Markov Chains

Geometric Algorithms
Lecture 12

CAS CS 132

Objectives

1. Motivate linear dynamical systems
2. Analyze Markov chains and their properties
3. Learn to solve for steady-states of Markov chains
4. Connect this to graphs and random walks

Keywords

linear dynamical systems

recurrence relations

linear difference equations

state vector

probability vector

stochastic matrix

Markov chain

steady-state vector

random walk

state diagram

Motivation

Change (or Waxing Poetic)

Change (or Waxing Poetic)

```
Things change.
```

Change (or Waxing Poetic)

Things change.

Things change from one state of affairs to another state of affairs.

Change (or Waxing Poetic)

Things change.

Things change from one state of affairs to another state of affairs.

Things change often in unpredictable ways.

Change (or Waxing Poetic)

Things change.

Things change from one state of affairs to another state of affairs.

Things change often in unpredictable ways.

If something changes unpredictably, what can we say about it?

Dynamical Systems

Dynamical Systems

Definition (Informal). A **dynamical system** is a thing (typically with interacting parts) that changes over time.

Dynamical Systems

Definition (Informal). A **dynamical system** is a thing (typically with interacting parts) that changes over time.

A dynamical system has *possible states* which it can be in as time elapses and its behavior is defined by a *evolution function*.

Dynamical Systems

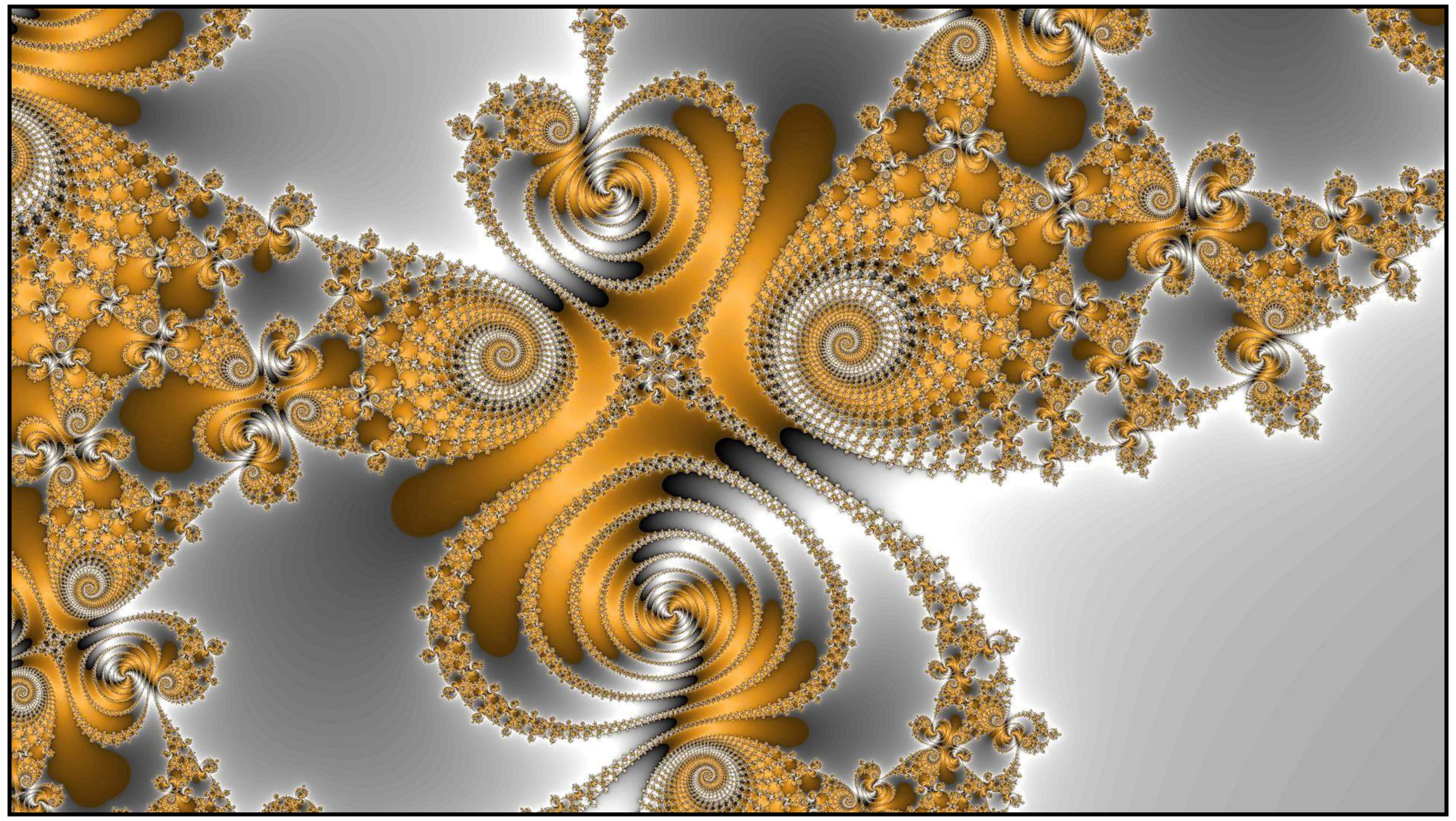
Definition (Informal). A **dynamical system** is a thing (typically with interacting parts) that changes over time.

A dynamical system has *possible states* which it can be in as time elapses and its behavior is defined by a *evolution function*.

Examples.

» economics (stocks)
» physical/chemical systems
» populations
» weather

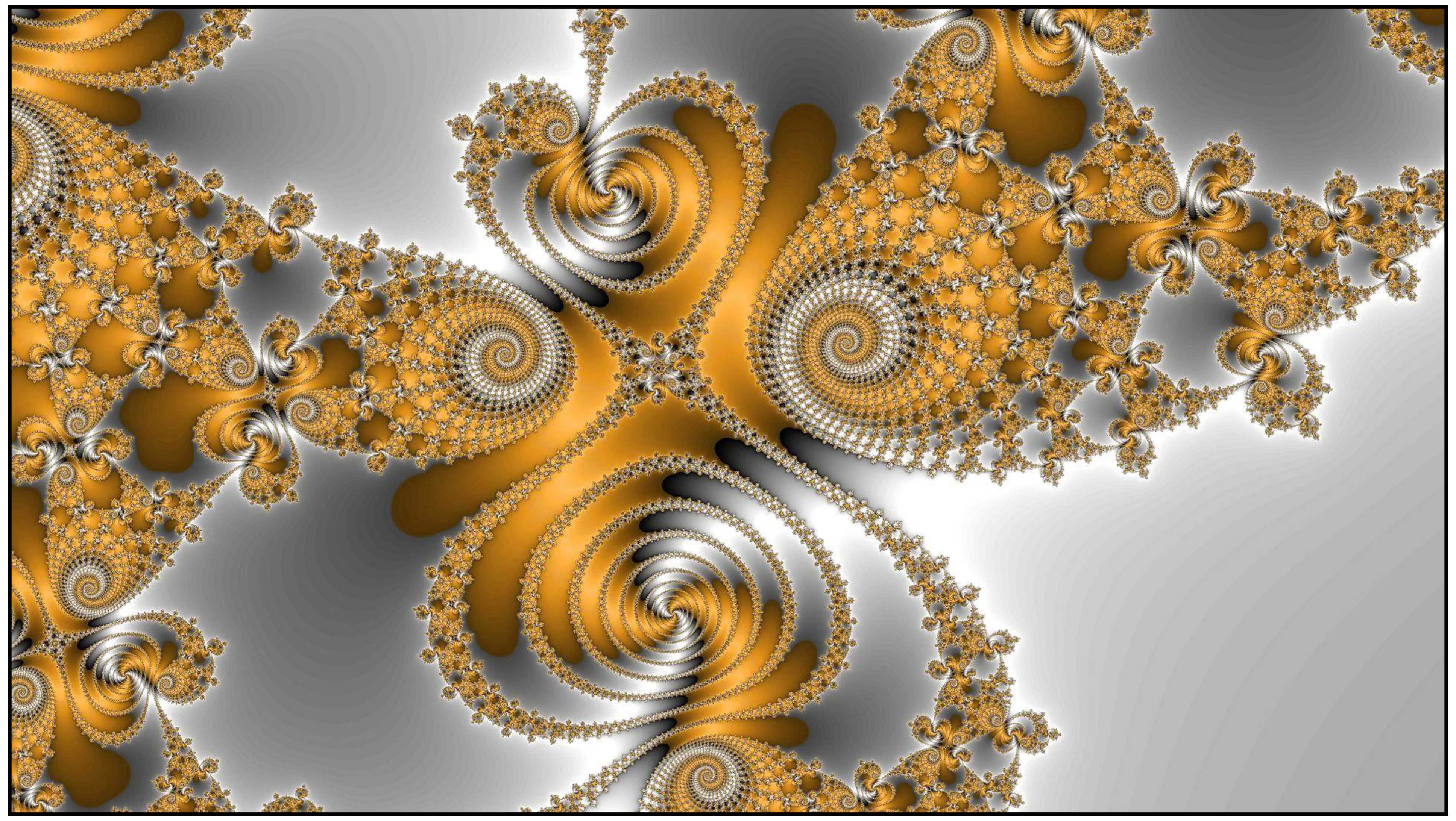
An Aside: Chaos Theory



https://commons.wikimedia.org/wiki/File:Fr137.jpg

An Aside: Chaos Theory

Complex systems like the weather or the economy look nearly random.

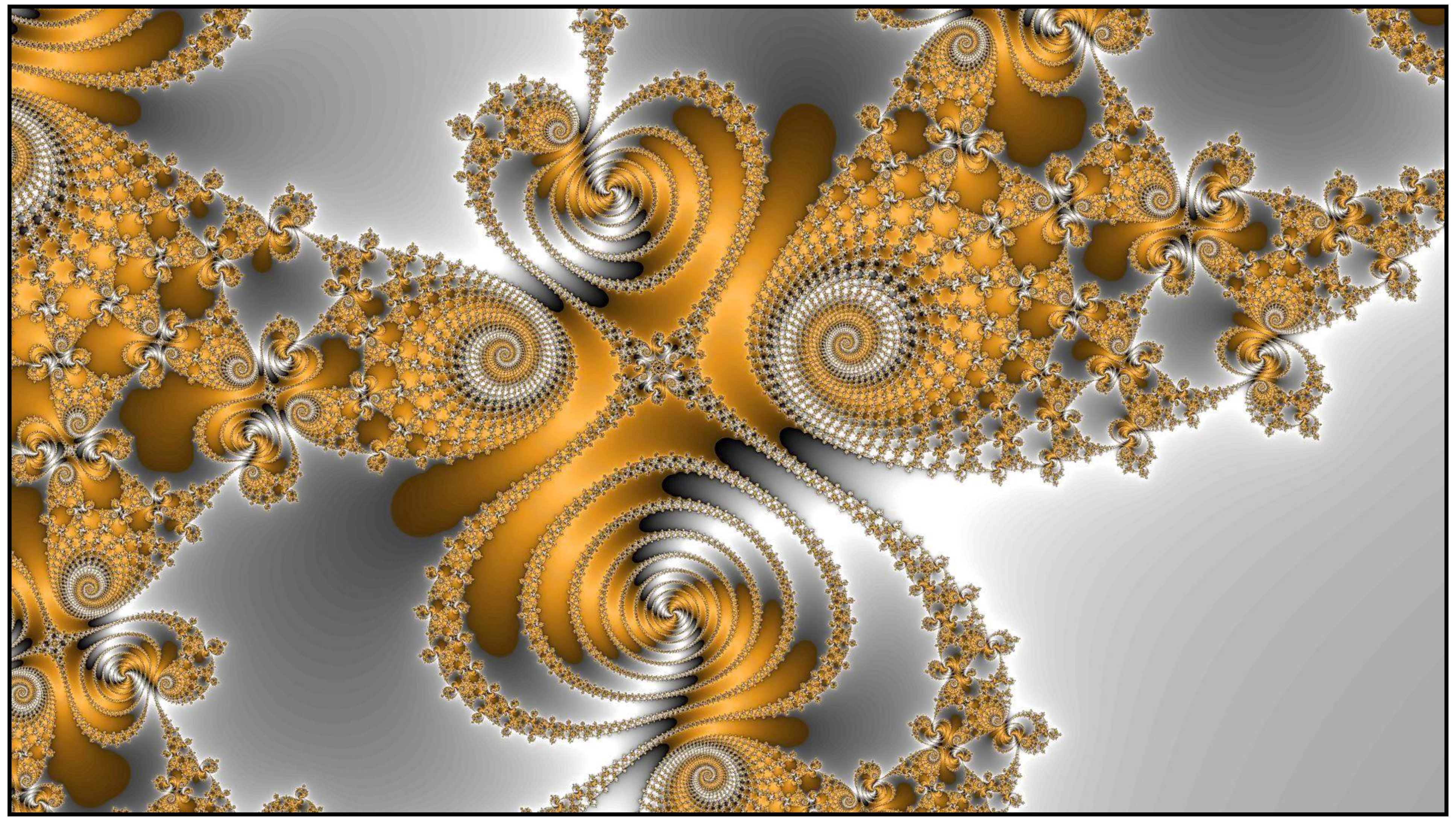


https://commons.wikimedia.org/wiki/File:Fr137.jpg

An Aside: Chaos Theory

Complex systems like the weather or the economy look nearly random.

But even in chaotic systems there are *underlying patterns* and *repeated structures*.



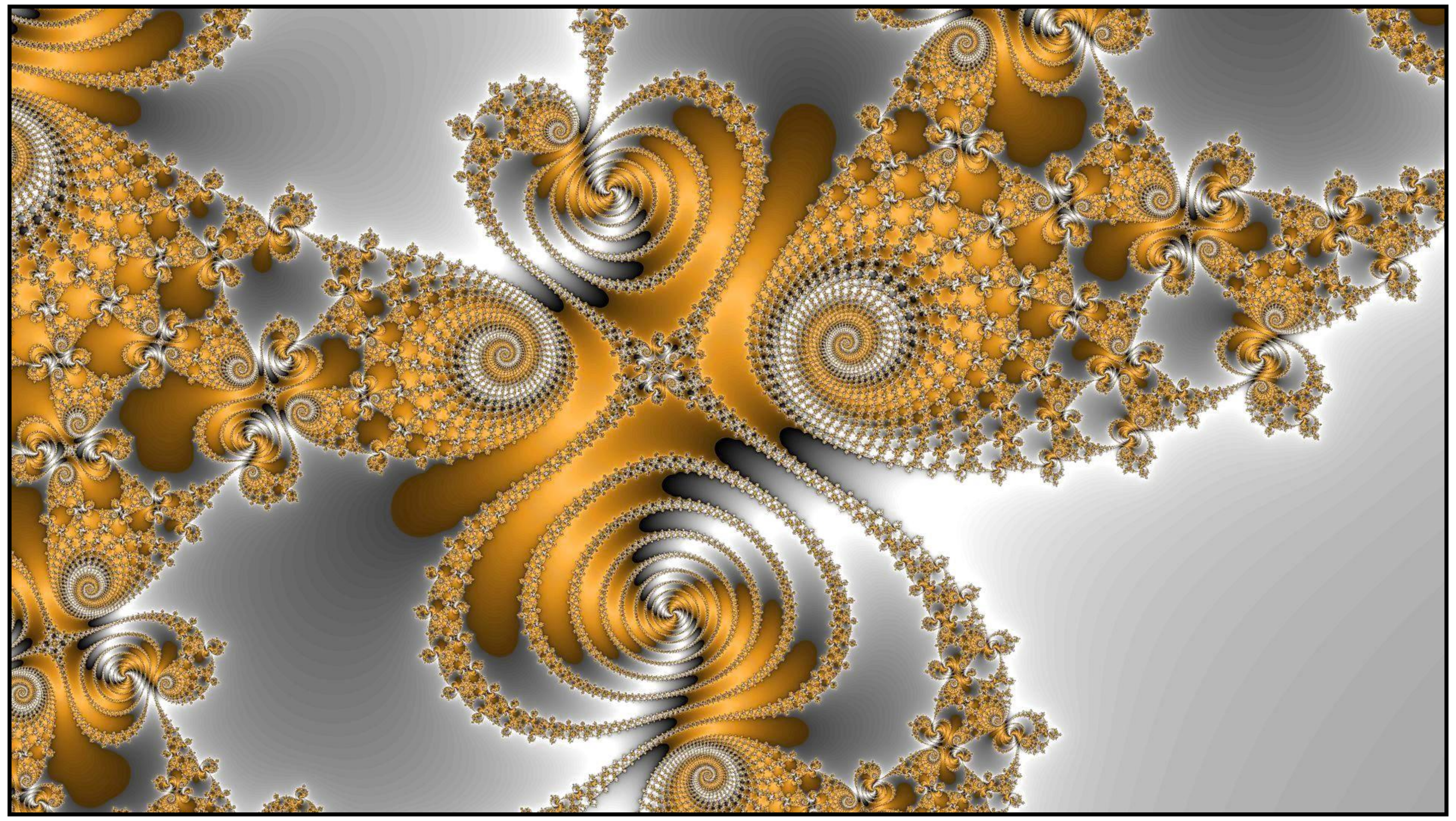
https://commons.wikimedia.org/wiki/File:Fr137.jpg

An Aside: Chaos Theory

Complex systems like the weather or the economy look nearly random.

But even in chaotic systems there are *underlying patterns* and *repeated structures*.

Often it's useful to consider chaotic systems in terms of global properties.



https://commons.wikimedia.org/wiki/File:Fr137.jpg

Motivating Questions

Motivating Questions

What does a dynamical system look like "in the long view?"

Motivating Questions

What does a dynamical system look like "in the long view?"

Does it reach a kind of equilibrium? (think heat diffusion)

Motivating Questions

What does a dynamical system look like "in the long view?"

Does it reach a kind of equilibrium? (think heat diffusion)

Or does some part of the system dominate over time? (think the population of rabbits without a predator)

(Linear) Dynamical Systems

Linear Dynamical Systems

Linear Dynamical Systems

Definition. A **(discrete time) linear dynamical system** is a described a $n \times n$ matrix A. It's **evolution function** is the matrix transformation $\mathbf{x} \mapsto A\mathbf{x}$.

Linear Dynamical Systems

Definition. A **(discrete time) linear dynamical system** is a described a $n \times n$ matrix A. It's **evolution function** is the matrix transformation $\mathbf{x} \mapsto A\mathbf{x}$.

The possible states of the system are vectors in $\mathbb{R}^n$.

Linear Dynamical Systems

Definition. A **(discrete time) linear dynamical system** is a described a $n \times n$ matrix A. It's **evolution function** is the matrix transformation $\mathbf{x} \mapsto A\mathbf{x}$.

The possible states of the system are vectors in $\mathbb{R}^n$.

Given an **initial state vector** $\mathbf{v}_0$, we can determine the **state vector** of the system after i time steps:

$$\mathbf{v}_i = A\mathbf{v}_{i-1}$$

Linear Dynamical Systems

Definition. A **(discrete time) linear dynamical system** is a described a $n \times n$ matrix A. It's **evolution function** is the matrix transformation $\mathbf{x} \mapsto A\mathbf{x}$.

The possible states of the system are vectors in $\mathbb{R}^n$.

A tells us how our system evolves over time.

Given an **initial state vector** $\mathbf{v}_0$, we can determine the **state vector** of the system after i time steps:

$$\mathbf{v}_i = A\mathbf{v}_{i-1}$$

State Vectors

$$\begin{aligned}
\mathbf{v}_1 &= A\mathbf{v}_0 \\
\mathbf{v}_2 &= A\mathbf{v}_1 = A(A\mathbf{v}_0) \\
\mathbf{v}_3 &= A\mathbf{v}_2 = A(AA\mathbf{v}_0) \\
\mathbf{v}_4 &= A\mathbf{v}_3 = A(AAA\mathbf{v}_0) \\
\mathbf{v}_5 &= A\mathbf{v}_4 = A(AAAA\mathbf{v}_0) \\
&\vdots
\end{aligned}$$

The state vector $\mathbf{v}_k$ tells us what the system looks like after a number k time steps.

This is also called a *recurrence relation* or a *linear difference function*.

How to: Determining State Vectors

Question. Determine the state vector $\mathbf{v}_i$ for the linear dynamical system with matrix A given the initial state vector $\mathbf{v}_0$.

Solution. Compute

$$\mathbf{v}_i = A^i \mathbf{v}_0$$

Matrix Powers in NumPy

```
numpy.linalg.matrix_power(a)
```

Matrix Powers in NumPy

```
numpy.linalg.matrix_power(a)
```

There is a function in NumPy for doing matrix powers.

Matrix Powers in NumPy

```
numpy.linalg.matrix_power(a)
```

There is a function in NumPy for doing matrix powers.

Use can use this when you need to take a large power of a matrix.

Matrix Powers in NumPy

`numpy.linalg.matrix_power(a)`

There is a function in NumPy for doing matrix powers.

Use can use this when you need to take a large power of a matrix.

It's much faster than doing each multiplication individually because it uses the "repeated squaring" trick

Matrix Powers in NumPy

```
numpy.linalg.matrix_power(a)
```

There is a function in NumPy for doing matrix powers.

Use can use this when you need to take a large power of a matrix.

It's much faster than doing each multiplication individually because it uses the "repeated squaring" trick

But be cautious of floating-point error.

Warm up: Population Dynamics

The Setup

The Setup

We're working for the census. We have 2023 population measurements for a <u>city</u> and a <u>suburb</u> which are geographically coincident.

The Setup

We're working for the census. We have 2023 population measurements for a city and a suburb which are geographically coincident.

We find by analyzing previous data that each year:

- 5% of the population moves from city → suburb
- 3% of the population moves from suburb → city

Fundamental Question

Can we make any predictions about the population of the city and suburb in 2043?

Note: No immigration, emigration, birth, death, etc. **The overall population stays fixed.**

Setting up Linear Equations

Setting up Linear Equations

If $city_0$ = 2023 city pop. = 600,000
and $suburb_0$ = 2023 suburb pop. = 400,000

Setting up Linear Equations

If city_0 = 2023 city pop. = 600,000
and suburb_0 = 2023 suburb pop. = 400,000

then the pop. in 2024 are given by:

$$\text{city}_1 = (0.95)\text{city}_0 + (0.03)\text{suburb}_0$$

$$\text{suburb}_1 = (0.05)\text{city}_0 + (0.97)\text{suburb}_0$$

Setting up Linear Equations

If city_0 = 2023 city pop. = 600,000
and suburb_0 = 2023 suburb pop. = 400,000

then the pop. in 2024 are given by:

$$\text{city}_1 = (0.95)\text{city}_0 + (0.03)\text{suburb}_0$$
$$\text{suburb}_1 = (0.05)\text{city}_0 + (0.97)\text{suburb}_0$$

people who stayed

people who left

Setting up a Matrix

$$\begin{bmatrix} \text{city}_1 \\ \text{suburb}_1 \end{bmatrix} = \begin{bmatrix} 0.95 & 0.3 \\ 0.05 & 0.97 \end{bmatrix} \begin{bmatrix} \text{city}_0 \\ \text{suburb}_0 \end{bmatrix} = \begin{bmatrix} 582,000 \\ 418,000 \end{bmatrix}$$

```
In 2024, we expect the population of the city
to decrease.
```

Setting up a Matrix

$$\begin{bmatrix} \text{city}_2 \\ \text{suburb}_2 \end{bmatrix} = \begin{bmatrix} 0.95 & 0.3 \\ 0.05 & 0.97 \end{bmatrix} \begin{bmatrix} \text{city}_1 \\ \text{suburb}_1 \end{bmatrix} = \begin{bmatrix} 565{,}440 \\ 434{,}560 \end{bmatrix}$$

In 2025, we expect the population of the city to *continue* to decrease.

Will it decrease indefinitely?

Setting up a Matrix

$$\begin{bmatrix} \text{city}_k \\ \text{suburb}_k \end{bmatrix} = \begin{bmatrix} 0.95 & 0.3 \\ 0.05 & 0.97 \end{bmatrix} \begin{bmatrix} \text{city}_{k-1} \\ \text{suburb}_{k-1} \end{bmatrix}$$

This is a *linear dynamical system.*

So we want to guess what the population will look like in 20 years, we need to compute

$$\begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix}^{20} \begin{bmatrix} \text{city}_0 \\ \text{suburb}_0 \end{bmatrix}$$

demo

Markov Chains

Stochastic Matrices

$$\begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix}$$

What's special about this matrix?

- Its entries are nonnegative.
- Its columns sum to 1.

This should make us think probability.

Stochastic Matrices

Definition. A $n \times n$ matrix is **stochastic** if its entries are nonnegative and its columns sum to 1.

Example.

$$\begin{bmatrix} 0.7 & 0.1 & 0.3 \\ 0.2 & 0.8 & 0.3 \\ 0.1 & 0.1 & 0.4 \end{bmatrix}$$

Markov Chains

Definition. A **Markov chain** is a linear dynamical system whose evolution function is given by a stochastic matrix.

(We can construct a "chain" of state vectors, where each state vector only depends on the one before it.)

Key Property of Stochastic Matrices

Key Property of Stochastic Matrices

Stochastic matrices redistribute the "stuff" in a vector.

Key Property of Stochastic Matrices

Stochastic matrices redistribute the "stuff" in a vector.

Theorem. For a stochastic matrix A and a vector $\mathbf{v}$,

$$\text{sum of entries of } \mathbf{v} \; = \; \text{sum of entries of } A\mathbf{v}$$

Key Property of Stochastic Matrices

The sum of the entries of $\mathbf{v}$ can be computed as

$$\mathbf{1}^T\mathbf{v} = \langle \mathbf{1}, \mathbf{v} \rangle$$

So the previous statement can be written

$$\mathbf{1}^T(A\mathbf{v}) = \mathbf{1}^T\mathbf{v}$$

Key Property of Stochastic Matrices

$$\mathbf{1}^T(A\mathbf{v}) = \mathbf{1}^T\mathbf{v}$$

A is stochastic

Let's verify this:

(I'll leave it as an exercise)

More General Solutions

More General Solutions

In our example, we analyzed the dynamics of a *particular* population.

More General Solutions

In our example, we analyzed the dynamics of a *particular* population.

What if we're interested more generally in the behavior of the process for *any* population?

More General Solutions

In our example, we analyzed the dynamics of a *particular* population.

What if we're interested more generally in the behavior of the process for *any* population?

We need to shift from a population vector to a population ***distribution*** vector.

Returning to the Example

$$\begin{bmatrix} \text{city}_k \\ \text{suburb}_k \end{bmatrix} = \begin{bmatrix} 0.95 & 0.3 \\ 0.05 & 0.97 \end{bmatrix} \begin{bmatrix} \text{city}_{k-1} \\ \text{suburb}_{k-1} \end{bmatrix}$$

Returning to the Example

$$\begin{bmatrix} \text{city}_k \\ \text{suburb}_k \end{bmatrix} = \begin{bmatrix} 0.95 & 0.3 \\ 0.05 & 0.97 \end{bmatrix}^k \begin{bmatrix} \text{city}_0 \\ \text{suburb}_0 \end{bmatrix}$$

Returning to the Example

$$\begin{bmatrix} \text{city}_k \\ \text{suburb}_k \end{bmatrix} = \begin{bmatrix} 0.95 & 0.3 \\ 0.05 & 0.97 \end{bmatrix}^k \begin{bmatrix} 600{,}000 \\ 400{,}000 \end{bmatrix}$$

Returning to the Example

$$\begin{bmatrix} \text{city}_k \\ \text{suburb}_k \end{bmatrix} = \begin{bmatrix} 0.95 & 0.3 \\ 0.05 & 0.97 \end{bmatrix}^k \begin{bmatrix} 600{,}000 \\ 400{,}000 \end{bmatrix}$$

```
But what if we start of with a different
population?
```

Returning to the Example

$$\begin{bmatrix} \text{city}_k \\ \text{suburb}_k \end{bmatrix} = \begin{bmatrix} 0.95 & 0.3 \\ 0.05 & 0.97 \end{bmatrix}^k \begin{bmatrix} 600{,}000 \\ 400{,}000 \end{bmatrix}$$

But what if we start of with a different population?

Do we have to do all our work over again?

Returning to the Example

$$\begin{bmatrix} \text{city}_k \\ \text{suburb}_k \end{bmatrix} = \begin{bmatrix} 0.95 & 0.3 \\ 0.05 & 0.97 \end{bmatrix}^k \begin{bmatrix} 0.6 \\ 0.4 \end{bmatrix}$$

`60% of pop. in city`

`40% of pop. in suburb`

`Not really.`

`But rather than thinking in terms of populations, we need to think about` **`how the population is distributed`**`.`

Probability Vectors

Probability Vectors

Definition. A **probability vector** is a vector whose entries sum to 1.

Probability Vectors

Definition. A **probability vector** is a vector whose entries sum to 1.

They represent

- discrete probability distributions
- distributions of collections of things

Probability Vectors

Definition. A **probability vector** is a vector whose entries sum to 1.

They represent

- discrete probability distributions
- distributions of collections of things

These are really the same thing.

Probability Vectors (Example)

The vector $\begin{bmatrix} 1/3 \\ 1/6 \\ 1/2 \end{bmatrix}$ represents the distribution

where we choose:

1 with probability $1/3$

2 with probability $1/6$

3 with probability $1/2$

Probability Vectors (Example)

The vector $\begin{bmatrix} 0.6 \\ 0.4 \end{bmatrix}$ represented the distribution of the population, but we can also think of this as:

If we choose a random person from the population we'll get someone:

in the city with probability 0.6

in the suburbs with probability 0.4

The point

The point

We'll be interested in the dynamics of Markov chains on probability vectors.

The point

We'll be interested in the dynamics of Markov chains on probability vectors.

Since stochastic matrices preserve $\mathbf{1}^T\mathbf{v}$, they *transform* one distribution into another.

The point

We'll be interested in the dynamics of Markov chains on probability vectors.

Since stochastic matrices preserve $\mathbf{1}^T\mathbf{v}$, they *transform* one distribution into another.

Can we say something about how the distribution changes in the long run?

Steady-State Vectors

Steady-State Vectors

Definition. A **steady-state vector** for a stochastic matrix A is a probability vector $\mathbf{q}$ such that

$$A\mathbf{q} = \mathbf{q}$$

A steady-state vector is *not changed* by the stochastic matrix. They describe equilibrium distributions.

Returning to the Example

Returning to the Example

How do we interpret a steady-state vector for our example?

Returning to the Example

How do we interpret a steady-state vector for our example?

The populations in the city and the suburb stay the same over time.

Returning to the Example

How do we interpret a steady–state vector for our example?

The populations in the city and the suburb stay the same over time.

The same number of people are moving into and out of the city each year.

Fundamental Questions

Do steady states exist?

Are they unique?

How do we find them?

Finding Steady-State Vectors

$$A\mathbf{q} = \mathbf{q}$$

Let's solve this equation for $\mathbf{q}$.

Finding Steady-State Vectors

$$A\mathbf{q} - \mathbf{q} = \mathbf{0}$$

Let's solve this equation for $\mathbf{q}$.

Finding Steady-State Vectors

$$A\mathbf{q} - I\mathbf{q} = \mathbf{0}$$

Let's solve this equation for $\mathbf{q}$.

Finding Steady-State Vectors

$$(A - I)\mathbf{q} = \mathbf{0}$$

Let's solve this equation for $\mathbf{q}$.

Finding Steady-State Vectors

$$(A - I)\mathbf{q} = \mathbf{0}$$

Let's solve this equation for $\mathbf{q}$.

This is a matrix equation.
So we know how to solve it.

How to: Steady-State Vectors

How to: Steady-State Vectors

Question. Determine if the Markov chain with stochastic matrix A has a steady-state vector. If it does, find such a vector.

How to: Steady-State Vectors

Question. Determine if the Markov chain with stochastic matrix A has a steady-state vector. If it does, find such a vector.

Solution. Solve the equation $(A - I)\mathbf{x} = \mathbf{0}$ and find a solution whose entries sum to 1 (this will be possible given a free variable).

How to: Steady-State Vectors

Question. Determine if the Markov chain with stochastic matrix A has a steady-state vector. If it does, find such a vector.

Solution. Solve the equation $(A - I)\mathbf{x} = \mathbf{0}$ and find a solution whose entries sum to 1 (this will be possible given a free variable).

If there is no such solution, the system does not have a steady state.

demo

Existence vs Convergence

If $(A - I)\mathbf{x} = \mathbf{0}$ infinitely many solutions, then it has a stable state.

This **does not** mean:

- » the stable state is unique
- » the system converges to this state

Convergence

Convergence

Definition. For a Markov chain with stochastic matrix A, an initial state $\mathbf{v}_0$ **converges** to the state $\mathbf{v}$ if $\lim_{k\to\infty} A^k\mathbf{v}_0 = \mathbf{v}$.

Convergence

Definition. For a Markov chain with stochastic matrix A, an initial state $\mathbf{v}_0$ **converges** to the state $\mathbf{v}$ if $\lim_{k\to\infty} A^k\mathbf{v}_0 = \mathbf{v}$.

As we repeatedly multiply $\mathbf{v}_0$ by A, we get closer and closer to $\mathbf{v}$ (in the limit).

Convergence

Definition. For a Markov chain with stochastic matrix A, an initial state $\mathbf{v}_0$ **converges** to the state $\mathbf{v}$ if $\lim_{k\to\infty} A^k \mathbf{v}_0 = \mathbf{v}$.

As we repeatedly multiply $\mathbf{v}_0$ by A, we get closer and closer to $\mathbf{v}$ (in the limit).

Example of Non-Convergence

Example of Non-Convergence

Non-Example. I is a stochastic matrix and

$$I\mathbf{v} = \mathbf{v}$$

for any choice of $\mathbf{v}$.

Example of Non-Convergence

Non-Example. I is a stochastic matrix and

$$I\mathbf{v} = \mathbf{v}$$

for any choice of $\mathbf{v}$.

So this system does not have a unique steady state.

Example of Non-Convergence

Non-Example. I is a stochastic matrix and

$$I\mathbf{v} = \mathbf{v}$$

for any choice of $\mathbf{v}$.

So this system does not have a unique steady state.

And no vectors converge to the same stable state.

Regular Stochastic Matrices

Regular Stochastic Matrices

Definition. A stochastic matrix A is **regular** if A^k has all positive entries for *some nonnegative* k.

Regular Stochastic Matrices

Definition. A stochastic matrix A is **regular** if A^k has all positive entries for *some nonnegative* k.

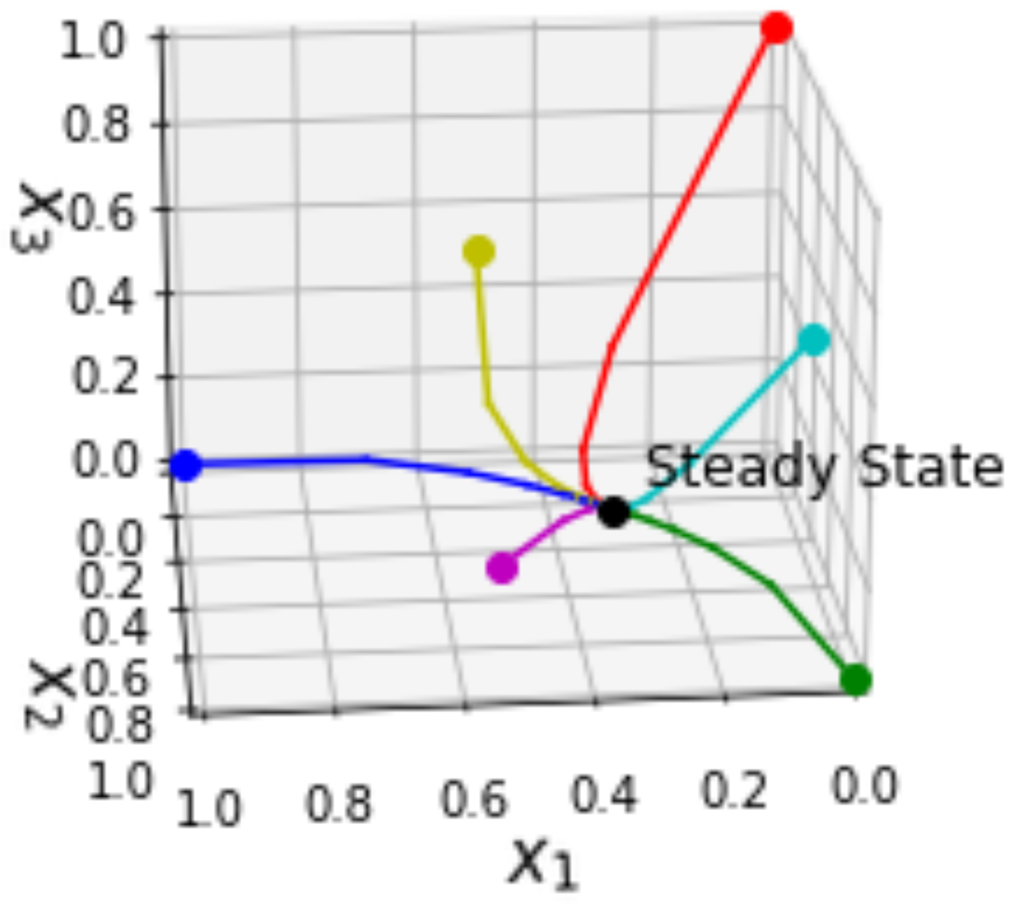
Theorem. A regular stochastic matrix P has a unique steady state, and

every probability vector converges to it

Mixing

This process of converging to a unique steady state is called "mixing."

This theorem says, after some amount of mixing, we'll be close to the stable state, **no matter where we started.**



How to: Regular Stochastic Matrices

Question. Show that A is regular, and then find it's unique steady state.

Solution. Find a power of A which has all positive entries, then solve the equation $(A - I)\mathbf{x} = \mathbf{0}$ as before.

Random Walks

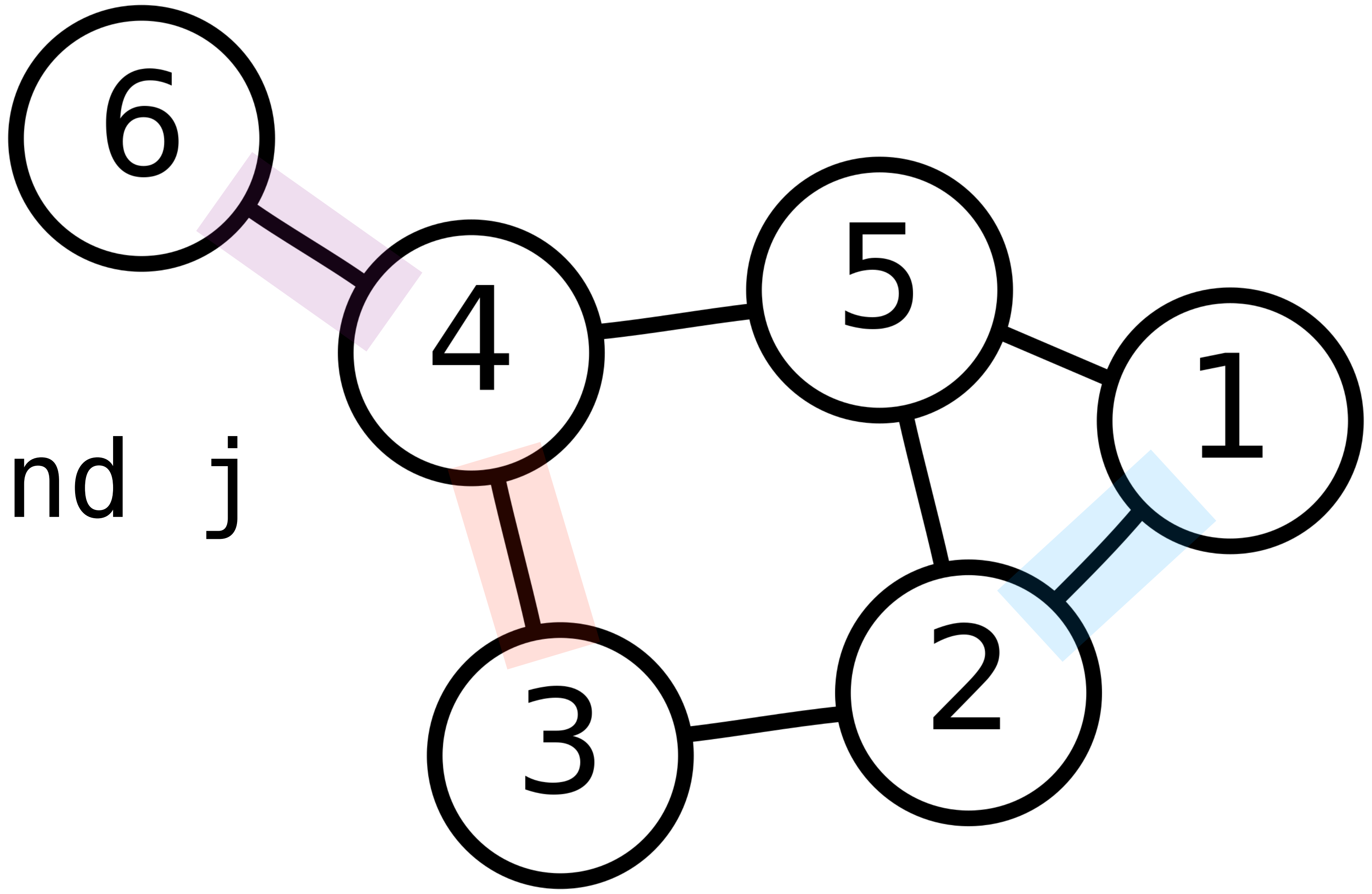
Recall: Adjacency Matrices

$$\begin{array}{c} \\ A_{21} \\ \\ A_{43} \\ \\ A_{64} \end{array} \overset{\quad A_{12} \qquad A_{34} \qquad A_{46}}{\begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}}$$

Let G be an undirected unweighted graph with its nodes labeled by numbers 1 through n.

We can create the **adjacency matrix** A for G as follows.

$$A_{ij} = \begin{cases} 1 & \text{there is an edge between i and j} \\ 0 & \text{otherwise} \end{cases}$$



Random Walk

Random Walk

A **random walk** on an undirected unweighted G starting at v is the following process:

Random Walk

A **random walk** on an undirected unweighted G starting at v is the following process:

» if v is connected to k nodes, roll a k–sided die

Random Walk

A **random walk** on an undirected unweighted G starting at v is the following process:

» if v is connected to k nodes, roll a k–sided die

» go to the kth vertex according to some order

Random Walk

A **random walk** on an undirected unweighted G starting at v is the following process:

» if v is connected to k nodes, roll a k–sided die

» go to the kth vertex according to some order

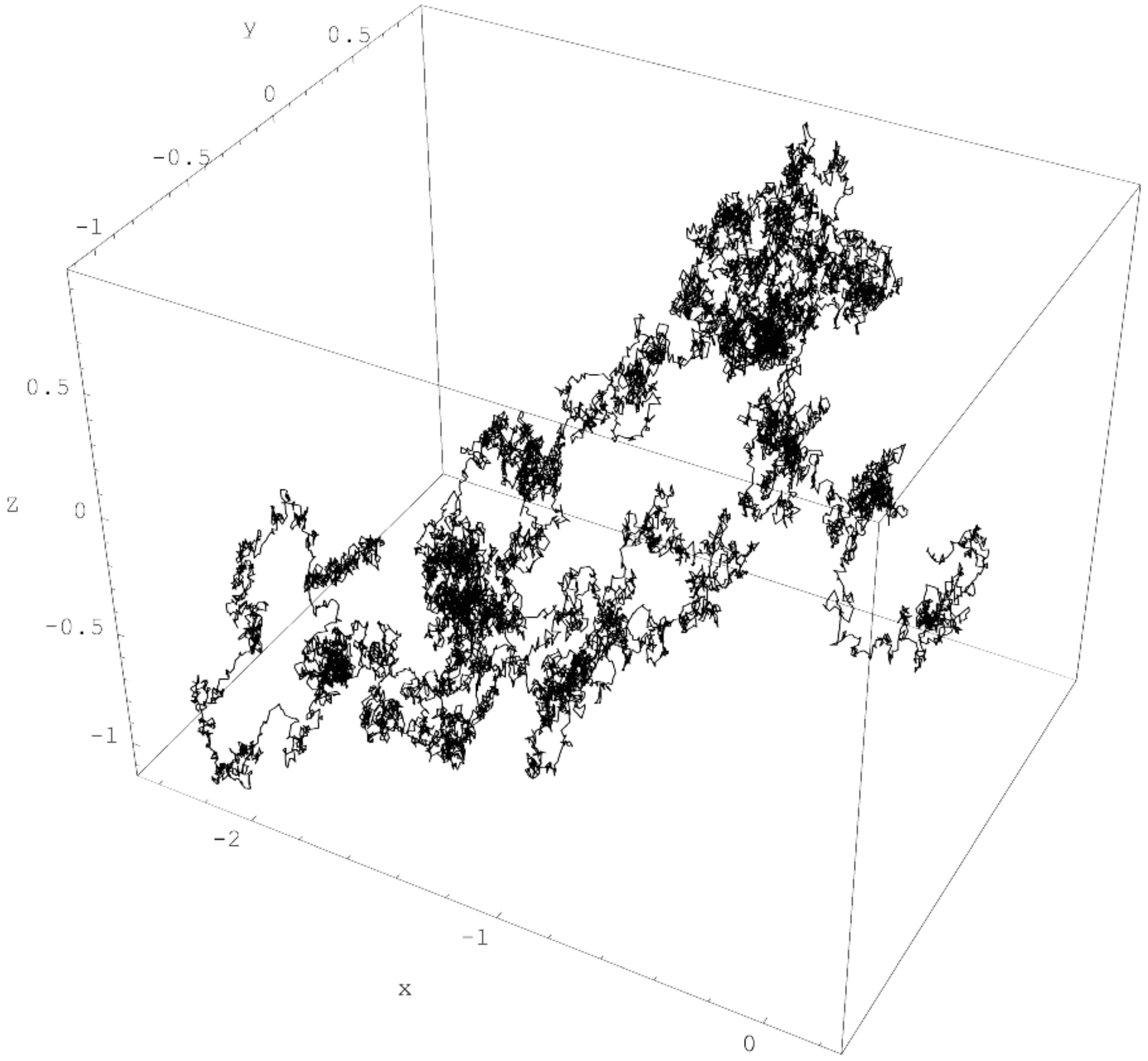
» repeat

Applications of Random Walks

Brownian Motion is a random walk in 3D space.

Random walks are to simulate complex systems in physics and in economics.

They are also used to design algorithms.



https://commons.wikimedia.org/wiki/File:Wiener_process_3d.png

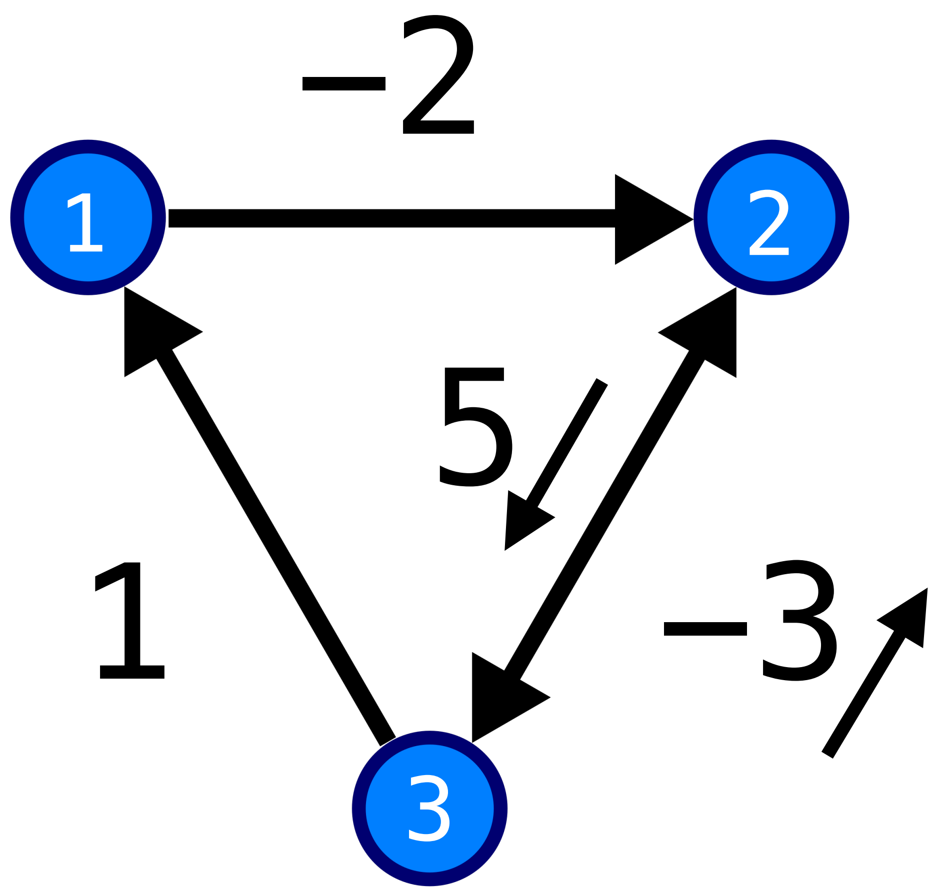
General Adjacency Matrices

We can extend the notion of an adjacency matrix to directed and weighted graphs.

$$A_{ij} = \begin{cases} w_{ji} & \text{there is an edge from j to i} \\ 0 & \text{otherwise} \end{cases}$$

Example.

$$\begin{bmatrix} 0 & 0 & 1 \\ -2 & 0 & -3 \\ 0 & 5 & 0 \end{bmatrix}$$



State Diagrams

Definition. A **state diagram** is a directed weighted graph whose adjacency matrix is stochastic.

Example.

Naming Convention Clash

The nodes of a state diagram are often called states.

The vectors which are dynamically updated according to a linear dynamical system are called state vectors.

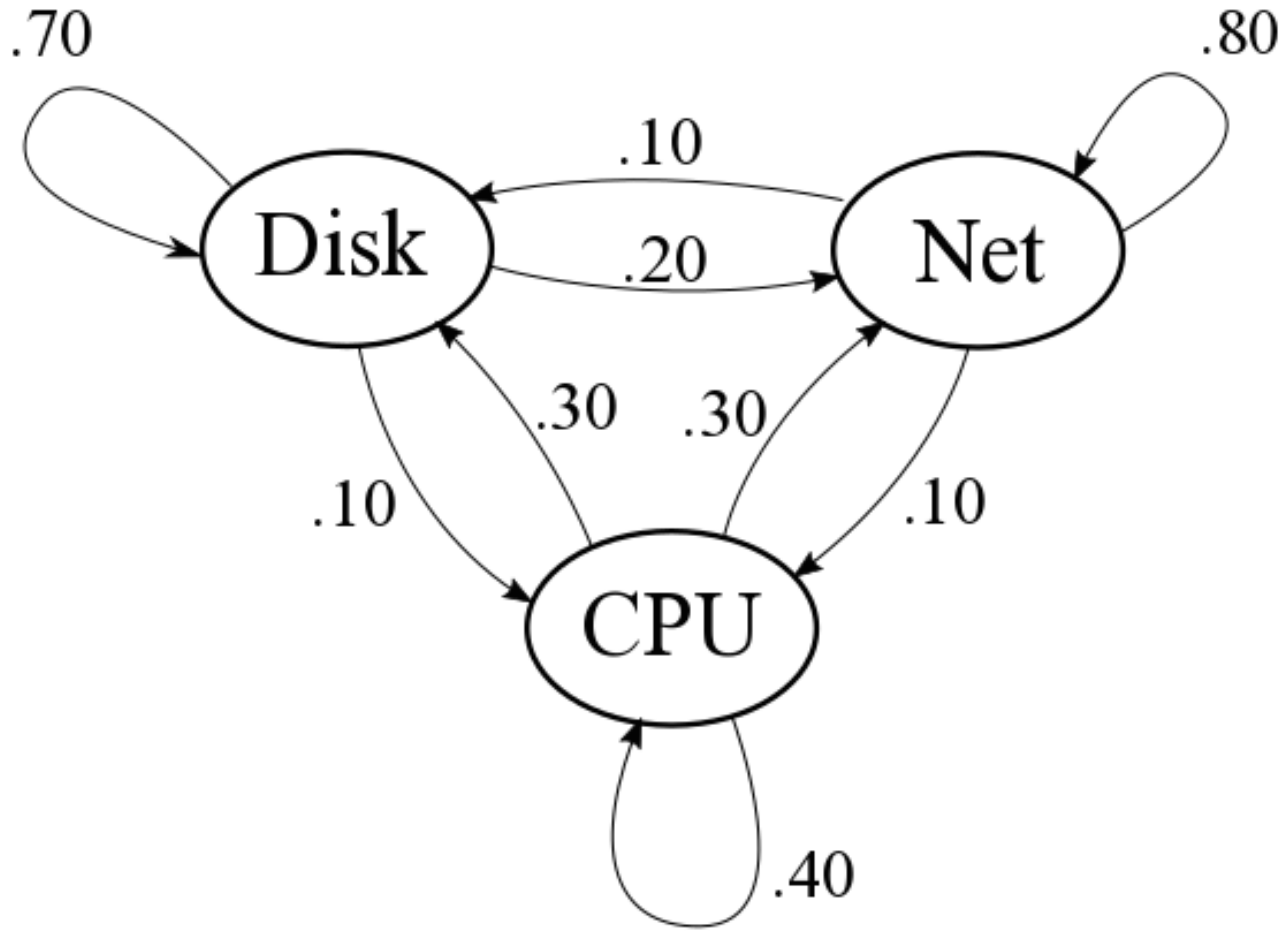
This is an unfortunate naming clash.

Example: Computer System

Imagine a computer system in which tasks request service from disk, network or CPU.

In the long term, which device is busiest?

This is about finding a stable state.



How To: State Diagram

Question. Given a state diagram, find the stable state for the corresponding linear dynamical system.

Solution. Find the adjacency matrix for the state diagram and go from there.

Random Walks as Linear Dynamical Systems

Once we have a stochastic matrix, we can reason about random walks *as linear dynamical systems*.

What are its steady states?

How do we interpret these steady states?

Random Walks on State Diagrams

A **random walk** on a state diagram starting at v is the following process:

» choose a node v is connected to according to the *distribution* given by the edge weights

» go to that node

» repeat

Random Walks on State Diagrams

A **random walk** on a state diagram starting at v is the following process:

» c… o

the…

Stable states of linear dynamical systems are stable states of random walks on state diagrams.

» g…

» repeat

Steady-States of Random Walks

Theorem (Advanced). Let A be the stochastic matrix for the graph G. The probability that a random walk starting at i of length k ends on node j is

$$(A^k \mathbf{e}_i)_j$$

the jth entry of the vector $A^k \mathbf{e}_i$

A transforms a distribution for length k walks to length $k+1$ walks.

Steady States of Random Walks

If a random walk goes on for a sufficiently long time, then the probability that we end up in a particular place becomes fixed.

If you wander for a sufficiently long time, it doesn't matter where you started.

Summary

Markov chains allow us to reason about dynamical systems that are dictated by some amount of randomness.

Stable states represent global equilibrium.

We can think of Markov chains as random walks on state diagrams.