Finite State Automata and Markov Chains

Reading

- E.A. Lee and S.A. Seshia, Introduction to Embedded Systems: CPS Approach, Second Edition, MIT Press, 2017
  - Book: https://ptolemy.berkeley.edu/books/leeseshia/releases/LeeSeshia_DigitalV2_2.pdf
  - Chapter 3

- Not exactly a standard DFA chapter, has a dynamical system bias, but similar to MDPs

A Simple Dynamics Model

- Suppose a car is moving in a straight line at $v\ m/s$
- How much will the car have travelled after $T\ s$?

$$vT\ m$$

- Suppose the car's position at time 0 is p_0 and at time T is p_T

$$p_T = p_0 + vT$$

- Suppose every T seconds velocity jumps up by $a\ m/s$
- How do we adapt the model (for discrete times when velocity is changed)?

$$p_{kT} = p_{(k-1)T} + v_{(k-1)T}T$$
$$v_{kT} = v_{(k-1)T} + a$$

  – where $k = 1,2,\ldots$

Elements of a dynamical system model

- Note: notation will change when we get to RL proper
- System has a state, denoted by $\boldsymbol{x} \in \mathbb{R}^n$
  - Captures position, velocity, acceleration, etc.
- Control inputs are denoted by $\boldsymbol{u} \in \mathbb{R}^p$
  - Captures throttle, steering, etc.
- Measurements are denoted by $\boldsymbol{y} \in \mathbb{R}^q$
  - Could measure states directly, e.g., odometry, GPS
  - Could be high-dimensional such as camera, LiDAR

State Evolution

- As time passes, the system state evolves based on the previous state and the current control inputs
- We typically model the state as a signal:

$$\boldsymbol{x}: \mathbb{R}_+ \to \mathbb{R}^n$$

  – i.e., for a given time t, $\boldsymbol{x}(t)$ returns the state at that time

- If we want to model the evolution of $\boldsymbol{x}$ in continuous time, we describe with ordinary differential equations:

$$\dot{\boldsymbol{x}} \coloneqq \frac{\partial \boldsymbol{x}(t)}{\partial t} = f\big(\boldsymbol{x}(t), \boldsymbol{u}(t)\big)$$

- Modern systems are digital, so a discrete-time model makes more sense (since controller is sampled at discrete times)

$$\boldsymbol{x}_{k+1} = f(\boldsymbol{x}_k, \boldsymbol{u}_k)$$

  – where k is incremented with the sampling rate (e.g., 10Hz)

State Evolution Example

- Going back to the position/velocity example:

$$p_{kT} = p_{(k-1)T} + v_{(k-1)T}T$$
$$v_{kT} = v_{(k-1)T} + a$$

- This is a discrete-time model where $\boldsymbol{x} = [p, v]^T, u_k = a$, so

$$f([x_1, x_2], u) = \begin{bmatrix} x_1 + x_2 T \\ x_2 + u \end{bmatrix}$$

- In this case, f is linear, so the system can also be written as

$$\boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}u_k$$

  – where $\boldsymbol{A} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

- Note that we implicitly dropped the T in the subscript
  – It is redundant, since k is chosen for a sampling rate of T

Measurement model

- Measurements are typically modeled as a function of the state:
$$\boldsymbol{y}_k = g(\boldsymbol{x}_k)$$
- In our example, if we can only measure position, then
$$\boldsymbol{y}_k = \boldsymbol{C}\boldsymbol{x}_k$$
  - where $\boldsymbol{C} = [1 \quad 0]$
- In case of more complex measurements, g may be quite complex or (as is often the case) unknown
  - In the F1/10 case, LiDAR measurements can be modeled as a function of the car state and the hallway dimensions
  - Modeling a camera would be significantly harder

General Model Form

- In its most general form, the model can be written as

$$\boldsymbol{x}_{k+1} = f(\boldsymbol{x}_k, \boldsymbol{u}_k)$$
$$\boldsymbol{y}_k = g(\boldsymbol{x}_k)$$

  - This model has the Markov property, i.e., the current state depends only on the previous state and control
    - It doesn't matter how we got to the previous state
- Given f and g, one needs to design a controller $\boldsymbol{u}_k = h(\boldsymbol{y}_k)$
  - E.g., to navigate the track as fast as possible
  - How do we pick the controls $\boldsymbol{u}_k$?
  - Minimize a cost function (surprise, surprise), e.g.,

$$J = \boldsymbol{x}_{k+H}^T \boldsymbol{Q} \boldsymbol{x}_{k+H} + \sum_{j=0}^{H-1} \boldsymbol{x}_{k+j}^T \boldsymbol{Q} \boldsymbol{x}_{k+j} + \boldsymbol{u}_{k+j}^T \boldsymbol{R} \boldsymbol{u}_{k+j} + \boldsymbol{x}_{k+j}^T \boldsymbol{N} \boldsymbol{u}_{k+j}$$

  - where H is a time horizon, $\boldsymbol{Q}, \boldsymbol{R}$ and $\boldsymbol{N}$ are user-defined matrices

Cost Function Considerations

- The cost function

$$J = \boldsymbol{x}_{k+H}^T \boldsymbol{Q} \boldsymbol{x}_{k+H} + \sum_{j=0}^{H-1} \boldsymbol{x}_{k+j}^T \boldsymbol{Q} \boldsymbol{x}_{k+j} + \boldsymbol{u}_{k+j}^T \boldsymbol{R} \boldsymbol{u}_{k+j} + \boldsymbol{x}_{k+j}^T \boldsymbol{N} \boldsymbol{u}_{k+j}$$

  - is known as the linear quadratic regulator (LQR)
  - Can be solved iteratively for linear systems
- Matrices $\boldsymbol{Q}, \boldsymbol{R}$ and $\boldsymbol{N}$ chosen to satisfy control requirements
  - e.g., reach a target, minimize fuel consumption
- Having a horizon allows to plan more complex strategies
  - E.g., mountain car is easily solved
- Optimal control is extremely well studied
  - Strong theory and optimality guarantees for linear systems
  - However, non-linear systems have no general solutions

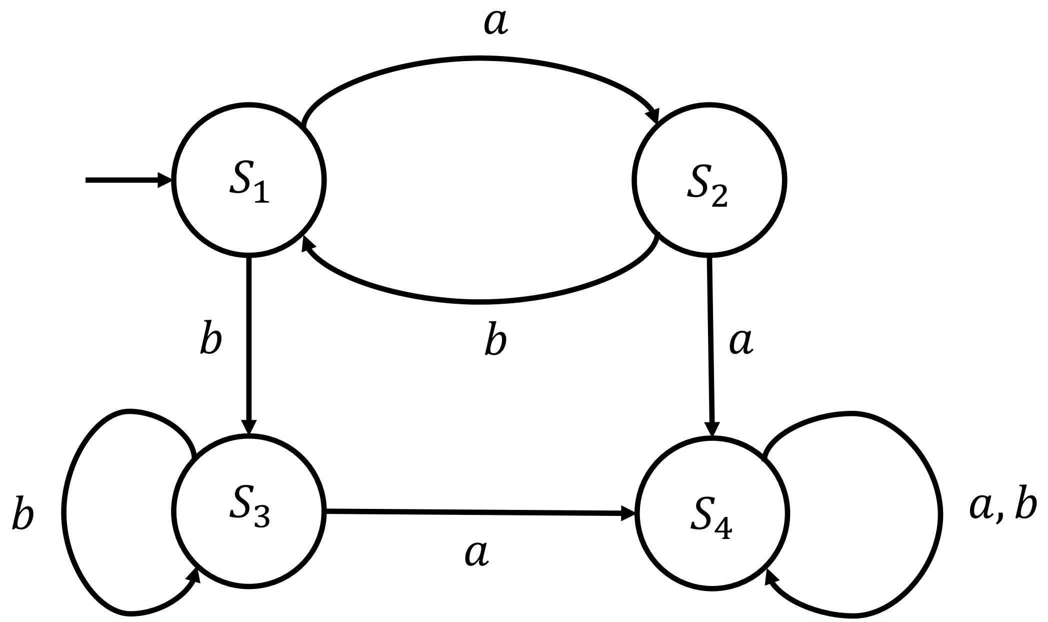
Reinforcement learning modeling and cost function

- Historically, RL theory has been based on finite state models
  - The f and g formulation is infinite-state
  - However, deep RL is increasingly (and surprisingly) able to work in infinite-state settings
  - RL models also have the Markov property
- Unlike optimal control, RL doesn't minimize a cost function
  - It maximizes a reward function
  - Mathematically, there is no difference
  - Maybe RL researchers are young and optimistic O.o

Finite State Machines

Rensselaer

- One of the fundamental models in computer science
- Also known as deterministic finite automata (DFA)
- Historically used to model computer programs
  - DFAs are not a perfect model but have served us well



Finite State Machines Formalization

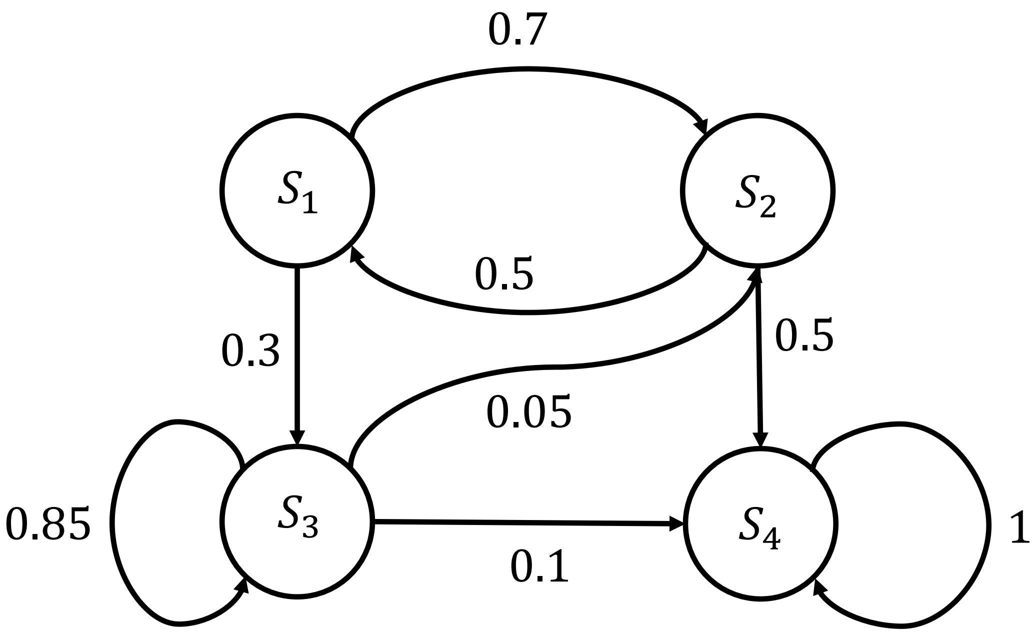
- A DFA is a tuple (A, S, S_0, δ, F), where
  - A is the input alphabet
  - S is the finite set of states
  - S_0 is the initial state
  - $\delta: S \times A \rightarrow S$ is the transition function
  - F is the (possibly empty) set of final (accepting) states
- For each state and input pair S and A, $\delta(S, A)$ outputs exactly one state
  - Hence the deterministic in the name
  - e.g., $\delta(S_1, a) = S_2$
- In a non-deterministic FA (NFA), δ can output 0 or more values
  - Every NFA can be converted to an equivalent DFA

DFA Expressivity

- DFAs are one of the simplest models of computation
  - E.g., simpler than pushdown automata, Turing machines
- At the same time, many problems are just extremely large DFAs
  - E.g., games are for the most part (very large) DFAs
  - E.g., in chess, every position is a state and every input (move/action) causes a transition to exactly one state
- Classical RL was actually developed for stochastic models, not deterministic
  - More expressive than DFAs
- To get there, we need to talk about Markov chains first

Markov Chains

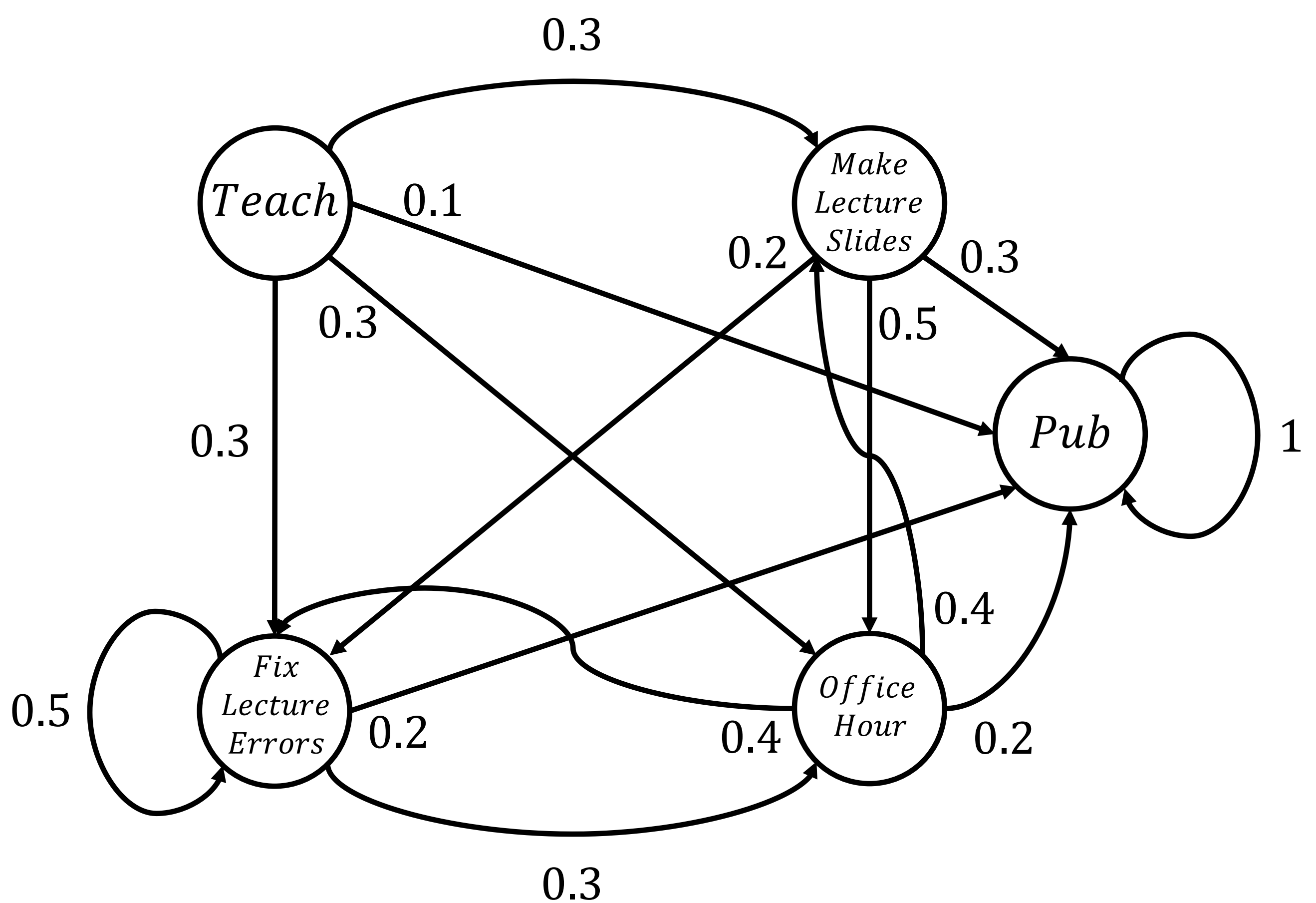
- Markov chains are effectively probabilistic automata
  - Formulation can be made more general, but we'll only need the finite-state version



- Each transition has an associated probability
  - E.g., probability of going from S_1 to S_2 is 0.7

Workday Example

- Markov chain describing my workday



Formalization

- A Markov Chain is a tuple (S, P, η), where
  - S is the finite set of states
  - $P: S \times S \rightarrow \mathbb{R}$ is the probabilistic transition function
  - $\eta: S \rightarrow \mathbb{R}$ is the initial state distribution
- Called Markov chain because the probability of the current state is determined only by the previous state

$$\mathbb{P}[S_t | S_{t-1}, S_{t-2}, \dots, S_0] = \mathbb{P}[S_t | S_{t-1}] = P(S_{t-1}, S_t)$$

  – where S_t denote the state after t steps

  – Examples:

$$\mathbb{P}[S_0 = Teach] = \eta(Teach)$$

$$\mathbb{P}[S_t = Pub | S_{t-1} = Office\ Hour, S_0 = Teach] =$$
$$\mathbb{P}[S_t = Pub | S_{t-1} = Office\ Hour] =$$
$$= 0.2$$

Examples

- What is the probability that I am at Pub two steps after $Teach$?
  - Need to look at all possible ways to get to Pub in two steps
- Formally:

$$\mathbb{P}[S_2 = Pub | S_0 = Teach] =$$
$$\mathbb{P}[S_1 = Pub, S_2 = Pub | S_0 = Teach] +$$
$$\mathbb{P}[S_1 = Office\ Hour, S_2 = Pub | S_0 = Teach] +$$
$$\mathbb{P}[S_1 = Fix\ Lecture\ Errors, S_2 = Pub | S_0 = Teach] +$$
$$\mathbb{P}[S_1 = Make\ Lecture\ Slides, S_2 = Pub | S_0 = Teach]$$

- Summing through all possibilities is called marginalization
- Recall the definition of conditional probability:

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A, B]}{\mathbb{P}[B]}$$

Examples

- What is the probability that I am at Pub two steps after $Teach$?
  - Need to look at all possible ways to get to Pub in two steps
- Formally:

$$\mathbb{P}[S_2 = Pub | S_0 = Teach] =$$
$$\mathbb{P}[S_1 = Pub, S_2 = Pub | S_0 = Teach] +$$
$$\mathbb{P}[S_1 = Office\ Hour, S_2 = Pub | S_0 = Teach] +$$
$$\mathbb{P}[S_1 = Fix\ Lecture\ Errors, S_2 = Pub | S_0 = Teach] +$$
$$\mathbb{P}[S_1 = Make\ Lecture\ Slides, S_2 = Pub | S_0 = Teach]$$

- Summing through all possibilities is called marginalization
- Probabilities are:

$$\mathbb{P}[S_1 = Pub, S_2 = Pub | S_0 = Teach] =$$
$$\mathbb{P}[S_2 = Pub | S_1 = Pub, S_0 = Teach] \mathbb{P}[S_1 = Pub | S_0 = Teach] =$$
$$\mathbb{P}[S_2 = Pub | S_1 = Pub] \mathbb{P}[S_1 = Pub | S_0 = Teach] = \ 0.1$$

Examples, cont'd

- What is the probability that I am at Pub two steps after $Teach$?
  - Need to look at all possible ways to get to Pub in two steps
- All possible paths
  - Pub, Pub
  - $Office\ Hour, Pub$
  - $Fix\ Lecture\ Errors, Pub$
  - $Make\ Lecture\ Slides, Pub$
- Probabilities are:

$$\mathbb{P}[S_1 = Office\ Hour, S_2 = Pub | S_0 = Teach] = 0.06$$
$$\mathbb{P}[S_1 = Fix\ Lecture\ Errors, S_2 = Pub | S_0 = Teach] = 0.06$$
$$\mathbb{P}[S_1 = Make\ Lecture\ Slides, S_2 = Pub | S_0 = Teach] = 0.09$$

- Total probability is $0.1 + 0.06 + 0.06 + 0.09 = 0.31$

Linear Algebra Aside

- Suppose we are given a square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$
- A vector $\boldsymbol{v}$ is said to be an eigenvector of $\boldsymbol{A}$ if

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}$$

  – Where $\lambda \in \mathbb{R}$ is a corresponding eigenvalue
- The matrix $\boldsymbol{A}$ has n eigenvectors, $\boldsymbol{v}_i$
  – And n corresponding eigenvalues, λ_i
- If eigenvalues are distinct, the eigenvectors form a basis in $\mathbb{R}^n$
  – i.e., any $\boldsymbol{x} \in \mathbb{R}^n$ can be written as a linear combination

$$\boldsymbol{x} = c_1 \boldsymbol{v}_1 + \cdots + c_n \boldsymbol{v}_n$$

- There may be repeated eigenvalues
- $\boldsymbol{A}$ is full rank iff $\lambda_i \neq 0$ for all i

Transition Matrix

Rensselaer

- We can store all transition probabilities in a matrix $\boldsymbol{P}$
- Entry P_{ij} denotes the probability of going from state i to j
- E.g., let states be ordered:
  $Teach, Office\ Hour, MLS, FLE, Pub$
- The transition matrix becomes:

$$\boldsymbol{P} = \begin{bmatrix} 0 & 0.3 & 0.3 & 0.3 & 0.1 \\ 0 & 0 & 0.4 & 0.4 & 0.2 \\ 0 & 0.5 & 0 & 0.2 & 0.3 \\ 0 & 0.3 & 0 & 0.5 & 0.2 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

- What properties does $\boldsymbol{P}$ have?

Transition Matrix Properties

- Each row must sum up to 1
  - Why?
  - For each state, transition probabilities mush sum up to 1
- Has an eigenvalue of 1
  - Why? What is the corresponding eigenvector?
  - Pick any row, p_j^T
  - Let $\mathbf{1} \in \mathbb{R}^{|S|}$ be a vector of all ones
  - What is $p_j^T \mathbf{1}$?
    - 1! So $\mathbf{1}$ is an eigenvector

Transition Matrix Properties, cont'd

Rensselaer

- Let $\boldsymbol{\eta}_t$ represent the probabilities the system is in any given state at time t
  - E.g., $\boldsymbol{\eta}_t = [1 \quad 0 \quad 0 \quad 0 \quad 0]^T$ means the state is $Teach$
- What happens if we multiply $\boldsymbol{\eta}_t^T \boldsymbol{P}$?

$$[1 \quad 0 \quad 0 \quad 0 \quad 0]\begin{bmatrix} 0 & 0.3 & 0.3 & 0.3 & 0.1 \\ 0 & 0 & 0.4 & 0.4 & 0.2 \\ 0 & 0.5 & 0 & 0.2 & 0.3 \\ 0 & 0.3 & 0 & 0.5 & 0.2 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} =$$

$$= [0 \quad 0.3 \quad 0.3 \quad 0.3 \quad 0.1]$$

- We get the distribution of states after one step, i.e., $\boldsymbol{\eta}_{t+1}^T$
  - What happens if we multiply $\boldsymbol{\eta}_t^T \boldsymbol{P}\boldsymbol{P}$?

Transition Matrix Properties, cont'd

- What happens if we multiply $\boldsymbol{\eta}_t^T \boldsymbol{P}\boldsymbol{P}$?

$$\boldsymbol{\eta}_t^T \boldsymbol{P}\boldsymbol{P} = \boldsymbol{\eta}_{t+1}^T \boldsymbol{P} = \boldsymbol{\eta}_{t+2}^T$$

- Now suppose you are given $\boldsymbol{\eta}_0$
  - The distribution at time 0
- How do you express $\boldsymbol{\eta}_t$ as a function of $\boldsymbol{\eta}_0$ and $\boldsymbol{P}$?

$$\boldsymbol{\eta}_t^T = \boldsymbol{\eta}_0^T \boldsymbol{P}^t$$

  - Can quickly compute state distributions over time
- What does this expression remind you of?
  - It's a linear system!

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{A}\boldsymbol{x}_k \\ \boldsymbol{\eta}_{t+1} &= \boldsymbol{P}^T \boldsymbol{\eta}_t \\ &= \left(\boldsymbol{P}^T\right)^{t+1} \boldsymbol{\eta}_0 \end{aligned}$$

Linear Algebra Aside, cont'd

- Suppose a square matrix $\boldsymbol{A}$ has eigenvalues $\lambda_1, \dots, \lambda_n$
- What are the eigenvalues of $\boldsymbol{A}^2$?

$$\lambda_1^2, \dots, \lambda_n^2$$

- Take any eigenvalue λ_i and corresponding eigenvector $\boldsymbol{v}_i$

$$\begin{aligned}\boldsymbol{A}\boldsymbol{A}\boldsymbol{v}_i &= \boldsymbol{A}\lambda_i\boldsymbol{v}_i \\ &= \lambda_i^2\boldsymbol{v}_i\end{aligned}$$

- In general, the eigenvalues of $\boldsymbol{A}^k$ are

$$\lambda_1^k, \dots, \lambda_n^k$$

  - The eigenvectors are the same as those of $\boldsymbol{A}$

Linear System Solution

- Consider a general discrete-time linear system

$$\boldsymbol{x}_k = \boldsymbol{A}^k \boldsymbol{x}_0$$

- Suppose $\boldsymbol{A}$ has distinct eigenvalues for simplicity
- Recall that the eigenvectors of $\boldsymbol{A}$ form a basis in $\mathbb{R}^n$, so

$$\boldsymbol{x}_0 = a_1 \boldsymbol{v}_1 + \cdots + a_n \boldsymbol{v}_n$$

- Then

$$\boldsymbol{A}^k \boldsymbol{x}_0 = a_1 \lambda_1^k \boldsymbol{v}_1 + \cdots + a_n \lambda_n^k \boldsymbol{v}_n$$

- Under what conditions does $\boldsymbol{x}_k$ converge to $\boldsymbol{0}$?

$|\lambda_i| < 1$, for all i

Transition Matrix Linear Systems

- Consider the transition matrix linear system

$$\boldsymbol{\eta}_{t+1} = \left(\boldsymbol{P}^T\right)^{t+1} \boldsymbol{\eta}_0$$

- We know that $\mathbf{1}$ is an eigenvector of $\boldsymbol{P}$
  - Also known as a right eigenvector
  - However, we are now interested in left eigenvectors
    - AKA eigenvectors of $\boldsymbol{P}^T$

$$\boldsymbol{v}^T \boldsymbol{P} = \left(\boldsymbol{P}^T \boldsymbol{v}\right)^T$$

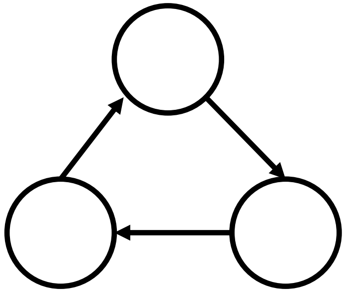
- It turns out that $\boldsymbol{P}$ also has a left eigenvalue of 1
  - Left and right eigenvalues are the same for square matrices
    - Eigenvectors may be different
- This means the system never converges to 0
  - But what does it converge to?

Transition Matrix Linear Systems, cont'd

- Consider a row vector $\boldsymbol{\mu}$ such that

$$\boldsymbol{\mu P} = \boldsymbol{\mu}$$

- Then $\boldsymbol{\mu}$ is an eigenvector corresponding to eigenvalue 1
  - There could be more than 1 such vectors
- What graph property determines whether there is a unique $\boldsymbol{\mu}$?
  - There is one $\boldsymbol{\mu}$ per closed communication class
    - i.e., loop in the graph that cannot be left
  - Formally, having one such class is known as irreducibility
- Another requirement is aperiodicity
  - If you have a periodic graph, you will never converge

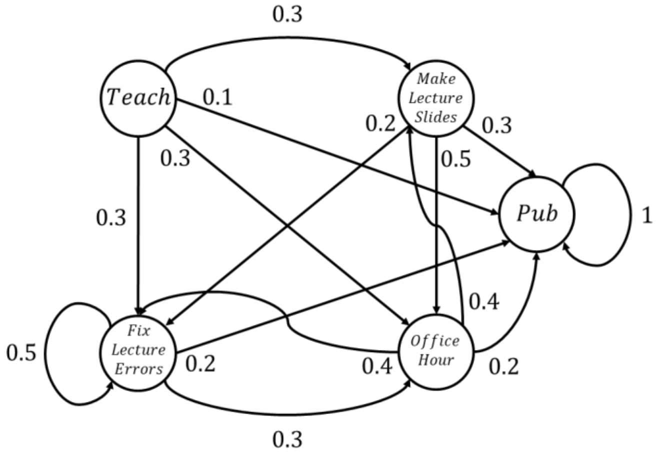


Stationary Distribution

- If you have an aperiodic, irreducible Markov chain, then there is a unique $\boldsymbol{\mu}$ such that

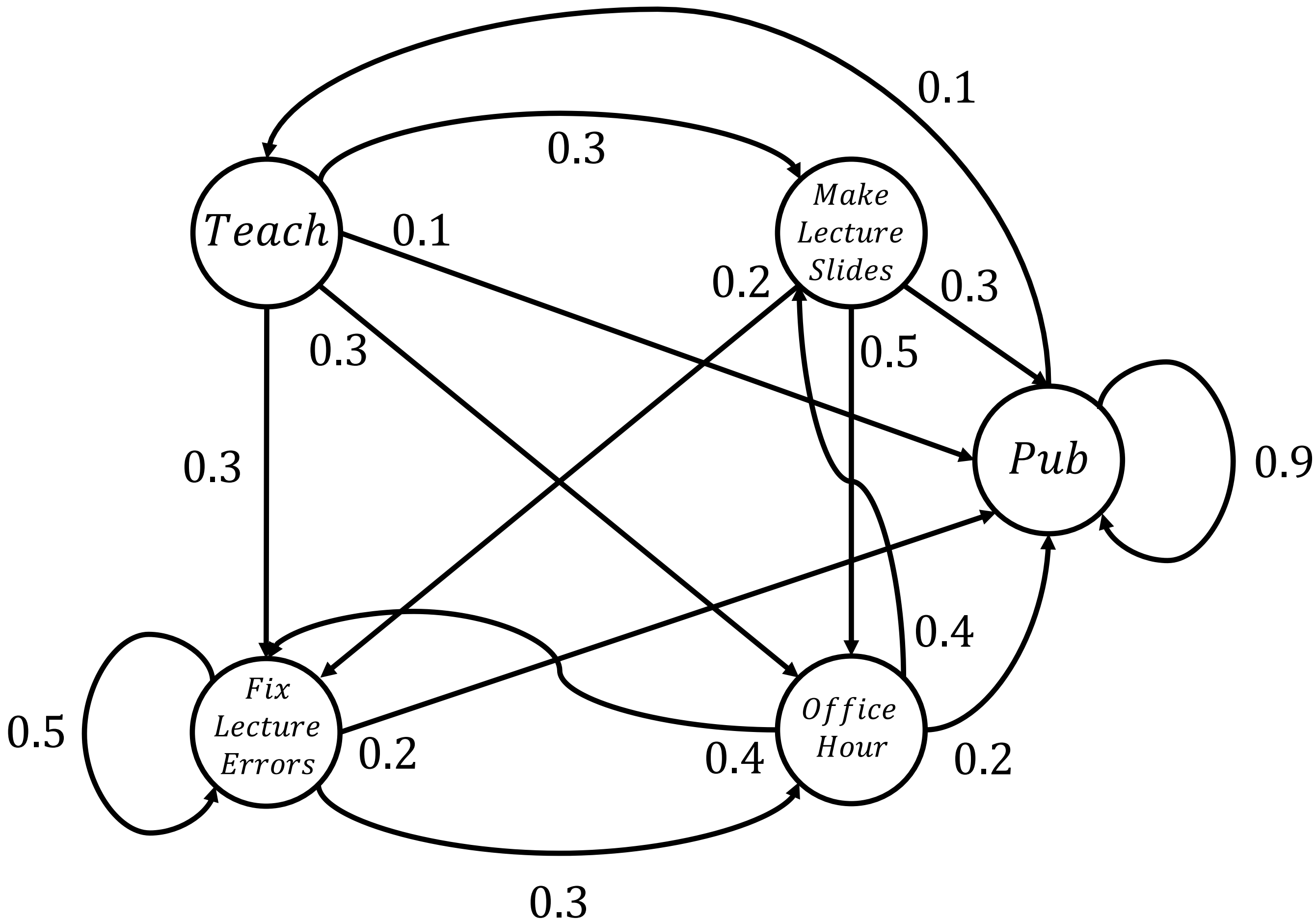
$$\boldsymbol{\mu P} = \boldsymbol{\mu}$$

- This is known as the stationary distribution
  - Each element of $\boldsymbol{\mu}$ denotes the *proportion* of time spent in that state in the long run
- What is $\boldsymbol{\mu}$ for the workday example?
  - It is $[0 \quad 0 \quad 0 \quad 0 \quad 1]$
  - Pub is an absorbing state
  - Every trace eventually gets to Pub
- $Teach$ is a transient state
  - You cannot go back to it



Workday Example, revisited

- Suppose I add a new transition from Pub to $Teach$



Stationary Distribution for Revisited Workday Example

- Stationary distribution is hard to derive by simply looking at the graph anymore
- Two ways of finding $\boldsymbol{\mu}$
  - Can either find left eigenvalues and eigenvectors of $\boldsymbol{P}$
    - Which eigenvalue does $\boldsymbol{\mu}$ correspond to?

      1
    - Might need to normalize eigenvector
  - Or just compute $\boldsymbol{P}^t$ for a big t and then compute $\boldsymbol{\eta}_0\boldsymbol{P}^t$
    - Recall $\boldsymbol{\mu}$ is the same for any initial $\boldsymbol{\eta}_0$
- For the revisited example

$$\boldsymbol{\mu} = [0.067 \quad 0.086 \quad 0.054 \quad 0.13 \quad 0.66]$$

- Still spending most time in Pub, but other states are also visited infinitely often