[12] The Eigenvector

Two interest-bearing accounts

Suppose Account 1 yields 5% interest and Account 2 yields 3% interest.

Represent balances in the two accounts by a 2-vector $\mathbf{x}^{(t)} = \begin{bmatrix} \text{amount in Account 1} \\ \text{amount in Account 2} \end{bmatrix}$.

$$\mathbf{x}^{(t+1)} = \begin{bmatrix} 1.05 & 0 \\ 0 & 1.03 \end{bmatrix} \mathbf{x}^{(t)}$$

Let A denote the matrix. It is diagonal.

To find out how, say, $\mathbf{x}^{(100)}$ compares to $\mathbf{x}^{(0)}$, we can use Equation repeatedly:

$$\begin{aligned} \mathbf{x}^{(100)} &= A\mathbf{x}^{(99)} \\ &= A(A\mathbf{x}^{(98)}) \\ &\vdots \\ &= \underbrace{A \cdot A \cdot \cdots \cdot A}_{100 \text{ times}} \mathbf{x}^{(0)} \\ &= A^{100}\mathbf{x}^{(0)} \end{aligned}$$

Two interest-bearing accounts

$$
\begin{aligned}
\mathbf{x}^{(100)} &= A\mathbf{x}^{(99)} \\
&= A(A\mathbf{x}^{(98)}) \\
&\vdots \\
&= \underbrace{A \cdot A \cdot \cdots \cdot A}_{100 \text{ times}} \mathbf{x}^{(0)} \\
&= A^{100}\mathbf{x}^{(0)}
\end{aligned}
$$

Since A is a diagonal matrix, easy to compute powers of A:

Two interest-bearing accounts

$$\begin{aligned}\mathbf{x}^{(100)} &= A\mathbf{x}^{(99)} \\ &= A(A\mathbf{x}^{(98)}) \\ &\vdots \\ &= \underbrace{A \cdot A \cdot \cdots \cdot A}_{100 \text{ times}} \mathbf{x}^{(0)} \\ &= A^{100}\mathbf{x}^{(0)}\end{aligned}$$

Since A is a diagonal matrix, easy to compute powers of A:

$$\begin{bmatrix} 1.05 & 0 \\ 0 & 1.03 \end{bmatrix} \begin{bmatrix} 1.05 & 0 \\ 0 & 1.03 \end{bmatrix} = \begin{bmatrix} 1.05^2 & 0 \\ 0 & 1.03^2 \end{bmatrix}$$

Two interest-bearing accounts

$$
\begin{aligned}
\mathbf{x}^{(100)} &= A\mathbf{x}^{(99)} \\
&= A(A\mathbf{x}^{(98)}) \\
&\vdots \\
&= \underbrace{A \cdot A \cdot \cdots \cdot A}_{100 \text{ times}} \mathbf{x}^{(0)} \\
&= A^{100}\mathbf{x}^{(0)}
\end{aligned}
$$

Since A is a diagonal matrix, easy to compute powers of A:

$$
\underbrace{\begin{bmatrix} 1.05 & 0 \\ 0 & 1.03 \end{bmatrix} \cdots \begin{bmatrix} 1.05 & 0 \\ 0 & 1.03 \end{bmatrix}}_{100 \text{ times}} = \begin{bmatrix} 1.05^{100} & 0 \\ 0 & 1.03^{100} \end{bmatrix} \approx \begin{bmatrix} 131.5 & 0 \\ 0 & 19.2 \end{bmatrix}
$$

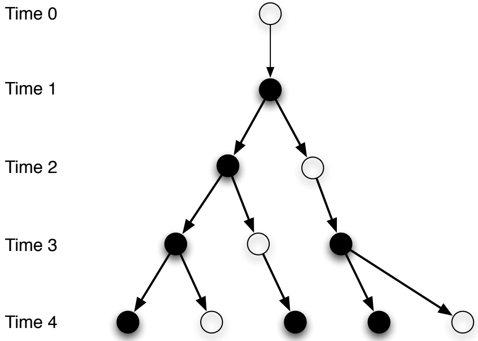
The takeaway:

$$
\begin{bmatrix} \text{Account 1 balance after } t \text{ years} \\ \text{Account 2 balance after } t \text{ years} \end{bmatrix} = \begin{bmatrix} 1.05^t \cdot (\text{initial Account 1 balance}) \\ 1.03^t \cdot (\text{initial Account 2 balance}) \end{bmatrix}
$$

Rabbit reproduction

To avoid getting into trouble, I'll pretend sex doesn't exist.

- Each month, each adult rabbit gives birth to one baby.
- A rabbit takes one month to become an adult.
- Rabbits never die.



$$\begin{bmatrix} \text{adults at time } t+1 \\ \text{juveniles at time } t+1 \end{bmatrix} = \underbrace{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}}_{A} \begin{bmatrix} \text{adults at time } t \\ \text{juveniles at time } t \end{bmatrix}$$

Use $\mathbf{x}^{(t)} = \begin{bmatrix} \text{number of adults after } t \text{ months} \\ \text{number of juveniles after } t \text{ months} \end{bmatrix}$

Then $\mathbf{x}^{(t+1)} = A\mathbf{x}^{(t)}$ where $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$.

$[1, 0],\ [1, 1],\ [2, 1],\ [3, 2],\ [5, 3],\ [8, 3], \ldots$

Analyzing rabbit reproduction

$\mathbf{x}^{(t+1)} = A\mathbf{x}^{(t)}$ where $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$.

As in bank-account example, $\mathbf{x}^{(t)} = A^t \mathbf{x}^{(0)}$.

How can this help us calculate how the entries of $x^{(t)}$ grow as a function of t? In the bank-account example, we were able to understand the behavior because A was a diagonal matrix. This time, A is not diagonal. However, there is a workaround:

Let $S = \begin{bmatrix} \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \\ 1 & 1 \end{bmatrix}$. Then $S^{-1}AS$ is the diagonal matrix $\begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix}$.

$$\begin{aligned} A^t &= \underbrace{A\,A\,\cdots A}_{t \text{ times}} \\ &= (S\Lambda S^{-1})(S\Lambda S^{-1})\cdots(S\Lambda S^{-1}) \\ &= S\Lambda^t S^{-1} \end{aligned}$$

Λ is a diagonal matrix $\Rightarrow$ easy to compute Λ^t.

If $\Lambda = \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix}$ then $\Lambda^t = \begin{bmatrix} \lambda_1^t & \\ & \lambda_2^t \end{bmatrix}$. Here $\Lambda = \begin{bmatrix} \frac{1+\sqrt{5}}{2} & \\ & \frac{1-\sqrt{5}}{2} \end{bmatrix}$.

Interpretation using change of basis

Interpretation: To make the analysis easier, we will use a change of basis

Basis consists of the two columns of the matrix S, $\mathbf{v}_1 = \begin{bmatrix} \frac{1+\sqrt{5}}{2} \\ 1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} \frac{1-\sqrt{5}}{2} \\ 1 \end{bmatrix}$

Let $\mathbf{u}^{(t)}$ = coordinate representation of $\mathbf{x}^{(t)}$ in terms of $\mathbf{v}_1$ and $\mathbf{v}_2$.

- (`rep2vec`) To go from repres. $\mathbf{u}^{(t)}$ to vector $\mathbf{x}^{(t)}$ itself, we multiply $\mathbf{u}^{(t)}$ by S.
- (Move forward one month) To go from $\mathbf{x}^{(t)}$ to $\mathbf{x}^{(t+1)}$, we multiply $\mathbf{x}^{(t)}$ by A.
- (`vec2rep`) To go to coord. repres., we multiply by S^{-1}.

Multiplying by the matrix $S^{-1}AS$ carries out the three steps above.

But $S^{-1}AS = \Lambda = \begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix}$ so $\mathbf{u}^{(t+1)} = \begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix} \mathbf{u}^{(t)}$

so

$$\mathbf{u}^{(t)} = \begin{bmatrix} (\frac{1+\sqrt{5}}{2})^t & 0 \\ 0 & (\frac{1-\sqrt{5}}{2})^t \end{bmatrix} \mathbf{u}^{(0)}$$

Eigenvalues and eigenvectors

For this topic, consider only matrices A such that row-label set = col-label set (*endomorphic* matrices).

Definition: If λ is a scalar and $\mathbf{v}$ is a nonzero vector such that $A\mathbf{v} = \lambda\mathbf{v}$, we say that λ is an *eigenvalue* of A, and $\mathbf{v}$ is a corresponding *eigenvector*.

Any nonzero vector in the eigenspace is considered an eigenvector. However, it is often convenient to require that the eigenvector have norm one.

Example: $\begin{bmatrix} 1.05 & 0 \\ 0 & 1.03 \end{bmatrix}$ has eigenvalues 1.05 and 1.03, and corresponding eigenvectors $[1, 0]$ and $[0, 1]$.

Example: $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ has eigenvalues $\lambda_1 = \frac{1+\sqrt{5}}{2}$ and $\lambda_2 = \frac{1-\sqrt{5}}{2}$, and corresponding eigenvectors $[\frac{1+\sqrt{5}}{2}, 1]$ and $[\frac{1-\sqrt{5}}{2}, 1]$.

Example: What does it mean when A has 0 as an eigenvalue? There is a nonzero vector $\mathbf{v}$ such that $A\mathbf{v} = 0\mathbf{v}$. That is, A's null space is nontrivial.

Last example suggests a way to find an eigenvector corresp. to eigenvalue 0: find nonzero vector in the null space. What about other eigenvalues?

Eigenvector corresponding to an eigenvalue

Suppose λ is an eigenvalue of A, with corresponding eigenvector $\mathbf{v}$. Then $A\mathbf{v} = \lambda\,\mathbf{v}$. That is, $A\mathbf{v} - \lambda\,\mathbf{v}$ is the zero vector.The expression $A\mathbf{v} - \lambda\,\mathbf{v}$ can be written as $(A - \lambda\,\mathbb{1})\mathbf{v}$,so $(A - \lambda\,\mathbb{1})\mathbf{v}$ is the zero vector. That means that $\mathbf{v}$ is a nonzero vector in the null space of $A - \lambda\,\mathbb{1}$. That means that $A - \lambda\,\mathbb{1}$ is not invertible.
Conversely, suppose $A - \lambda\,\mathbb{1}$ is not invertible. It is square, so it must have a nontrivial null space. Let $\mathbf{v}$ be a nonzero vector in the null space. Then $(A - \lambda\mathbb{1})\mathbf{v} = \mathbf{0}$, so $A\mathbf{v} = \lambda\mathbf{v}$.
We have proved the following:
Lemma: Let A be a square matrix.

- The number λ is an eigenvalue of A if and only if $A - \lambda\,\mathbb{1}$ is not invertible.
- If λ is in fact an eigenvalue of A then the corresponding eigenspace is the null space of $A - \lambda\,\mathbb{1}$.

Corollary

If λ is an eigenvalue of A then it is an eigenvalue of A^T.

Similarity

Definition: Two matrices A and B are *similar* if there is an invertible matrix S such that $S^{-1}AS = B$.
Proposition: Similar matrices have the same eigenvalues.
Proof: Suppose λ is an eigenvalue of A and $\mathbf{v}$ is a corresponding eigenvector. By definition, $A\mathbf{v} = \lambda\,\mathbf{v}$. Suppose $S^{-1}AS = B$, and let $\mathbf{w} = S^{-1}\mathbf{v}$. Then

$$\begin{aligned} B\mathbf{w} &= S^{-1}AS\mathbf{w} \\ &= S^{-1}ASS^{-1}\mathbf{v} \\ &= S^{-1}A\mathbf{v} \\ &= S^{-1}\lambda\,\mathbf{v} \\ &= \lambda\,S^{-1}\mathbf{v} \\ &= \lambda\,\mathbf{w} \end{aligned}$$

which shows that λ is an eigenvalue of B.

Example of similarity

Example: We will see later that the eigenvalues of the matrix $A = \begin{bmatrix} 6 & 3 & -9 \\ 0 & 9 & 15 \\ 0 & 0 & 15 \end{bmatrix}$ are its diagonal elements (6, 9, and 15) because U is upper triangular. The matrix $B = \begin{bmatrix} 92 & -32 & -15 \\ -64 & 34 & 39 \\ 176 & -68 & -99 \end{bmatrix}$ has the property that $B = S^{-1}AS$ where $S = \begin{bmatrix} -2 & 1 & 4 \\ 1 & -2 & 1 \\ -4 & 3 & 5 \end{bmatrix}$. Therefore the eigenvalues of B are also 6, 9, and 15.

Diagonalizability

Definition: If A is similar to a diagonal matrix, i.e. if there is an invertible matrix S such that $S^{-1}AS = \Lambda$ where Λ is a diagonal matrix, we say A is *diagonalizable*.

Equation $S^{-1}AS = \Lambda$ is equivalent to equation $A = S\Lambda S^{-1}$, which is the form used in the analysis of rabbit population. How is diagonalizability related to eigenvalues?

- Eigenvalues of a diagonal matrix $\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$ are its diagonal entries.
- If matrix A is similar to Λ then the eigenvalues of A are the eigenvalues of Λ
- Equation $S^{-1}AS = \Lambda$ is equivalent to $AS = S\Lambda$. Write S in terms of columns:

$$\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} \left[\mathbf{v}_1 \middle| \cdots \middle| \mathbf{v}_n \right] = \left[\mathbf{v}_1 \middle| \cdots \middle| \mathbf{v}_n \right] \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

- The argument goes both ways: if $n \times n$ matrix A has n linearly independent eigenvectors then A is diagonalizable.

Diagonalizability

Definition: If A is similar to a diagonal matrix, i.e. if there is an invertible matrix S such that $S^{-1}AS = \Lambda$ where Λ is a diagonal matrix, we say A is *diagonalizable*.

Equation $S^{-1}AS = \Lambda$ is equivalent to equation $A = S\Lambda S^{-1}$, which is the form used in the analysis of rabbit population. How is diagonalizability related to eigenvalues?

- Eigenvalues of a diagonal matrix $\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$ are its diagonal entries.
- If matrix A is similar to Λ then the eigenvalues of A are the eigenvalues of Λ
- Equation $S^{-1}AS = \Lambda$ is equivalent to $AS = S\Lambda$. Write S in terms of columns:

$$\left[\begin{array}{c|c|c} A\mathbf{v}_1 & \cdots & A\mathbf{v}_n \end{array}\right] = \left[\begin{array}{c|c|c} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{array}\right] \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

- The argument goes both ways: if $n \times n$ matrix A has n linearly independent eigenvectors then A is diagonalizable.

Diagonalizability

Definition: If A is similar to a diagonal matrix, i.e. if there is an invertible matrix S such that $S^{-1}AS = \Lambda$ where Λ is a diagonal matrix, we say A is *diagonalizable*.

Equation $S^{-1}AS = \Lambda$ is equivalent to equation $A = S\Lambda S^{-1}$, which is the form used in the analysis of rabbit population. How is diagonalizability related to eigenvalues?

- Eigenvalues of a diagonal matrix $\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$ are its diagonal entries.
- If matrix A is similar to Λ then the eigenvalues of A are the eigenvalues of Λ
- Equation $S^{-1}AS = \Lambda$ is equivalent to $AS = S\Lambda$. Write S in terms of columns:

$$\left[\begin{array}{c|c|c} A\mathbf{v}_1 & \cdots & A\mathbf{v}_n \end{array}\right] = \left[\begin{array}{c|c|c} \lambda_1\mathbf{v}_1 & \cdots & \lambda_n\mathbf{v}_n \end{array}\right]$$

  Columns $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of S are eigenvectors. Because S is invertible, the eigenvectors are linearly independent.
- The argument goes both ways: if $n \times n$ matrix A has n linearly independent eigenvectors then A is diagonalizable.

Diagonalizability Theorem

Diagonalizability Theorem: An $n \times n$ matrix A is diagonalizable iff it has n linearly independent eigenvectors.

Example: Consider the matrix $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Its null space is trivial so zero is not an eigenvalue. For any 2-vector $\begin{bmatrix} x \\ y \end{bmatrix}$, we have $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x+y \\ y \end{bmatrix}$. Suppose λ is an eigenvector. Then for some vector $[x, y]$,

$$\lambda[x, y] = [x + y, y]$$

Therefore $\lambda\, y = y$. Therefore $y = 0$. Therefore every eigenvector is in Span $\{[1, 0]\}$. Thus the matrix does not have two linearly independent eigenvectors, so it is not diagonalizable.

Interpretation using change of basis, revisited

Idea used for rabbit problem can be used more generally.

Suppose A is a diagonalizable matrix. Then $A = S\Lambda S^{-1}$ for a diagonal matrix Λ and invertible matrix S.
Suppose $\mathbf{x}^{(0)}$ is a vector. The equation $\mathbf{x}^{(t+1)} = A\,\mathbf{x}^{(t)}$ then defines $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \ldots$.
Then

$$
\begin{aligned}
\mathbf{x}^{(t)} &= \underbrace{A\;A\;\cdots\;A}_{t\text{ times}}\;\mathbf{x}^{(0)} \\
&= (S\Lambda S^{-1})(S\Lambda S^{-1})\cdots(S\Lambda S^{-1})\;\mathbf{x}^{(0)} \\
&= S\Lambda^t S^{-1}\;\mathbf{x}^{(0)}
\end{aligned}
$$

Interpretation: Let $\mathbf{u}^{(t)}$ be the coordinate representation of $\mathbf{x}^{(t)}$ in terms of the columns of S. Then we have the equation $\mathbf{u}^{(t+1)} = \Lambda\,\mathbf{u}^{(t)}$. Therefore

$$
\begin{aligned}
\mathbf{u}^{(t)} &= \underbrace{\Lambda\;\Lambda\;\cdots\;\Lambda}_{t\text{ times}}\;\mathbf{u}^{(0)} \\
&= \Lambda^t\;\mathbf{u}^{(0)}
\end{aligned}
$$

If $\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$ then

$$\Lambda^t = \begin{bmatrix} \lambda_1^t & & \\ & \ddots & \\ & & \lambda_n^t \end{bmatrix}$$

Interpretation using change of basis, re-revisited

Suppose $n \times n$ matrix A is diagonalizable, so it has linearly independent eigenvectors. Suppose eigenvalues are $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ and corresponding eigenvectors are $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$. The eigenvectors form a basis for $\mathbb{R}^n$, so any vector $\mathbf{x}$ can be written as a linear combination:

$$\mathbf{x} = \alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n$$

Left-multiply by A on both sides of the equation:

$$\begin{aligned} A\mathbf{x} &= A(\alpha_1 \mathbf{v}_1) + A(\alpha_2 \mathbf{v}_2) + \cdots + A(\alpha_n \mathbf{v}_n) \\ &= \alpha_1 A\mathbf{v}_1 + \alpha_2 A\mathbf{v}_2 + \cdots + \alpha_n A\mathbf{v}_n \\ &= \alpha_1 \lambda_1 \mathbf{v}_1 + \alpha_2 \lambda_2 \mathbf{v}_2 + \cdots + \alpha_n \lambda_n \mathbf{v}_n \end{aligned}$$

Applying the same reasoning to $A(A\mathbf{x})$, we get

$$A^2\mathbf{x} = \alpha_1 \lambda_1^2 \mathbf{v}_1 + \alpha_2 \lambda_2^2 \mathbf{v}_2 + \cdots + \alpha_n \lambda_n^2 \mathbf{v}_n$$

More generally, for any nonnegative integer t,

$$A^t\mathbf{x} = \alpha_1 \lambda_1^t \mathbf{v}_1 + \alpha_2 \lambda_2^t \mathbf{v}_2 + \cdots + \alpha_n \lambda_n^t \mathbf{v}_n$$

If $|\lambda_1|$ is bigger than the other eigenvalues then eventually λ_1^t will be *much* bigger than $\lambda_2^t, \ldots, \lambda_n^t$, so first term will dominate. For a large enough value of t, $A^t\mathbf{x}$ will be approximately $\alpha_1 \lambda_1^t \mathbf{v}_1$.

The Internet Worm of 1988

Robert T. Morris, Jr. He wrote a program that exploited some known security holes in unix to spread running copies of itself through the Internet. Whenever a worm (running copy) on one computer managed to break into another computer, it would spawn a worm on the other computer.

It was intended to remain undetected but eventually it took down most of the computers on the Internet. The reason is that each computer was running many independent instances of the program.
He took steps to prevent this.

- Each worm would check whether there was another worm running on same computer.
- If so, worm would set a flag indicating it was supposed to die.
- However, with probability $1/7$ the worm would designate itself immortal.

Does the number of worms grow? If so, how fast?

Modeling the Worm

Suppose Internet consists of just three computers in a triangular network.

In each iteration, each worm has probability $1/10$ of spawning a child worm on each neighboring computer. if it is a mortal worm, with probability $1/7$ it becomes immortal, and otherwise it dies.

Worm population represented by a vector $\mathbf{x} = [x_1, y_1, x_2, y_2, x_3, y_3]$: for $i = 1, 2, 3$, x_i is the expected number of mortal worms at computer i, and y_i is the expected number of immortal worms at computer i.

For $t = 0, 1, 2, \ldots,$, let $\mathbf{x}^{(t)} = (x_1^{(t)}, y_1^{(t)}, x_2^{(t)}, y_2^{(t)}, x_3^{(t)}, y_3^{(t)})$.
Any mortal worm at computer 1 is a child of a worm at computer 2 or computer 3. Therefore the expected number of mortal worms at computer 1 after $t+1$ iterations is $1/10$ times the expected number of worms at computers 2 and 3 after t iterations. Therefore

$$x_1^{(t+1)} = \frac{1}{10}x_2^{(t)} + \frac{1}{10}y_2^{(t)} + \frac{1}{10}x_3^{(t)} + \frac{1}{10}y_3^{(t)}$$

With probability $1/7$, a mortal worm at computer 1 becomes immortal. The previously immortal worms stay immortal.Therefore

$$y_1^{(t+1)} = \frac{1}{7}x_1^{(t)} + y_1^{(t)}$$

Modeling the Worm

Therefore

$$x_1^{(t+1)} = \frac{1}{10}x_2^{(t)} + \frac{1}{10}y_2^{(t)} + \frac{1}{10}x_3^{(t)} + \frac{1}{10}y_3^{(t)}$$

With probability 1/7, a mortal worm at computer 1 becomes immortal. The previously immortal worms stay immortal.Therefore

$$y_1^{(t+1)} = \frac{1}{7}x_1^{(t)} + y_1^{(t)}$$

The equations for $x_2^{(t+1)}$ and $y_2^{(t+1)}$ and $x_3^{(t+1)}$ and $y_3^{(t+1)}$ are similar. We therefore get

$$\mathbf{x}^{(t+1)} = A\mathbf{x}^{(t)}$$

where A is the matrix

$$A = \begin{bmatrix} 0 & 0 & 1/10 & 1/10 & 1/10 & 1/10 \\ 1/7 & 1 & 0 & 0 & 0 & 0 \\ 1/10 & 1/10 & 0 & 0 & 1/10 & 1/10 \\ 0 & & 1/7 & 1 & 0 & 0 \\ 1/10 & 1/10 & 1/10 & 1/10 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/7 & 1 \end{bmatrix}$$

Analyzing the worm, continued

$$\mathbf{x}^{(t+1)} = A\mathbf{x}^{(t)}$$

where A is the matrix

$$A = \begin{bmatrix} 0 & 0 & 1/10 & 1/10 & 1/10 & 1/10 \\ 1/7 & 1 & 0 & 0 & 0 & 0 \\ 1/10 & 1/10 & 0 & 0 & 1/10 & 1/10 \\ 0 & & 1/7 & 1 & 0 & 0 \\ 1/10 & 1/10 & 1/10 & 1/10 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/7 & 1 \end{bmatrix}$$

This matrix has linearly independent eigenvectors, and its largest eigenvalue is about 1.034. Because this is larger than 1, we can infer that the number of worms will grow exponentially with the number of iterations.

t	1.034^t
100	29
200	841
500	$20,000,000$
600	$600,000,000$

Modeling population movement

Dance-club dynamics: At the beginning of each song,

- 56% of the people standing on the side go onto the dance floor, and
- 12% of the people on the dance floor leave it.

Suppose that there are a hundred people in the club. Assume nobody enters the club and nobody leaves. What happens to the number of people in each of the two locations?

Represent state of system by

$$\mathbf{x}^{(t)} = \begin{bmatrix} x_1^{(t)} \\ x_2^{(t)} \end{bmatrix} = \begin{bmatrix} \text{number of people standing on side after } t \text{ songs} \\ \text{number of people on dance floor after } t \text{ songs} \end{bmatrix}$$

$$\begin{bmatrix} x_1^{(t+1)} \\ x_2^{(t+1)} \end{bmatrix} = \begin{bmatrix} .44 & .12 \\ .56 & .88 \end{bmatrix} \begin{bmatrix} x_1^{(t)} \\ x_2^{(t)} \end{bmatrix}$$

Diagonalize: $S^{-1}AS = \Lambda$ where

$$A = \begin{bmatrix} .44 & .12 \\ .56 & .88 \end{bmatrix}, S = \begin{bmatrix} 0.209529 & -1 \\ 0.977802 & 1 \end{bmatrix}, \Lambda = \begin{bmatrix} 1 & 0 \\ 0 & 0.32 \end{bmatrix}$$

Analyzing dance-floor dynamics

$$
\begin{aligned}
\begin{bmatrix} x_1^{(t)} \\ x_2^{(t)} \end{bmatrix} &= \left(S\Lambda S^{-1}\right)^t \begin{bmatrix} x_1^{(0)} \\ x_2^{(0)} \end{bmatrix} \\
&= S\Lambda^t S^{-1} \begin{bmatrix} x_1^{(0)} \\ x_2^{(0)} \end{bmatrix} \\
&= \begin{bmatrix} .21 & -1 \\ .98 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & .32 \end{bmatrix}^t \begin{bmatrix} .84 & .84 \\ -.82 & .18 \end{bmatrix} \begin{bmatrix} x_1^{(0)} \\ x_2^{(0)} \end{bmatrix} \\
&= \begin{bmatrix} .21 & -1 \\ .98 & 1 \end{bmatrix} \begin{bmatrix} 1^t & 0 \\ 0 & .32^t \end{bmatrix} \begin{bmatrix} .84 & .84 \\ -.82 & .18 \end{bmatrix} \begin{bmatrix} x_1^{(0)} \\ x_2^{(0)} \end{bmatrix} \\
&= 1^t(.84x_1^{(0)} + .84x_2^{(0)}) \begin{bmatrix} .21 \\ .98 \end{bmatrix} + (0.32)^t(-.82x_1^{(0)} + .18x_2^{(0)}) \begin{bmatrix} -1 \\ 1 \end{bmatrix} \\
&= 1^t \underbrace{\left(x_1^{(0)} + x_2^{(0)}\right)}_{\text{total population}} \begin{bmatrix} .18 \\ .82 \end{bmatrix} + (0.32)^t \left(-.82x_1^{(0)} + .18x_2^{(0)}\right) \begin{bmatrix} -1 \\ 1 \end{bmatrix}
\end{aligned}
$$

Analyzing dance-floor dynamics, continued

$$\begin{bmatrix} x_1^{(t)} \\ x_2^{(t)} \end{bmatrix} = \underbrace{\left(x_1^{(0)} + x_2^{(0)}\right)}_{\text{total population}} \begin{bmatrix} .18 \\ .82 \end{bmatrix} + (0.32)^t \left(-.82x_1^{(0)} + .18x_2^{(0)}\right) \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

The numbers of people in the two locations after t songs depend on the *initial* numbers of people in the two locations.

However, the dependency grows weaker as the number of songs increases: $(0.32)^t$ gets smaller and smaller, so the second term in the sum matters less and less.

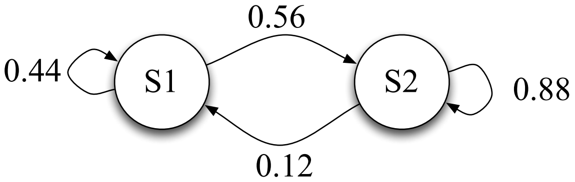
After ten songs, $(0.32)^t$ is about 0.00001.

The first term in the sum is $\begin{bmatrix} .18 \\ .82 \end{bmatrix}$ times the total number of people. This shows that, as the number of songs increases, the proportion of people on the dance floor gets closer and closer to 82%.

Modeling Randy

Without changing math, we switch interpretations. Instead of modeling whole dance-club population, we model one person, Randy.

Randy's behavior captured in transition diagram:



State S1 represents Randy being on the side. State S2 represents Randy being on the dance floor.

After each song, Randy follows one of the arrows from current state. Which arrow? Chosen randomly according to probabilities on the arrows (*transition probabilities*).

For each state, labels on arrows from that state must sum to 1.

Where is Randy?

Even if we know where Randy starts at time 0, we can't predict with certainty where he will be at time t. However, for each time t, we can calculate the probability distribution for his location.

Since there are two possible locations (off floor, on floor), the probability distribution is given by a 2-vector $\mathbf{x}^{(t)} = \begin{bmatrix} x_1^{(t)} \\ x_2^{(t)} \end{bmatrix}$ where $x_1^{(t)} + x_2^{(t)} = 1$.

Probability distribution for Randy's location at time $t+1$ is related to probability distribution for Randy's location at time t:

$$\begin{bmatrix} x_1^{(t+1)} \\ x_2^{(t+1)} \end{bmatrix} = \begin{bmatrix} .44 & .12 \\ .56 & .88 \end{bmatrix} \begin{bmatrix} x_1^{(t)} \\ x_2^{(t}) \end{bmatrix}$$

Using earlier analysis,

$$\begin{aligned} \begin{bmatrix} x_1^{(t)} \\ x_2^{(t)} \end{bmatrix} &= \left(x_1^{(0)} + x_2^{(0)}\right) \begin{bmatrix} .18 \\ .82 \end{bmatrix} + (0.32)^t \left(-.82x_1^{(0)} + .18x_2^{(0)}\right) \begin{bmatrix} -1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} .18 \\ .82 \end{bmatrix} + (0.32)^t \left(-.82x_1^{(0)} + .18x_2^{(0)}\right) \begin{bmatrix} -1 \\ 1 \end{bmatrix} \end{aligned}$$

Where is Randy?

$$\begin{bmatrix} x_1^{(t)} \\ x_2^{(t)} \end{bmatrix} = \begin{bmatrix} .18 \\ .82 \end{bmatrix} + (0.32)^t \left(-.82x_1^{(0)} + .18x_2^{(0)} \right) \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

If we know Randy starts off the dance floor at time 0 then $x_1^{(0)} = 1$ and $x_2^{(0)} = 0$.
If we know Randy starts on the dance floor at time 0 then $x_1^{(0)} = 0$ and $x_2^{(0)} = 1$.

In either case, we can plug in to equation to get exact probability distribution for time t.

But after a few songs, the starting location doesn't matter much—the probability distribution gets very close to $\begin{bmatrix} .18 \\ .82 \end{bmatrix}$ in either case.

This is called Randy's *stationary distribution*.

It doesn't mean Randy stays in one place—we expect him to move back and forth all the time. It means that the probability distribution for his location after t steps depends less and less on t.

From Randy to spatial locality in CPU memory fetches

We again switch interpretations without changing the math.

CPU uses caches and prefetching to improve performance.
To help computer architects, it is useful to model CPU access patterns.

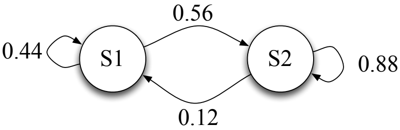
After accessing location x, CPU usually accesses location $x + 1$.
Therefore simple model is:

$$\text{Probability}[\text{address requested at time } t + 1 \text{ is } 1 + \text{address requested at time } t] = .6$$

However, a slightly more sophisticated model predicts much more accurately.

Observation: Once consecutive addresses have been requested in timesteps t and $t + 1$, it is very likely that the address requested in timestep $t + 2$ is also consecutive.

Use same model as used for Randy.

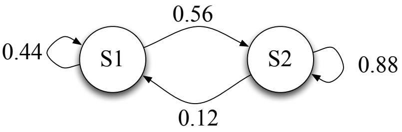


State S1 = CPU is requesting nonconsecutive addresses.
State S2 = CPU is requesting consecutive addresses.

From Randy to spatial locality in CPU memory fetches

Observation: Once consecutive addresses have been requested in timesteps t and $t+1$, it is very likely that the address requested in timestep $t+2$ is also consecutive.

Use same model as used for Randy.



State S1 = CPU is requesting nonconsecutive addresses.
State S2 = CPU is requesting consecutive addresses.

Once CPU starts requesting consecutive addresses, it tends to stay in that mode for a while. This tendency is captured by the model.

As with Randy, after a while the probability distribution is $[0.18, 0.82]$. Being in the first state means the CPU is issuing the first of a run of consecutive addresses (possibly of length 1) Since the system is in the first state roughly 18% of the time, the average length of such a run is $1/0.18$.

Various such calculations can be useful in designing architectures and improving performance.

Markov chains

An *n-state Markov chain* is a system such that

- At each time, the system is in one of n states, say $1, \ldots, n$, and
- there is a matrix A such that, if at some time t the system is in state j then for $i = 1, \ldots, n$, the probability that the system is in state i at time $t+1$ is $A[i,j]$.

That is, $A[i,j]$ is the probability of transitioning from j to i, the $j \to i$ *transition probability*.

A is called the *transition matrix* of the Markov chain.

$A[1,1] + A[2,1] + \cdots + A[n,1] = \text{Probability}(1 \to 1) + \text{Probability}(1 \to 2) + \cdots + \text{Probability}(1 \to n) = 1$

Similarly, every column's elements must sum to 1.

Called a *left stochastic matrix* (common convention is to use right stochastic matrices, where every row's elements sum to 1).

Example: $\begin{bmatrix} .44 & .12 \\ .56 & .88 \end{bmatrix}$ is the transition matrix for a two-state Markov chain.

The biggest Markov chain in the world

Randy's web-surfing behavior: From whatever page he's viewing, he selects one of the links uniformly at random and follows it.

Defines a Markov chain in which the states are web pages.

Idea: Suppose this Markov chain has a stationary distribution.

- Find the stationary distribution $\Rightarrow$ probabilities for all web pages.
- Use each web page's probability as a measure of the page's importance.
- When someone searches for "matrix book", which page to return? Among all pages with those terms, return the one with highest probability.

Advantages:

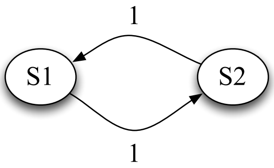
- Computation of stationary distribution is independent of search terms: can be done once and subsequently used for all searches.
- Potentially could use power method to compute stationary distribution.

Pitfalls: There might not be a stationary distribution, or maybe there are several, and how would you compute one?

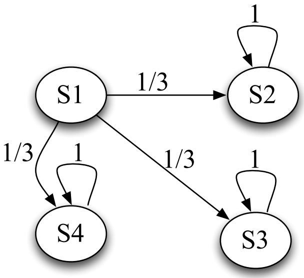
Stationary distributions

Stationary distribution: probability distribution after t iterations gets closer and closer to stationary distribution as t increases.

No stationary distribution:



Several stationary distributions:



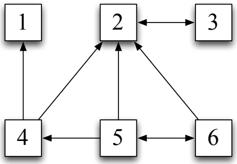
Let A be the transition matrix of a Markov chain. Because column sums are 1, $[1, 1, \ldots, 1] * A = [1, 1, \ldots, 1]$.

Thus 1 is an eigenvalue of A^T. Therefore 1 is an eigenvalue of A.

Can show that there is no eigenvalue with absolute value bigger than 1. However, could be eigenvalues of equal absolute value.

Solve the problem with a hack: In each step, with probability 0.15, Randy just teleports to a web page chosen uniformly at random.

Mix of two distributions



Following random links:

$A_1 =$

	1	2	3	4	5	6
1	1			$\frac{1}{2}$		
2			1	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{2}$
3		1				
4					$\frac{1}{3}$	
5						$\frac{1}{2}$
6					$\frac{1}{3}$	

Uniform distribution: transition matrix of the form

$A_2 =$

	1	2	3	4	5	6
1	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
2	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
3	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
4	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
5	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
6	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Use a mix of the two: incidence matrix is

$$A = 0.85 * A_1 + 0.15 * A_2$$

To find the stationary distribution, find the eigenvector $\mathbf{v}$ corresponding to eigenvalue 1.

How? Use power method, which requires repeated matrix-vector multiplications.

Clever approach to matrix-vector multiplication

$$A = 0.85 * A_1 + 0.15 * A_2$$

$$\begin{aligned} A\,\mathbf{v} &= (0.85 * A_1 + 0.15 * A_2)\mathbf{v} \\ &= 0.85 * (A_1\,\mathbf{v}) + 0.15 * (A_2\,\mathbf{v}) \end{aligned}$$

- Multiplying by A_1: use sparse matrix-vector multiplication you implemented in `Mat`
- Multiplying by A_2: Use the fact that

$$A_2 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{bmatrix}$$