm{e}^{hA} - I}{h} = A + \frac{1}{2!}hA^2 + \frac{1}{3!}h^2A^3 + \cdots$$

This converges to A as $h \to 0$, so

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{tA} = \mathrm{e}^{tA}A$$

Ex. Using the lemma, show that e^{A} is invertible with inverse e^{-A}

Fix $n \times n$ matrix A and let $B = -A$

Evidently A and B commute (i.e., $AB = BA$), so

$$\mathrm{e}^{A}\mathrm{e}^{B} = \mathrm{e}^{A+B} = \mathrm{e}^{A-A} = \mathrm{e}^{0}$$

Moreover,

$$\mathrm{e}^{0} = I + \sum_{k \geqslant 0} \frac{0^k}{k!} = I$$

Hence $\mathrm{e}^{A}\mathrm{e}^{-A} = I$, which proves the claim

Continuous time dynamical systems

Recall:

- a discrete dynamical system is a pair (U, S), where U is a set and S is a self-map on U
- trajectories are sequences $(S^t u)_{t \geqslant 0} = (u, Su, S^2 u, \ldots)$, where $u \in U$ is the initial condition

What is the continuous time equivalent?

We consider a pair $(U, (S_t)_{t \geqslant 0})$ where U is any set and S_t is a self-map on U for each $t \in \mathbb{R}_+$

The interpretation is that if $u \in U$ is the current state of the system, then $S_t u$ will be the state after t units of time

The map $t \mapsto S_t u$ is the trajectory from u

Example. For the savings balance $u_t = \mathrm{e}^{rt} u_0$, we take $U = \mathbb{R}$ and $S_t u = \mathrm{e}^{rt} u$

Then $S_t u$ is the state at time t given initial deposit u

To understand the pair $(U, (S_t)_{t \geqslant 0})$ as a continuous time dynamical system, we require

1. that S_0 is the identity map and
2. the **semigroup property**: for all $t, t' \geqslant 0$,

$$S_{t+t'} = S_{t'} \circ S_t$$

Meaning: if we

- start at u
- move forward to $u_t := S_t u$ and
- move again to $S_{t'} u_t$ after another t' units of time

the outcome is the same as moving from u to $S_{t+t'}\, u$ in one step

Linear initial value problems

Let A be $n \times n$ and u_t', u_t be column vectors in $\mathbb{R}^n$

Proposition. The unique solution of the n-dimensional IVP

$$u_t' = Au_t, \qquad u_0 \in \mathbb{R}^n \text{ given} \tag{2}$$

in the set of continuous functions $t \mapsto u_t$ mapping $\mathbb{R}_+$ to $\mathbb{R}^n$ is

$$u_t = \mathrm{e}^{tA} u_0 \qquad (t \geqslant 0). \tag{3}$$

Proof: That $u_t := \mathrm{e}^{tA} u_0$ solves (2) follows from slide 9

Uniqueness can be proved using an argument similar to that used to solve the exercise on slide 4

The last proposition motivates us to study flows of the form

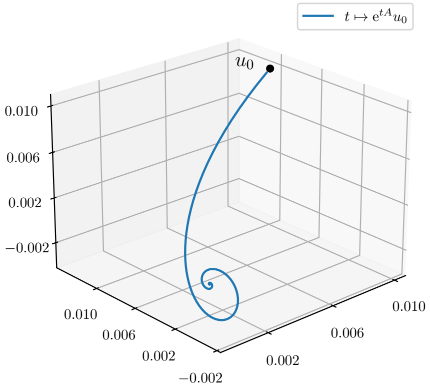
$$t \mapsto u_t, \quad u_t = \mathrm{e}^{tA} u_0 \qquad (t \geqslant 0) \tag{4}$$

where

- A is $n \times n$
- u_0 is a vector in $\mathbb{R}^n$ (initial condition)
- u_t is the "state" of the system at time t

The next slide illustrates for

$$A := \begin{pmatrix} -2.0 & -0.4 & 0 \\ -1.4 & -1.0 & 2.2 \\ 0.0 & -2.0 & -0.6 \end{pmatrix} \tag{5}$$



$t \mapsto e^{tA} u_0$
u_0
0.010
0.006
0.002
−0.002
0.010
0.006
0.002
−0.002
0.010
0.006
0.002

Stability

How do these exponential flows depend on A?

For example, when do we have

$$u_t := e^{tA} u_0 \to 0 \quad \text{as } t \to \infty$$

Two options

1. analyze this flow at every u_0
2. directly consider the matrix-valued flow $t \mapsto e^{tA}$

Below we take the second option, ask when $e^{tA} \to 0$

Suppose first that A is diagonalizable with $A = P^{-1}DP$

- $D = \mathrm{diag}_j(\lambda_j)$ contains the eigenvalues of A

Recall from slide 9 that for any $t \geqslant 0$,

$$\mathrm{e}^{tA} = \mathrm{e}^{tP^{-1}DP} = P^{-1}\mathrm{e}^{tD}P \tag{6}$$

and, moreover,

$$\mathrm{e}^{tD} = \mathrm{diag}(\mathrm{e}^{t\lambda_1}, \ldots, \mathrm{e}^{t\lambda_n})$$

Hence long run dynamics of e^{tA} fully determined by

$$t \mapsto \mathrm{e}^{t\lambda_j} \text{ for } j = 1, \ldots, n$$

So how does $\mathrm{e}^{t\lambda}$ evolve over time when $\lambda \in \mathbb{C}$?

To answer this question we write $\lambda = a + ib$ to obtain

$$\mathrm{e}^{t\lambda} = \mathrm{e}^{ta}(\cos(tb) + i\sin(tb)).$$

Hence

$$\mathrm{e}^{t\lambda} \to 0 \text{ as } t \to \infty \quad \iff \quad \operatorname{Re}\lambda < 0$$

$$\therefore \quad \mathrm{e}^{tA} \to 0 \text{ as } t \to \infty \quad \iff \quad \operatorname{Re}\lambda_j < 0 \text{ for all } \lambda_j \in \sigma(A)$$

Equivalently, $\mathrm{e}^{tA} \to 0$ if and only if $s(A) < 0$, where

$$s(A) := \max_{\lambda \in \sigma(A)} \operatorname{Re}\lambda$$

is called the **spectral bound** of A

The last result illustrated the importance of the spectral bound

Letting $\|\cdot\|$ be the matrix norm, we have

Lemma. For each $n \times n$ matrix A and $\tau > 0$ we have

$$\tau s(A) = s(\tau A)$$

Moreover,

$$\mathrm{e}^{s(A)} = \rho(\mathrm{e}^{A}) \quad \text{and} \quad s(A) = \lim_{t \to \infty} \frac{1}{t} \ln \|\mathrm{e}^{tA}\|$$

Ex. Confirm that $\rho(\mathrm{e}^A) = \mathrm{e}^{s(A)}$

<u>Proof</u>: Recall from slide 9 that

$$\lambda \in \sigma(A) \text{ if and only if } \mathrm{e}^\lambda \in \sigma(\mathrm{e}^A)$$

From the definition $s(A) := \max_{\lambda \in \sigma(A)} \operatorname{Re} \lambda$, we have

$$\rho(\mathrm{e}^A) = \max_{\lambda \in \sigma(\mathrm{e}^A)} |\lambda| = \max_{\lambda \in \sigma(A)} |\mathrm{e}^\lambda| = \max_{\lambda \in \sigma(A)} \mathrm{e}^{\operatorname{Re} \lambda} = \mathrm{e}^{s(A)}$$

The next theorem extends our stability result for the diagonal case

Theorem. For any square matrix A, the following statements are equivalent:

1. $s(A) < 0$
2. $\|\mathrm{e}^{tA}\| \to 0$ as $t \to \infty$
3. $\exists\ M, \omega > 0$ such that $\|\mathrm{e}^{tA}\| \leqslant M\mathrm{e}^{-t\omega}$ for all $t \geqslant 0$
4. $\int_0^\infty \|\mathrm{e}^{tA}u_0\|^p \,\mathrm{d}t < \infty$ for all $p \geqslant 1$ and $u_0 \in \mathbb{R}^n$

Let's sketch the proof that $s(A) < 0$ implies $\mathrm{e}^{tA} \to 0$ as $t \to \infty$

Suppose $s(A) < 0$

Fix $\varepsilon > 0$ such that $s(A) + \varepsilon < 0$ and

Recall that

$$s(A) = \lim_{t \to \infty} \frac{1}{t} \ln \|\mathrm{e}^{tA}\|$$

Hence $\exists$ a $T < \infty$ such that

$$\frac{1}{t} \ln \|\mathrm{e}^{tA}\| \leqslant s(A) + \varepsilon \text{ for all } t \geqslant T$$

Equivalently, for t large, we have $\|\mathrm{e}^{tA}\| \leqslant \mathrm{e}^{t(s(A)+\varepsilon)}$

The claim follows

Semigroup terminology

Advanced treatments of continuous time systems often begin with operator semigroups

Let's briefly describe these and connect them to things we have studied earlier

Let X be a finite set and let $(S_t)_{t \geqslant 0}$ be a subset of $\mathcal{L}(\mathbb{R}^{\mathsf{X}})$ indexed by $t \in \mathbb{R}_+$

The family $(S_t)_{t \geqslant 0}$ is called an **operator semigroup** on $\mathbb{R}^{\mathsf{X}}$ if

1. $S_0 = I$, where I is the identity,
2. $S_{t+t'} = S_t \circ S_{t'}$, and
3. $t \mapsto S_t$ is continuous as a map from $\mathbb{R}_+$ to $\mathcal{L}(\mathbb{R}^{\mathsf{X}})$

Given an operator semigroup $(S_t)_{t \geqslant 0}$ on $\mathcal{L}(\mathbb{R}^{\mathsf{X}})$, does there always exist a "vector field" type object that "generates" $(S_t)_{t \geqslant 0}$?

When X is finite, the answer is affirmative

This object is called the **infinitesimal generator** of the semigroup and is defined by

$$A = \lim_{t \downarrow 0} \frac{S_t - S_0}{t} = \lim_{t \downarrow 0} \frac{S_t - I}{t} \tag{7}$$

At $u \in U$, the vector Au indicates the instantaneous change in the state

Example. Fix A in $\mathcal{L}(\mathbb{R}^{\mathsf{X}})$ and let $(S_t)_{t \geqslant 0}$ be defined by $S_t = \mathrm{e}^{tA}$

Then $(S_t)_{t \geqslant 0}$ is an operator semigroup on $\mathbb{R}^{\mathsf{X}}$

To verify this we take $X = \{x_1, \ldots, x_n\}$ and S_t and A as $n \times n$ matrices

The operator semigroup properties now follow directly the lemma on slide 9

For example, S_t is continuous in t because it is differentiable in t

The infinitesimal generator is

$$\lim_{t \downarrow 0} \frac{S_t - S_0}{t} = \lim_{t \downarrow 0} \frac{\mathrm{e}^{tA} - \mathrm{e}^0}{t} = \frac{\mathrm{d}}{\mathrm{d}t} \mathrm{e}^{tA} \Big|_{t=0} = A\mathrm{e}^{0A} = A$$

The next slide shows that this is the only example of an operator semigroup on $\mathbb{R}^{\mathsf{X}}$ when $|\mathsf{X}| < \infty$

Proposition. If $(S_t)_{t\geqslant 0}$ is an operator semigroup on $\mathbb{R}^{\mathsf{X}}$ and X is finite, then

1. there exists an $A \in \mathcal{L}(\mathbb{R}^{\mathsf{X}})$ such that $S_t = \mathrm{e}^{tA}$ for all $t \geqslant 0$, and
2. A is the infinitesimal generator of $(S_t)_{t\geqslant 0}$.

Semigroups of this form are called **exponential semigroups**

Put differently: in finite dimensions, the only operator semigroups are exponential semigroups

Markov Semigroups

We are now ready to specialize to the Markov case, where dynamics evolve in distribution space

Let $|\mathsf{X}| = n$ and let $(X_t)_{t \geqslant 0}$ be P-Markov on X for some $P \in \mathcal{M}(\mathbb{R}^{\mathsf{X}})$

The marginal distributions of $(X_t)_{t \geqslant 0}$ evolve according to the linear difference system $\psi_{t+1} = \psi_t P$

We now seek a continuous time analog in the form of linear differential equations that drive the evolution of distributions

To this end we define an $n \times n$ matrix Q to be an **intensity matrix** when

$$Q(x, x') \geqslant 0 \text{ whenever } x \neq x' \quad \text{and} \quad \sum_{x'} Q(x, x') = 0 \text{ for all } x \in \mathsf{X}$$

Example. The matrix

$$Q := \begin{pmatrix} -2 & 1 & 1 \\ 0 & -1 & 1 \\ 2 & 1 & -3 \end{pmatrix}$$

is an intensity matrix, since

- off-diagonal terms are nonnegative and
- rows sum to zero

We call $\mathcal{D}(\mathsf{X})$ **invariant** for

$$\psi_t' = \psi_t \, Q, \qquad \psi_0 \in \mathcal{D}(\mathsf{X}) \text{ given.} \tag{8}$$

if the solution $(\psi_t)_{t \geqslant 0}$ remains in $\mathcal{D}(\mathsf{X})$ for all $t \geqslant 0$

- ψ_t and ψ_t' are understood to be row vectors

By the result on slide 16, we can rephase by stating that $\mathcal{D}(\mathsf{X})$ is invariant for (8) whenever

$$\psi_0 \in \mathcal{D}(\mathsf{X}) \quad \Longrightarrow \quad \psi_0 \, \mathrm{e}^{tQ} \in \mathcal{D}(\mathsf{X}) \text{ for all } t \geqslant 0 \tag{9}$$

Proposition. Let Q be $n \times n$ and set $P_t := \mathrm{e}^{tQ}$ for each $t \geqslant 0$ The following statements are equivalent:

1. Q is an intensity matrix.
2. P_t is a stochastic matrix for all $t \geqslant 0$.
3. the set of distributions $\mathcal{D}(\mathsf{X})$ is invariant for the IVP (8).

Meaning: the set of $n \times n$ intensity matrices coincides with the set of continuous time Markov models on X

Any specification outside this class fails to generate flows in distribution space.

Proof: See the book

Markov Semigroups

The family $(P_t)_{t \geqslant 0} = (\mathrm{e}^{tQ})_{t \geqslant 0}$ that solves $\psi_t' = \psi_t Q$ is an exponential semigroup

When Q is an intensity matrix, it is also called the **Markov semigroup** generated by Q

- Q is also called the infinitesimal generator of $(P_t)_{t \geqslant 0}$

$(P_t)_{t \geqslant 0}$ satisfies the semigroup property

$$P_{s+t} = P_s P_t \quad \text{for all } s, t \geqslant 0$$

This can be written more explicitly as

$$P_{s+t}(x, x') = \sum_{z \in \mathsf{X}} P_s(x, z) P_t(z, x')$$

for $s, t \geqslant 0$ and $x, x' \in \mathsf{X}$

- called the **Chapman–Kolmogorov equation**

The probability of moving from x to x' over $s + t$ units of time equals

1. the probability of moving from x to z over s units of time
2. and then z to x' over t units of time

summed over all z

Let $C(\mathbb{R}_+, \mathsf{X})$ be the set of right-continuous functions from $\mathbb{R}_+$ to X and let $(P_t)_{t \geqslant 0}$ be a Markov semigroup in $\mathcal{L}(\mathbb{R}^{\mathsf{X}})$

A **continuous time Markov chain** generated by $(P_t)_{t \geqslant 0}$ is a $C(\mathbb{R}_+, \mathsf{X})$-valued random element $(X_t)_{t \geqslant 0}$ that satisfies

$$\mathbb{P}\{X_{s+t} = x' \,|\, \mathcal{F}_s\} = P_t(X_s, x') \qquad \text{for all } s, t \geqslant 0 \text{ and } x' \in \mathsf{X} \tag{10}$$

where $\mathcal{F}_s := (X_\tau)_{0 \leqslant \tau \leqslant s}$ is the history of the process up to time s

We will call a continuous time Markov chain $(X_t)_{t \geqslant 0}$ **Q-Markov** when (10) holds and Q is the infinitesimal generator of $(P_t)_{t \geqslant 0}$

Let $(X_t)_{t \geqslant 0}$, Q and P_t be as above

Conditioning on $X_s = x$, we get

$$P_t(x, x') = \mathbb{P}\{X_{s+t} = x' \mid X_s = x\} \qquad (s, t \geqslant 0,\ x, x' \in \mathsf{X})$$

In what follows, $\mathbb{P}_x$ and $\mathbb{E}_x$ denote probabilities and expectations conditional on $X_0 = x$

Given $h \in \mathbb{R}^{\mathsf{X}}$, we have

$$\mathbb{E}_x\, h(X_t) = \sum_{x'} P_t(x, x') h(x') =: (P_t h)(x)$$

This expression mirrors the discrete time case

A jump chain construction

We now describe a standard method for constructing continuous time Markov chains by using three components:

1. an initial condition $\psi \in \mathcal{D}(\mathsf{X})$,
2. a **jump matrix** $\Pi \in \mathcal{M}(\mathbb{R}^{\mathsf{X}})$, and
3. a **rate function** λ mapping X to $(0, \infty)$.

The process (X_t)

- starts at state x, which is drawn from ψ
- waits there for an exponential time W with rate $\lambda(x)$ and
- updates to a new state x' drawn from $\Pi(x, \cdot)$

We take x' as the new state for the process and repeat

Algorithm 1: Jump chain algorithm

draw Y_0 from ψ, set $J_0 = 0$ and $k = 1$
while $t < \infty$ **do**
  draw W_k independently from Exp$(\lambda(Y_{k-1}))$
  $J_k \leftarrow J_{k-1} + W_k$
  $X_t \leftarrow Y_{k-1}$ for all t in $[J_{k-1}, J_k)$
  draw Y_k from $\Pi(Y_{k-1}, \cdot)$
  $k \leftarrow k + 1$
end

- (W_k) is called the sequence of **wait times**
- the sums $J_k = \sum_{i=1}^{k} W_i$ are called the **jump times** and
- (Y_k) is called the **embedded jump chain**

Let $I \in \mathcal{L}(\mathbb{R}^{\mathsf{X}})$ be the identity matrix ($I(x, x') = \mathbb{1}\{x = x'\}$)

Define $Q \in \mathcal{L}(\mathbb{R}^{\mathsf{X}})$ via

$$Q(x, x') = \lambda(x)(\Pi(x, x') - I(x, x')) \qquad (x, x' \in \mathsf{X})$$

Ex. Check that Q is an intensity matrix

Proposition. The process $(X_t)_{t \geqslant 0}$ generated by the jump chain algorithm is Q-Markov

Proof uses the Kolmogorov backward equation

- see the book for details

Some intuition for

$$Q(x, x') = \lambda(x)(\Pi(x, x') - I(x, x'))$$

If $x \neq x'$, the rate of flow from x to x' is

$$\lambda(x)\Pi(x, x') = Q(x, x')$$

What about $x = x'$?

The jump matrix Π is constructed s.t. $\Pi(x, x) = 0$

- at jump times, we actually jump (don't stay at x)

Rate of flow out of x is $\lambda(x)$

Hence the rate of flow from x to x is

$$-\lambda(x) = Q(x, x)$$

Application: inventory dynamics

Let X_t be a firm's inventory at time t

When current stock is $x > 0$, customers arrive at rate $\lambda(x) > 0$

- wait time for the next customer is $\mathrm{Exp}(\lambda(x))$

The k-th customer demands U_k units, where each U_k is an independent draw from a fixed distribution φ on $\mathbb{N}$

Inventory falls by $U_k \wedge X_t$

When inventory hits zero the firm orders b units of new stock

The wait time for new stock is $\mathrm{Exp}(\lambda(0))$

Let Y = inventory size after the next jump, given current stock x

If $x > 0$, then Y is a draw from the distribution of $x - U \wedge x$ where $U \sim \varphi$

If $x = 0$, then $Y \equiv b$

Hence Y is a draw from $\Pi(x, \cdot)$, where $\Pi(0, y) = \mathbb{1}\{y = b\}$ and, for $0 < x \leqslant b$,

$$\Pi(x, y) = \begin{cases} 0 & \text{if } x \leqslant y \\ \mathbb{P}\{x - U = y\} & \text{if } 0 < y < x \\ \mathbb{P}\{U \geqslant x\} & \text{if } y = 0 \end{cases} \tag{11}$$

Ex. Prove that Π is a stochastic matrix on $\mathsf{X} := \{0, 1, \ldots, b\}$

We can simulate the inventory process $(X_t)_{t \geqslant 0}$ via the jump chain algorithm

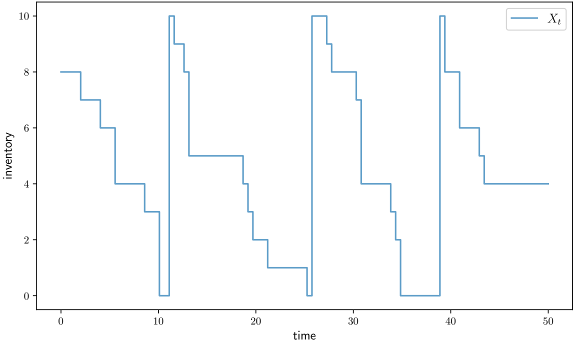
- (W_k) is the wait time for customers / new inventory and
- (Y_k) is the level of inventory immediately after each jump

By the proposition on slide 40, the process (X_t) is Q-Markov with

$$Q(x, x') = \lambda(x)(\Pi(x, x') - I(x, x'))$$

The next slide shows a simulation when orders are geometric, so that

$$\varphi(k) = \mathbb{P}\{U = k\} = (1 - \alpha)^{k-1}\alpha \qquad (k \in \mathbb{N}, \ \alpha \in (0, 1)).$$



X_t
inventory
time
0
2
4
6
8
10
0
10
20
30
40
50

Valuation with constant discounting

Consider

$$v(x) := \mathbb{E}_x \int_0^\infty \mathrm{e}^{-t\delta} h(X_t)\,\mathrm{d}t \qquad (x \in \mathsf{X})$$

for some $\delta \in \mathbb{R}$ and $h \in \mathbb{R}^{\mathsf{X}}$

- $(X_t)_{t \geqslant 0}$ is Q-Markov on X and $P_t = \mathrm{e}^{tQ}$

Interpretation:

- $h(X_t)$ is an instantaneous reward at t
- δ is a fixed discount rate
- $v(x)$ is lifetime value conditional on starting at x

Proposition. If $\delta > 0$, then v is finite, $\delta I - Q$ is bijective,

$$(\delta I - Q)^{-1} \geqslant 0 \quad \text{and} \quad v = (\delta I - Q)^{-1} h$$

In addition, v is the unique fixed point of

$$Uw = h + (Q + (1 - \delta) I) w \qquad \left(w \in \mathbb{R}^{\mathsf{X}} \right)$$

and U is order stable on $\mathbb{R}^{\mathsf{X}}$

Proof: Letting $A := Q - \delta I$, we claim that $s(A) < 0$

Using the result for spectral bounds in slide 22, we have

$$\begin{aligned} e^{s(Q-\delta I)} = \rho(e^{Q-\delta I}) &= \rho(e^{Q} e^{-\delta I}) \\ &= \rho(e^{Q} e^{-\delta} I) \\ &= e^{-\delta}\rho(e^{Q}) = e^{-\delta}\rho(P_1) = e^{-\delta} \end{aligned}$$

Therefore $s(Q - \delta I) = -\delta$

$$\therefore \quad s(A) = s(Q - \delta I) < 0$$

We have just shown that $s(A) = s(Q - \delta I) < 0$

Hence A has nonzero determinant and is therefore nonsingular

$$\therefore \quad -A = \delta I - Q \text{ is nonsingular / bijective}$$

Also, $s(A) < 0$ and the stability result on slide 24 yield

$$v(x) = \int_0^\infty \mathrm{e}^{-t\delta}\, \mathbb{E}_x\, h(X_t)\, \mathrm{d}t$$

$$= \int_0^\infty \mathrm{e}^{-t\delta}\, (P_t\, h)(x)\, \mathrm{d}t$$

$$= \int_0^\infty \mathrm{e}^{-t\delta} (\mathrm{e}^{tQ}\, h)(x)\, \mathrm{d}t = \int_0^\infty (\mathrm{e}^{tA}\, h)(x)\, \mathrm{d}t < \infty$$

We have

$$v = \int_0^\infty \mathrm{e}^{\tau A} h \, \mathrm{d}\tau = \int_0^t \mathrm{e}^{\tau A} h \, \mathrm{d}\tau + \int_t^\infty \mathrm{e}^{\tau A} h \, \mathrm{d}\tau$$

But

$$\begin{aligned} \int_t^\infty \mathrm{e}^{\tau A} h \, \mathrm{d}\tau &= \int_0^\infty \mathrm{e}^{(t+\tau) A} h \, \mathrm{d}\tau \\ &= \int_0^\infty \mathrm{e}^{tA} \mathrm{e}^{\tau A} h \, \mathrm{d}\tau = \mathrm{e}^{tA} \int_0^\infty \mathrm{e}^{\tau A} h \, \mathrm{d}\tau = \mathrm{e}^{tA} v \end{aligned}$$

$$\therefore \quad v = \int_0^t \mathrm{e}^{\tau A} h \, \mathrm{d}\tau + \mathrm{e}^{tA} v$$

Rearranging $v = \int_0^t \mathrm{e}^{\tau A} h \, \mathrm{d}\tau + \mathrm{e}^{tA} v$ and dividing by $t > 0$ yields

$$-\frac{\mathrm{e}^{tA} - I}{t} v = \frac{1}{t} \int_0^t \mathrm{e}^{\tau A} h \, \mathrm{d}\tau \tag{12}$$

By the fundamental theorem of calculus,

$$\lim_{t \to 0} \frac{1}{t} \int_0^t \mathrm{e}^{\tau A} h \, \mathrm{d}\tau = \frac{\mathrm{d}}{\mathrm{d}t} \int_0^t \mathrm{e}^{\tau A} h \, \mathrm{d}\tau \, \Big|_{t=0} = \mathrm{e}^{0A} h = I \, h = h$$

As a result, taking $t \to 0$ in (12)

$$-Av = -A\mathrm{e}^{0A} v = -\frac{\mathrm{d}}{\mathrm{d}t} \mathrm{e}^{tA} v \, \Big|_{t=0} = \lim_{t \to 0} -\frac{\mathrm{e}^{tA} - I}{t} v = h$$

We have shown that

1. $A := Q - \delta I$ is bijective and
2. $-Av = h$

Hence

$$v = -A^{-1}h = (-A)^{-1}h = (\delta I - Q)^{-1}h$$

From $v(x) = \mathbb{E}_x \int_0^\infty \mathrm{e}^{-t\delta} h(X_t) \,\mathrm{d}t$ we have

$$h \geqslant 0 \implies (\delta I - Q)^{-1}h \geqslant 0$$

Hence $(\delta I - Q)^{-1} \geqslant 0$, as claimed

It remains only to show that v is the unique fixed point of

$$Uw = h + (Q + (1 - \delta)I)w$$

and U is order stable on $\mathbb{R}^{\mathsf{X}}$

The first claim is true because

$$\begin{aligned}
Uw = w \iff & h + Qw + w - \delta w = w \\
\iff & h + Qw - \delta w = 0 \\
\iff & (\delta I - Q)w = h \\
\iff & w = (\delta I - Q)^{-1} h = v
\end{aligned}$$

To prove that U is order stable, we need to show that U is upward and downward stability on $\mathbb{R}^{\mathsf{X}}$

For upward stability, suppose that $w \in \mathbb{R}^{\mathsf{X}}$ and $Uw \geqslant w$

Then $h + Aw \geqslant 0$, or $-Aw \leqslant h$

But $-A^{-1} \geqslant 0$, so $w \leqslant -A^{-1}h = v$ and upward stability holds

The proof of downward stability is similar

Continuous time Markov decision processes

Fix two finite sets A and X, called the state and action spaces respectively

Informally, a continuous time Markov decision process is an optimization problem where the aim is to maximize

$$v(x) := \mathbb{E}_x \int_0^\infty \mathrm{e}^{-t\delta} r(X_t, A_t) \, \mathrm{d}t$$

where

- $X_t \in \mathsf{X}$ is the state
- $A_t \in \Gamma(X_t) \subset \mathsf{A}$ is the action

Formally...

A **continuous time Markov decision process** is a tuple $\mathcal{C} = (\Gamma, \delta, r, Q)$ consisting of

1. a nonempty **feasible correspondence** Γ from $\mathsf{X} \to \mathsf{A}$, which in turn defines the **feasible state-action pairs**

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

2. a constant $\delta > 0$, referred to as the **discount rate**
3. a function r from G to $\mathbb{R}$, referred to as the **reward function** and
4. an **intensity kernel** Q from G to X; that is, a map Q from $\mathsf{G} \times \mathsf{X}$ to $\mathbb{R}$ satisfying

$$\sum_{x' \in \mathsf{X}} Q(x, a, x') = 0 \quad \text{for all } (x, a) \text{ in } \mathsf{G}$$

and $Q(x, a, x') \geqslant 0$ whenever $x \neq x'$

Intuition: at state x with action a over the short interval from t to $t + h$,

- the controller receives instantaneous reward $r(x, a)h$ and
- the state transitions to state x' with probability $Q(x, a, x')h + o(h)$

The set of **feasible policies** is

$$\Sigma := \{\sigma \in \mathsf{A}^{\mathsf{X}} : \sigma(x) \in \Gamma(x) \text{ for all } x \in \mathsf{X}\} \tag{13}$$

Choosing policy σ from Σ means that we respond to state X_t with action $A_t := \sigma(X_t)$ at every $t \in \mathbb{R}_+$

Lifetime Values

Under policy σ, the state evolves according to the intensity matrix

$$Q_\sigma(x, x') := Q(x, \sigma(x), x')$$

Letting

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

the **lifetime value** of following σ starting from state x is defined as

$$v_\sigma(x) = \mathbb{E}_x \int_0^\infty \mathrm{e}^{-\delta t} r_\sigma(X_t) \, \mathrm{d}t$$

where $(X_t)_{t \geqslant 0}$ is Q_σ-Markov with $X_0 = x$

We call v_σ the **σ-value function**

Lemma. The σ-value function associated with $\sigma \in \Sigma$ obeys

$$v_\sigma = (\delta I - Q_\sigma)^{-1} r_\sigma$$

In addition, v_σ is the unique fixed point of

$$T_\sigma \, v = r_\sigma + (Q_\sigma + (1 - \delta) I) v.$$

and T_σ is order stable on $\mathbb{R}^{\mathsf{X}}$

This follows directly from

- $\delta > 0$
- the result on slide 47

Provides a straightforward method for computing v_σ

A policy $\sigma \in \Sigma$ is called v-**greedy** for $\mathcal{C}$ if

$$\sigma(x) \in \operatorname*{argmax}_{a \in \Gamma(x)} \left\{ r(x, a) + \sum_{x'} v(x') Q(x, a, x') \right\} \quad \text{for all } x \in \mathsf{X}. \tag{14}$$

A v-greedy policy chooses actions optimally to trade off

- high current rewards versus
- high rate of flow into future states with high values

The discount factor does not appear in (14) because the trade-off is instantaneous

Algorithm 2: Continuous time Howard policy iteration

input $\sigma_0 \in \Sigma$, an initial guess of σ^*

$k \leftarrow 0$

$\varepsilon \leftarrow 1$

while $\varepsilon > 0$ **do**

- $v_k \leftarrow (\delta I - Q_{\sigma_k})^{-1} r_{\sigma_k}$
- $\sigma_{k+1} \leftarrow$ a v_k-greedy policy
- $\varepsilon \leftarrow \mathbb{1}\{\sigma_k \neq \sigma_{k+1}\}$
- $k \leftarrow k + 1$

end

return σ_k

Optimality

For a continuous time MDP $\mathcal{C} = (\Gamma, \delta, r, Q)$ with σ-value functions $\{v_\sigma\}$,

- the **value function** generated by $\mathcal{C}$ is $v^* := \bigvee_\sigma v_\sigma$, and
- a policy is called **optimal** for $\mathcal{C}$ if $v_\sigma = v^*$.

A function $v \in \mathbb{R}^{\mathsf{X}}$ is said to satisfy a **Hamilton–Jacobi–Bellman** (**HJB**) equation if

$$\delta v(x) = \max_{a \in \Gamma(x)} \left\{ r(x, a) + \sum_{x'} v(x') Q(x, a, x') \right\} \tag{15}$$

for all $x \in \mathsf{X}$

We say that $\mathcal{C}$ obeys **Bellman's principle of optimality** if

$$\sigma \in \Sigma \text{ is optimal for } \mathcal{C} \quad \iff \quad \sigma \text{ is } v^*\text{-greedy}$$

Theorem. If $\mathcal{C} = (\Gamma, \delta, r, Q)$ is a continuous time MDP, then

1. the value function v^* is the unique solution to the HJB equation in $\mathbb{R}^{\mathsf{X}}$,
2. $\mathcal{C}$ obeys Bellman's principle of optimality, and
3. $\mathcal{C}$ has at least one optimal policy.

In addition, continuous time HPI converges to an optimal policy in finitely many steps

Proof: Let $\mathcal{C} = (\Gamma, \delta, r, Q)$ be a fixed continuous time MDP with lifetime values $\{v_\sigma\}$ and value function v^*

Consider the order stable ADP $\mathcal{A} := (\mathbb{R}^{\mathsf{X}}, \{T_\sigma\})$ with

$$T_\sigma \, v = r_\sigma + (Q_\sigma + (1 - \delta)I)v.$$

The ADP Bellman max-operator is $T := \bigvee_\sigma T_\sigma$, which can be written more explicitly as

$$(Tv)(x) = \max_{a \in \Gamma(x)} \left\{ r(x, a) + \sum_{x'} v(x')Q(x, a, x') \right\} + (1 - \delta)v(x) \tag{16}$$

For each $v \in \mathbb{R}^{\mathsf{X}}$, the set of v-max-greedy policies is nonempty

Since Σ is finite, it follows that $\mathcal{A}$ is max-stable

Hence an optimal policy always exists and the value function v^* is the unique fixed point of T in $\mathbb{R}^{\mathsf{X}}$

The last statement is equivalent to the assertion that v^* is the unique element of $\mathbb{R}^{\mathsf{X}}$ satisfying

$$v^*(x) = \max_{a \in \Gamma(x)} \left\{ r(x,a) + \sum_{x'} v^*(x') Q(x,a,x') \right\} + (1-\delta) v^*(x)$$

Rearranging this expression confirms that v^* is the unique solution to the HJB equation in $\mathbb{R}^{\mathsf{X}}$.

A policy is optimal for $\mathcal{A}$ (and hence $\mathcal{C}$) if and only if $T_\sigma v^* = T v^*$

This proves the claim that $\mathcal{C}$ obeys Bellman's principle of optimality

The continuous time HPI routine in slide 61 is just ADP max-HPI specialized to the current setting

Hence, continuous time HPI converges to an optimal policy in finitely many steps

Application: job search

We study continuous time job search with separation

A worker can be either unemployed (state 0) or employed (state 1)

When the worker is employed, she can be fired at any time

Firing occurs at rate $\alpha > 0$

- for $h \approx 0$, probability of being fired over $[t, t+h]$ is $\approx \alpha h$

When unemployed, the worker receives

- flow unemployment compensation c and
- job offers at rate κ

She discounts the future at rate $\delta > 0$

Job offers are at wage w in finite set W

Conditional on current w, the next offer is drawn from $P(w, \cdot)$

For the state space we set

$$\mathsf{X} = \{0, 1\} \times \mathsf{W} \text{ with typical state } x = (s, w)$$

Here

- s is binary and indicates current employment status
- w is the current wage

Let

$$\lambda(x) = \lambda(s, w) = \mathbb{1}\{s = 0\}\kappa + \mathbb{1}\{s = 1\}\alpha$$

denote the state-dependent jump rate

Let $a \in \mathsf{A} := \{0, 1\}$ indicate the action (reject, accept)

Let $\Pi(x, a, x')$ represent the jump probabilities, with

$$\Pi((0, w), a, (0, w')) = P(w, w')(1 - a) \qquad \text{(unemployed to unemployed)}$$

$$\Pi((0, w), a, (1, w')) = P(w, w')a \qquad \text{(unemployed to employed)}$$

$$\Pi((1, w), a, (0, w')) = P(w, w') \qquad \text{(employed to unemployed)}$$

$$\Pi((1, w), a, (1, w')) = 0 \qquad \text{(employed to employed)}$$

The probability assigned to the last line is zero because a jump from $s = 1$ occurs when the worker is fired

Motivated by the jump chain construction of intensity matrices in, we set

$$Q(x, a, x') = \lambda(x)(\Pi(x, a, x') - I(x, x'))$$

Fix $\sigma \in \Sigma := \{0, 1\}^{\mathsf{X}}$

The operator

$$Q_\sigma(x, x') := \lambda(x)(\Pi(x, \sigma(x), x') - I(x, x'))$$

is an intensity matrix for the jump chain under policy σ

- inventory is Q_σ-Markov under policy σ

If we define

$$r(x, a) = r((s, w), a) = c\mathbb{1}\{s = 0\} + w\mathbb{1}\{s = 1\},$$

then lifetime value is given by

$$v_\sigma(x) = \mathbb{E}_x \int_0^\infty \mathrm{e}^{-\delta t} r_\sigma(X_t) \, \mathrm{d}t,$$

where $(X_t)_{t \geqslant 0}$ is Q_σ-Markov and $X_0 = x$

With Γ defined by $\Gamma(x) = \mathsf{A}$ for all $x \in \mathsf{X}$, the tuple $\mathcal{C} = (\Gamma, \delta, r, Q)$ is a continuous time MDP

By the result on slide 63, An optimal policy exists and can be computed with HPI in a finite number of iterations

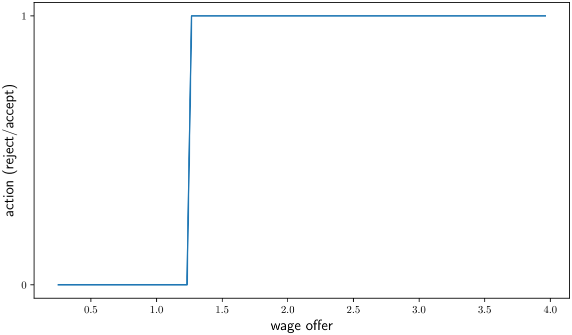


Figure: Continuous time job search policy

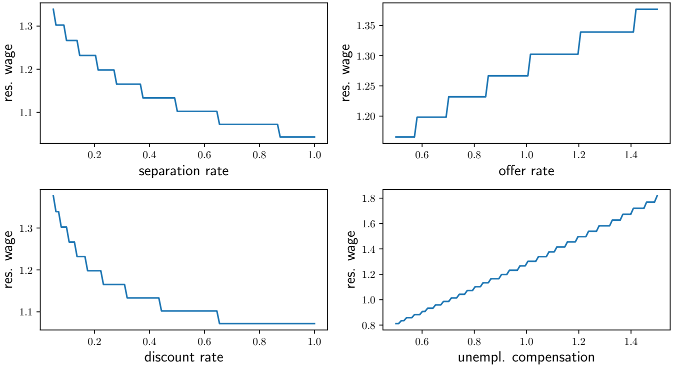


Figure: Continuous time job search reservation wage