Some applications of linear algebra (only regression is examinable):

- Regression (week 12 p23-27; §6.6): fitting curves to experimental data.
- Dynamical systems and differential equations (§5.6-5.7): stock price example, week 10 p17-19.
- Multivariate calculus: the derivative of a multivariate function is a linear transformation.

Some applications of linear algebra (only regression is examinable):

- Regression (week 12 p23-27; §6.6): fitting curves to experimental data.
- Dynamical systems and differential equations (§5.6-5.7): stock price example, week 10 p17-19.
- Multivariate calculus: the derivative of a multivariate function is a linear transformation.
- Algebraic graph theory (friendship graph example, Homework 3 Q7): if every two people who are friends have a common friends, and every two people who are not friends have b common friends, then the eigenvalues of the associated matrix gives very strong conditions on the values of a and b (see “strongly regular graph”).

Some applications of linear algebra (only regression is examinable):

- Regression (week 12 p23-27; §6.6): fitting curves to experimental data.
- Dynamical systems and differential equations (§5.6-5.7): stock price example, week 10 p17-19.
- Multivariate calculus: the derivative of a multivariate function is a linear transformation.
- Algebraic graph theory (friendship graph example, Homework 3 Q7): if every two people who are friends have a common friends, and every two people who are not friends have b common friends, then the eigenvalues of the associated matrix gives very strong conditions on the values of a and b (see "strongly regular graph").
- Compressed sensing: the (i, j)-entry of the "Netflix matrix" is how much person i likes movie j. Very few entries in this matrix are known, because each person has only seen a small number of movies. Can we guess the entries for the movies that someone has not seen, and therefore recommend him a movie he would like? The guesses assume that the matrix has low rank, i.e. very few factors decide whether someone likes a movie.

§4.9: Markov Chains - applying Linear Algebra to Probability (non-examinable, related to possible final year projects with Dr. Pang)

Example: Every day you eat at one of three places: Starbucks, AAB, or Hall Canteen.

- After eating at Starbucks, you return the next day with probability $\frac{1}{2}$, or go to the other two places with probability $\frac{1}{4}$ each.
- After eating at AAB, you are equally likely to visit each of the three places the next day.
- After eating at Hall Canteen, you don't return the next day; you go to Starbucks with probability $\frac{2}{3}$, or AAB with probability $\frac{1}{3}$.

§4.9: Markov Chains - applying Linear Algebra to Probability (non-examinable, related to possible final year projects with Dr. Pang)

Example: Every day you eat at one of three places: Starbucks, AAB, or Hall Canteen.

- After eating at Starbucks, you return the next day with probability $\frac{1}{2}$, or go to the other two places with probability $\frac{1}{4}$ each.
- After eating at AAB, you are equally likely to visit each of the three places the next day.
- After eating at Hall Canteen, you don't return the next day; you go to Starbucks with probability $\frac{2}{3}$, or AAB with probability $\frac{1}{3}$.

The features of this model that means it is a Markov chain:

- Each day, the "system" is in one of a finite number of possible states.
- The state of the system tomorrow depends only on the state of the system today, not on the past or which day it is.

These features mean that a Markov chain can be encoded in a transition matrix K, whose rows and columns are indexed by the possible states:

$$i, j\text{-entry of } K = \text{ probability of moving from } j \text{ to } i.$$

(Warning: this is the algebraist's convention, as in the textbook. In the probabilist's convention (and mine), the probability above is k_{ji} - i.e. the matrix is transposed.)

Example: In our previous example, the transition matrix is

$$\text{Destination:}\left\{\begin{array}{l} \\ \text{Starbucks} \\ \text{AAB} \\ \text{Hall} \end{array}\right. \begin{array}{c} \overbrace{\begin{array}{ccc}\text{Starbucks} & \text{AAB} & \text{Hall}\end{array}}^{\text{Source:}} \\ \begin{bmatrix} 1/2 & 1/3 & 2/3 \\ 1/4 & 1/3 & 1/3 \\ 1/4 & 1/3 & 0 \end{bmatrix} \end{array}.$$

Because a Markov chain is encoded in a single matrix, it is fast and easy to run a Markov chain on a computer. Hence Markov chains are widely used for simulation.

Here's why the transition matrix is useful:

Example: Suppose you went to Starbucks today. What is the probability of you going to AAB after 2 days?

Answer: There are three possibilities:

$$
\text{Destination: }
\begin{array}{l}
\\ \text{Starbucks} \\ \text{AAB} \\ \text{Hall}
\end{array}
\overbrace{
\begin{array}{ccc}
\text{Starbucks} & \text{AAB} & \text{Hall} \\
\end{array}
}^{\text{Source: Day Two}}
\quad
\overbrace{
\begin{array}{ccc}
\text{Starbucks} & \text{AAB} & \text{Hall} \\
\end{array}
}^{\text{Source: Day One}}
$$

$$
\begin{array}{c|ccc}
 & \text{Starbucks} & \text{AAB} & \text{Hall} \\
\hline
\text{Starbucks} & 1/2 & 1/3 & 2/3 \\
\text{AAB} & 1/4 & 1/3 & 1/3 \\
\text{Hall} & 1/4 & 1/3 & 0
\end{array}
\qquad
\begin{array}{c|ccc}
 & \text{Starbucks} & \text{AAB} & \text{Hall} \\
\hline
 & 1/2 & 1/3 & 2/3 \\
 & 1/4 & 1/3 & 1/3 \\
 & 1/4 & 1/3 & 0
\end{array}
$$

Starbucks $\to$ Starbucks $\to$ AAB: $\frac{1}{2}\frac{1}{4}$

Starbucks $\to$ AAB $\to$ AAB: $\frac{1}{4}\frac{1}{3}$

Starbucks $\to$ Hall $\to$ AAB: $\frac{1}{4}\frac{2}{3}$

So total probability $= \frac{1}{2}\frac{1}{4} + \frac{1}{4}\frac{1}{3} + \frac{1}{4}\frac{2}{3} =$ entry of K^2 in row AAB, column Starbucks.

Similarly, the probability of moving from state j to state i in t steps is the (i, j) entry of K^t.

We are interested in the long term behaviour, i.e. what is the matrix K^t for large t? There are many different ways to investigate this, and one of them is to use eigenvalues.

It can be proved that all eigenvalues of a transition matrix are between -1 and 1, and 1 is always an eigenvalue (because the columns are probabilities summing to 1, see week 5 p26).

Suppose in addition that K is diagonalisable, with eigenvectors $\mathbf{v}_1, \ldots \mathbf{v}_n$ corresponding to $1, \lambda_2 \ldots, \lambda_n$. Also suppose that $|\lambda_i| < 1$ for all $i > 1$ (i.e. the multiplicity of 1 is 1, and -1 is not an eigenvalue). Then, for any weights $c_1, \ldots, c_n$:

$$\begin{aligned} K^t(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 \cdots + c_n\mathbf{v}_n) &= c_1K^t(\mathbf{v}_1) + c_2K^t(\mathbf{v}_2) + \cdots + c_nK^t(\mathbf{v}_n) \\ &= c_1\mathbf{v}_1 + c_2\lambda_2^t(\mathbf{v}_2) + \cdots + c_n\lambda_n^t(\mathbf{v}_n) \to c_1\mathbf{v}_1 \text{ as } t \to \infty, \end{aligned}$$

because all $\lambda_i^t \to 0$.

So the eigenvector of eigenvalue 1 describes the limiting distribution, and the eigenvalue with largest absolute value, after 1 controls the convergence rate.

From previous page: if i) the transition matrix K is diagonalisable, ii) the multiplicity of the eigenvalue 1 is 1, iii) -1 is not an eigenvalue, then the eigenvector of eigenvalue 1 describes the limiting distribution, and the eigenvalue with largest absolute value, after 1 controls the convergence rate. (The other eigenvectors are also interesting: see next page.)

Because the vast majority of matrices is diagonalisable, the result is true without i). And there are easy, natural conditions on the transition probabilities to ensure ii), iii).

From previous page: if i) the transition matrix K is diagonalisable, ii) the multiplicity of the eigenvalue 1 is 1, iii) -1 is not an eigenvalue, then the eigenvector of eigenvalue 1 describes the limiting distribution, and the eigenvalue with largest absolute value, after 1 controls the convergence rate. (The other eigenvectors are also interesting: see next page.)

Because the vast majority of matrices is diagonalisable, the result is true without i). And there are easy, natural conditions on the transition probabilities to ensure ii), iii).

Problem: for many Markov chains in real applications, it is very hard to find the eigenvalues. (Many chains depend on parameters, and the eigenvalues are not nice functions of the parameters.)

What my friends and I decided to do:

- Find some matrices K whose eigenvalues and eigenvectors are easy to find (e.g. matrices for "nice" linear transformations on abstract vector spaces);
- Study the Markov chains whose transition matrix is K;
- Make a connection between our chains and the chains in real applications.

Why eigenvectors of the transition matrix are interesting:

Example: You spend \$30 if you go to Starbucks, \$25 if you go to AAB, and \$20 if you go to Hall Canteen. Suppose you went to Starbucks today. What is the average (expected value) that you will spend tomorrow?

Answer: There are three possibilities:

$$\begin{array}{ccc} \text{Starbucks} & \text{AAB} & \text{Hall} \\ \end{array}$$
$$\begin{bmatrix} 30 & 25 & 20 \end{bmatrix}$$

$$\begin{array}{c} \\ \text{Destination:} \left\{ \begin{array}{l} \text{Starbucks} \\ \text{AAB} \\ \text{Hall} \end{array} \right. \end{array} \begin{array}{c} \overbrace{\begin{array}{ccc} \text{Starbucks} & \text{AAB} & \text{Hall} \end{array}}^{\text{Source:}} \\ \begin{bmatrix} 1/2 & 1/3 & 2/3 \\ 1/4 & 1/3 & 1/3 \\ 1/4 & 1/3 & 0 \end{bmatrix} \end{array}$$

Starbucks: $\frac{1}{2}30$ AAB: $\frac{1}{4}25$ Hall: $\frac{1}{4}20$

So average $= \frac{1}{2}30 + \frac{1}{4}25 + \frac{1}{4}20 =$ entry of $\begin{bmatrix} 30 & 25 & 20 \end{bmatrix} K$ in the column for Starbucks.

Why eigenvectors of the transition matrix are interesting:

Example: You spend \$30 if you go to Starbucks, \$25 if you go to AAB, and \$20 if you go to Hall Canteen. Suppose you went to Starbucks today. What is the average (expected value) that you will spend tomorrow?

Answer: There are three possibilities:

$$\begin{array}{ccc} \text{Starbucks} & \text{AAB} & \text{Hall} \\ \end{array}$$
$$\begin{bmatrix} 30 & 25 & 20 \end{bmatrix}$$

$$\begin{array}{cc} & \overbrace{\begin{array}{ccc}\text{Starbucks} & \text{AAB} & \text{Hall}\end{array}}^{\text{Source:}} \\ \text{Destination:} \left\{ \begin{array}{l} \text{Starbucks} \\ \text{AAB} \\ \text{Hall} \end{array} \right. & \begin{bmatrix} 1/2 & 1/3 & 2/3 \\ 1/4 & 1/3 & 1/3 \\ 1/4 & 1/3 & 0 \end{bmatrix} \end{array}$$

Starbucks: $\frac{1}{2}30$ AAB: $\frac{1}{4}25$ Hall: $\frac{1}{4}20$

So average $= \frac{1}{2}30 + \frac{1}{4}25 + \frac{1}{4}20 =$ entry of $\begin{bmatrix} 30 & 25 & 20 \end{bmatrix} K$ in the column for Starbucks.

Similarly, the expected value of a function f after t steps is $\begin{bmatrix} f(1) & \ldots & f(n) \end{bmatrix} K^t$.
So, if $\begin{bmatrix} f(1) & \ldots & f(n) \end{bmatrix}$ is an eigenvector of K^T corresponding to the eigenvalue λ, then the expected value of f after t steps is λ^t multiplied by its original value, which is very easy to calculate.

What I have:

- A general formula for the eigenvalues and some eigenvectors of many matrices K and their transposes. The matrices K depend on many many parameters.

What I need (i.e. what you might do if you do a project with me):

- Focus on a sequence K_d of these matrices - i.e. fix all the parameters except d;
- Use my general formula to calculate the eigenvalues and eigenvectors of K_d;
- Using the linear transformation represented by K_d, think of a sequence of Markov chains whose transition matrix is K_d;
- By taking linear combinations of the eigenvectors from the formula, find the expected value of some “nice” functions for these Markov chains.

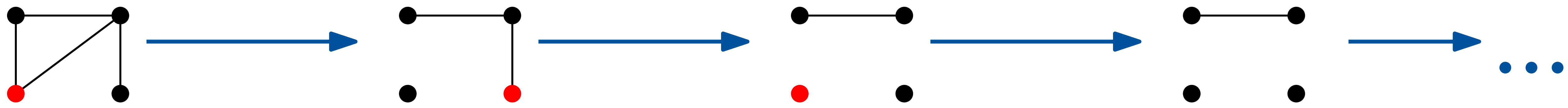
What I have:

- A general formula for the eigenvalues and some eigenvectors of many matrices K and their transposes. The matrices K depend on many many parameters.

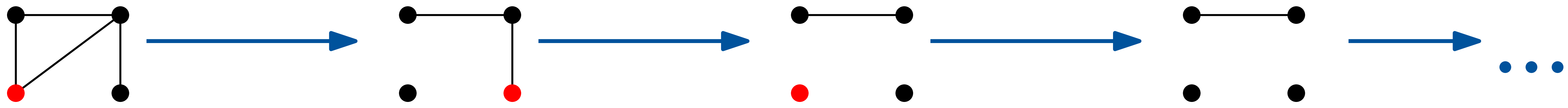
What I need (i.e. what you might do if you do a project with me):

- Focus on a sequence K_d of these matrices - i.e. fix all the parameters except d;
- Use my general formula to calculate the eigenvalues and eigenvectors of K_d;
- Using the linear transformation represented by K_d, think of a sequence of Markov chains whose transition matrix is K_d;
- By taking linear combinations of the eigenvectors from the formula, find the expected value of some "nice" functions for these Markov chains.

Example: One sequence K_d is the transition matrix for this Markov chain: the states are graphs with d vertices, and at each step, disconnect a random vertex. Here is one possibility in the case $d = 4$:



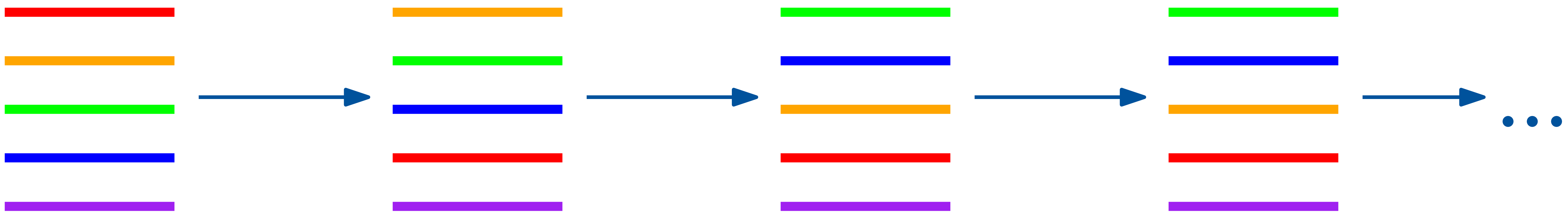
Example: One sequence K_d is the transition matrix for this Markov chain: the states are graphs with d vertices, and at each step, disconnect a random vertex. Here is one possibility in the case $d = 4$:



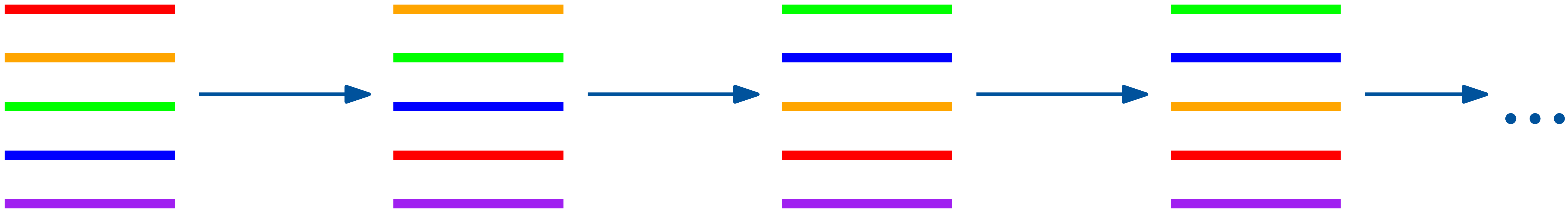
One eigenvector of K_d^T describes the function $G \mapsto$ number of edges in G, and its corresponding eigenvalue is $\frac{d-2}{d}$. So the expected number of edges after t steps is $(\frac{d-2}{d})^t$ multiplied by the starting number of edges.

Another eigenvector of K_d^T describes the function $G \mapsto$ number of triangles in G, and its corresponding eigenvalue is $\frac{d-3}{d}$. So the expected number of triangles after t steps is $(\frac{d-3}{d})^t$ multiplied by the starting number of triangles.

Example: Another sequence K_d is the transition matrix for the "top to random shuffle" of d playing cards: the states are all possible orders of d cards, and at each step, remove the top card and reinsert it at a random position. (This is related to dynamic storage allocation schemes.) Here is a possibility for d=5:



Example: Another sequence K_d is the transition matrix for the "top to random shuffle" of d playing cards: the states are all possible orders of d cards, and at each step, remove the top card and reinsert it at a random position. (This is related to dynamic storage allocation schemes.) Here is a possibility for d=5:



By using two eigenvectors of different eigenvalues, I can show that, after t steps, the probability of the bottom two cards being in the same relative order as they started is $\left(1-\left(\frac{d-2}{d}\right)^t\right)\frac{1}{2}$.

I also have related results about the "riffle shuffle" (the way we usually shuffle: split the deck into two and then "riffle" them together).