Recitation 1. February 12

Focus: recognizing vector spaces, rules of matrix multiplication.

Definition. A *real vector space* V is a set endowed with operations of adding two vectors and multiplying a vector by a real number such that the following holds for any three vectors u, v, w and any real numbers a, b:

- $(u+v)+w = u+(v+w)$;
- $u+v = v+u$;
- there exists a *zero vector* $0 \in V$ such that $v+0 = v$;
- $a(bv) = (ab)v$;
- $1v = v$;
- $(a+b)v = av+bv$;
- $a(u+v) = au+av$.

Elements of a vector space are called *vectors*.

Remark. The most basic rule that you should remember: **row column**. It shows the order in which you write or compute, e.g.:

- The first index denotes the row, the second number the column.
- You multiply a row by a column to get a number.
- An $n \times m$ matrix has n rows and m columns.

Notation. We will denote by A^T the transpose of a matrix A.

1. Is this a vector space? Why / why not? Which natural operations you considered when checking axioms?
   a) The line $y = x$.
   b) The line $y = x+1$.
   c) The union of the x and y axes.
   d) The unit circle $\{(x,y) \mid x^2+y^2 = 1\}$.
   e) The set of 5×5 matrices with the element in position $(3,3)$ being 0.
   f) Functions of the form $f(x) = ax^2+bx+c$.
   g) Functions $f(x)$ with $f(7) = 0$.
   h) Functions $f(x)$ with $f(0) = 7$.

2. *Rules of matrix multiplication. (Section 2.4 of Strang.)* Consider the following matrices:

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 3 & -1 & 2 \end{pmatrix},\ B = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 3 & 2 \end{pmatrix},\ C = \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix},\ D = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix},\ E = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}.$$

   Which of these matrix operations are allowed, and what are the results?
   a) AB
   b) AB^T
   c) $B^T A$
   d) $(A+B)C$
   e) $(A+B)C^T$
   f) $C(A+B)$
   g) DB
   h) BD

i) AE

j) EA

k) CAE

3. When you multiply an $n \times m$ matrix by an $m \times l$ matrix, what are the dimensions of the resulting matrix?

Solution:

4. When can a matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be written as $X^T X$ for some other matrix $X = \begin{pmatrix} p & 0 \\ q & r \end{pmatrix}$? Assume that $b \neq 0$. What are p, q, r in terms of a, b, c, d when possible?

Solution:

5. Consider the following matrices:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},\ B = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix},\ C = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},\ D = \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix}.$$

How is each row of BA, CA, DA related to the rows of A? How is each column of AB, AC, AD related to the columns of A?

Solution:

6. *In this problem, we will practice block multiplication. (Page 75 of Strang.)* Consider the following column vector c and a 3×3 matrix A with columns a_1, a_2, a_3:

$$c = \begin{pmatrix} \lambda \\ \mu \\ \nu \end{pmatrix},\ A = \begin{pmatrix} | & | & | \\ a_1 & a_2 & a_3 \\ | & | & | \end{pmatrix}.$$

Write the result of matrix multiplication rA as a linear combination of the column vectors a_1, a_2, a_3. What if we write a matrix R as three rows $R = \begin{pmatrix} - & r_1 & - \\ - & r_2 & - \\ - & r_3 & - \end{pmatrix}$ and multiply R by A?

Solution:

Recitation 1. Solution

Focus: recognizing vector spaces, rules of matrix multiplication.

We will provide a formal definition of a vector space for the sake of honesty, but for the purpose of solving the first problem of this worksheet, you will need to check just three properties: when you add two elements, you get an element of the same set (being closed under addition); when you multiply an element by a scalar, the result is also in the same set (being closed under multiplication by scalars); and zero should belong to this set. These properties are boxed. Once again: we do not expect you to be able to recite all the axioms, although we definitely would be impressed if you are :)

Definition. A *real vector space* V is a set endowed with operations of adding two vectors and multiplying a vector by a real numbe such that the following holds for any three vectors u, v, w and any real numbers a, b:

- $(u+v)+w = u+(v+w)$;
- $u+v = v+u$;
- there exists a *zero vector* $0 \in V$ such that for any vector v, we have $v+0=v$;
- $a(bv) = (ab)v$;
- $1v = v$;
- $(a+b)v = av + bv$;
- $a(u+v) = au + av$.

Elements of a vector space are called *vectors*.

Remark. The most basic rule that you should remember: **row column**. It shows the order in which you write or compute, e.g.:

- The first index denotes the row, the second number the column.
- You multiply a row by a column to get a number.
- An $n \times m$ matrix has n rows and m columns.

Notation. We will denote by A^T the transpose of a matrix A.

1. Is this a vector space? Why / why not? Which natural operations you considered when checking axioms?

   a) The line $y = x$.
   b) The line $y = x + 1$.
   c) The union of the x and y axes.
   d) The unit circle $\{(x, y) \mid x^2 + y^2 = 1\}$.
   e) The set of 5×5 matrices with the element in position $(3, 3)$ being 0.
   f) Functions of the form $f(x) = ax^2 + bx + c$.
   g) Functions $f(x)$ with $f(7) = 0$.
   h) Functions $f(x)$ with $f(0) = 7$.
   i) *Tricky question.* Newtonian universe.

Solution:

a) Yes.

b) No, because the set is not closed under addition. For example, the points $\binom{-1}{0}$ and $\binom{0}{1}$ belong to the line, but their sum $\binom{-1}{1}$ does not. Neither is the set closed under multiplication or has zero.

c) No, because the set is not closed under addition. Example of points is the same as above. However, it is closed under scaling and contains zero.

d) No, for the same reasons as in part (b).

e) Yes. Adding two matrices or multiplying such a matrix by a number does not affect the property that the middle element is zero.

f) Yes, the set of quadratic polynomials is a vector space.

g) Yes.

h) No, because the set is not closed under addition: if you add two functions f and g with $f(0) = g(0) = 7$, then their sum evaluates to 14 at 0.

i) No, we don't have a zero, because there is no natural reference point.

2. *Rules of matrix multiplication. (Section 2.4 of Strang.)* Consider the following matrices:

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 3 & -1 & 2 \end{pmatrix}, \; B = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 3 & 2 \end{pmatrix}, \; C = \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}, \; D = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}, \; E = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}.$$

Which of these matrix operations are allowed, and what are the results?

a) AB
b) AB^T
c) $B^T A$
d) $(A+B)C$
e) $(A+B)C^T$
f) $C(A+B)$
g) DB
h) BD
i) AE
j) EA
k) CAE

Solution: In order to multiply two matrices, number of columns in the first should be equal to number of rows in the second.

a) AB not allowed: we cannot multiply a 2×3 matrix by a 2×3 matrix.

b) $AB^T = \begin{pmatrix} 1 & 2 & 0 \\ 3 & -1 & 2 \end{pmatrix} \begin{pmatrix} -1 & -1 \\ 0 & 3 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} -1 & 5 \\ -1 & -2 \end{pmatrix}.$

c) $B^T A = \begin{pmatrix} -1 & -1 \\ 0 & 3 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 3 & -1 & 2 \end{pmatrix} = \begin{pmatrix} -4 & -1 & -2 \\ 9 & -3 & 6 \\ 7 & 0 & 4 \end{pmatrix}.$

d) $(A+B)C$ not allowed.

e) $(A+B)C^T$ not allowed.

f) $C(A+B) = \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} 0 & 2 & 1 \\ 2 & 2 & 4 \end{pmatrix} = \begin{pmatrix} 0 & -4 & -2 \\ -4 & -4 & -8 \end{pmatrix}.$

g) DB not allowed.

h) BD not allowed.

i) $AE = \begin{pmatrix} 1 & 2 & 0 \\ 3 & -1 & 2 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 7 \\ 9 \end{pmatrix}$.

j) EA not allowed.

k) $CAE = C(AE) = \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} 7 \\ 9 \end{pmatrix} = \begin{pmatrix} -14 \\ -18 \end{pmatrix}$.

3. When you multiply an $n \times m$ matrix by an $m \times l$ matrix, what are the dimensions of the resulting matrix?

Solution: $(n \times m)(m \times l) \to (n \times l)$.

4. When can a matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be written as $X^T X$ for some other matrix $X = \begin{pmatrix} p & 0 \\ q & r \end{pmatrix}$? Assume that $b \neq 0$. What are p, q, r in terms of a, b, c, d when possible?

Solution: In order to describe the condition, we compare matrix elements in the desired equality $A = X^T X$:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} p & q \\ 0 & r \end{pmatrix} \begin{pmatrix} p & 0 \\ q & r \end{pmatrix} = \begin{pmatrix} p^2 + q^2 & qr \\ qr & r^2 \end{pmatrix}.$$

One can note that matrix elements of $X^T X$ are dot products of the columns of X.

One can immediately observe that A should satisfy $a \geq 0$ and $d \geq 0$, because square of a real number is always nonnegative, and $b = qr = c$, that is A is necessarily symmetric. In addition, if either a or d is zero, then $b = c = 0$ as well, because vanishing of a implies that $q = 0$, so $b = c = qr = 0$, and similarly for d. There will be one more condition which we will find later.

Now we turn to expressing p, q and r in terms of a, b, c and d:

- $r = \sqrt{d}$;
- if $d \neq 0$, then $q = \frac{b}{r} = \frac{b}{\sqrt{d}}$ and $p = \sqrt{a - q^2} = \sqrt{a - \frac{b^2}{d}}$, hence we have an additional constraint that $a - \frac{b^2}{d} \geq 0$;
- if $d = 0$, then $r = 0$ and p and q are any real numbers such that the vector $\begin{pmatrix} p \\ q \end{pmatrix}$ lies on the circle of radius a.

5. Consider the following matrices:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, B = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix}, C = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, D = \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix}.$$

How is each row of BA, CA, DA related to the rows of A? How is each column of AB, AC, AD related to the columns of A?

Solution: I will only write the solution for rows, because what happens to columns is exactly the same after you change the order of multiplication.

- The first row of BA is twice the first row of A, and the second is minus the second row of A.
- The first row of CA is the second row of A, while the second row is zero.
- The first row of DA is the second row of A and the second row of DA is minus the second row of A.

So you can see that multiplying a matrix A by another matrix on the left performs row operations. Similarly, right multiplication performs column operations. You will see more in the next problem.

6. *In this problem, we will practice block multiplication. (Page 75 of Strang.)* Consider the following column vector c and a 3×3 matrix A with columns a_1, a_2, a_3:

$$c = \begin{pmatrix} \lambda \\ \mu \\ \nu \end{pmatrix}, A = \begin{pmatrix} | & | & | \\ a_1 & a_2 & a_3 \\ | & | & | \end{pmatrix}.$$

Write the result of matrix multiplication rA as a linear combination of the column vectors a_1, a_2, a_3. What if we write a matrix R as three rows $R = \begin{pmatrix} - & r_1 & - \\ - & r_2 & - \\ - & r_3 & - \end{pmatrix}$ and multiply R by A?

Solution: First compute Ac, and note that a_1, a_2, a_3 are all 3-vectors:

$$Ac = \begin{pmatrix} | & | & | \\ a_1 & a_2 & a_3 \\ | & | & | \end{pmatrix} \begin{pmatrix} \lambda \\ \mu \\ \nu \end{pmatrix} = \lambda a_1 + \mu a_2 + \nu a_3.$$

This is a particular case of block multiplication.

Now calculate RA – this turns out to be the usual rule of matrix multiplication:

$$RA = \begin{pmatrix} - & r_1 & - \\ - & r_2 & - \\ - & r_3 & - \end{pmatrix} \begin{pmatrix} | & | & | \\ a_1 & a_2 & a_3 \\ | & | & | \end{pmatrix} = \begin{pmatrix} r_1a_1 & r_1a_2 & r_1a_3 \\ r_2a_1 & r_2a_2 & r_2a_3 \\ r_3a_1 & r_3a_2 & r_3a_3 \end{pmatrix}.$$

♡ Week 2 Extra ♡

Focus: LU decomposition (aka Gaussian elimination), orthogonal matrices.

LU factorization of a matrix A is a way of writing A as a product of two matrices $\boxed{A = LU}$, where L is a lower-diagonal matrix with units on the diagonal and U is an upper-diagonal matrix.

Definition. A square matrix A is called *orthogonal* if $A^T A = I$. (The unit matrix is denoted by I.)

1. Compute the result of multiplying a row $r = \begin{pmatrix} r_1 & \cdots & r_n \end{pmatrix}$ by a matrix M with rows $M_1, \ldots, M_n$.

Solution:

2. If R is a $k \times n$ matrix and M is a matrix with n rows $M_1, \ldots, M_n$, what are the rows of RM in terms of $M_1, \ldots, M_n$ and the matrix coefficients of R?

Solution:

3. *Bonus. LU factorization = Gaussian elimination.* Solve the system of linear equations using LU factorization:

$$\begin{cases} x + 2y + 3z = 1, \\ \quad\ \ y + \ z = 2, \\ 3x + y - z = 3. \end{cases}$$

Solution:

4. *Not all matrices can be written in LU form.* Show directly why these matrix equations are both impossible (empty spaces mean zeroes):

   a) $\begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ l & 1 \end{pmatrix} \begin{pmatrix} d & e \\ 0 & f \end{pmatrix}$;

   b) $\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & & \\ l & 1 & \\ m & n & 1 \end{pmatrix} \begin{pmatrix} d & e & g \\ & f & h \\ & & i \end{pmatrix}$.

Solution:

5. *Orthogonal matrices.* Find $A^T A$ if the columns of A are unit vectors, all mutually perpendicular (in this situation, we say that the vectors are *orthonormal*). What if we ask that the rows of A are orthonormal?

Solution:

6. *Bonus.* Say that a square matrix A is factored as a product $A = B^{-1}C$. Perform the same row operation on both B and C, for example add the first row to the second: $B \mapsto B'$, $C \mapsto C'$. Show that this operation does not change the result, that is $A = B^{-1}C = (B')^{-1}C'$.

Hint: think about the first two problems.

Solution:

7. *Binomial formula for matrices.* Show that $(A+B)^2$ is different from $A^2 + 2AB + B^2$ when

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix}, \; B = \begin{pmatrix} 1 & 0 \\ 3 & 0 \end{pmatrix}.$$

Write down the correct rule: $(A+B)^2 = A^2 + \cdots + B^2$. Can you generalize the rule to $(A+B)^n$?

Solution:

Week 3 Review Session

Focus: rules of matrix multiplication, orthogonal matrices, rotation matrices.

1. When you multiply an $n \times m$ matrix by an $m \times l$ matrix, what are the dimensions of the resulting matrix?

Solution:

2. *Zero scalar vs zero vector vs zero matrix.* Let A be an $n \times m$ matrix, B be an $m \times l$ matrix, v be a column m-vector and r be a row m-vector, for example:

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 1 & 0 & 2 \end{pmatrix}, \; B = \begin{pmatrix} 2 & -2 \\ -1 & 1 \\ -1 & 1 \end{pmatrix}, \; v = \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix}, \; r = \begin{pmatrix} 1 & 2 & 0 \end{pmatrix}.$$

In this case, products AB, Av and rv are all zero, so we can write $AB = 0$, $Av = 0$, $rv = 0$. But would it make sense to write $AB = Av = rv = 0$? Why / why not? Do the results of those operations belong to the same vector space?

Solution:

3. Answer the following questions. Provide explanations.

   a) Is the identity matrix always square? (By the way it can be stored with one parameter.)

   b) Do rectangular matrices have inverses?

   c) Do all square matrices have inverses?

   d) What is the condition for a 2×2 matrix to have inverse?

Solution:

4. *Parallel planes.*

   a) Consider the set of points $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ in $\mathbb{R}^3$ satisfying condition $2x + 3y + 4z = 0$. Describe this geometric object. Find a normal vector to it.

   b) What if we consider the equation $2x + 3y + 4z = 1$? Why is the normal the same?

Solution:

5. *Row and column operations as matrix multiplication. (Inspired by problem 2.4.8 from Strang.)* Consider the following matrices:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \; B = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix}, \; C = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \; D = \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix}.$$

   How is each row of BA, CA, DA related to the rows of A? How is each column of AB, AC, AD related to the columns of A?

Solution:

6. *(Problem 2.4.5 from Strang.)* Compute A^2 and A^3. Make a prediction for A^5 amd A^n:

   a) $A = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$;

   b) $A = \begin{pmatrix} 2 & 2 \\ 0 & 0 \end{pmatrix}$.

Solution:

7. *Binomial formula for matrices. Matrices do not commute. (Problem 2.4.6 from Strang.)* Show that $(A+B)^2$ is different from $A^2 + 2AB + B^2$ when

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix}, \ B = \begin{pmatrix} 1 & 0 \\ 3 & 0 \end{pmatrix}.$$

Write down the correct rule: $(A+B)^2 = A^2 + \cdots + B^2$. Can you generalize the rule to $(A+B)^n$?

Solution:

8. *(Problem 2.4.7 from Strang.)* Is the following true or false? Give counterexamples when false. Matrices A, B and C are such that all the operations are well-defined.

   a) If columns 1 and 3 of B are the same, then so are columns 1 and 3 of AB.

   b) If rows 1 and 3 of B are the same, then so are rows 1 and 3 of AB.

   c) If rows 1 and 3 of A are the same, then so are rows 1 and 3 of ABC.

   d) $(AB)^2 = A^2B^2$.

Solution:

9. *Orthogonal matrices.* Find A^TA if the columns of A are unit vectors, all mutually perpendicular (in this situation, we say that the vectors are *orthonormal*). What if we ask that the rows of A are orthonormal?

Solution:

10. *Defining a matrix by its image. Rotation matrices.* Work out these questions for 2×2 matrices.

a) If we want a matrix A to send vector $\left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)$ to twice itself and vector $\left(\begin{smallmatrix}0\\1\end{smallmatrix}\right)$ to $\left(\begin{smallmatrix}a\\b\end{smallmatrix}\right)$, then what are the matrix entries of A?

b) If we want a matrix B to send vector $\left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)$ $\left(\begin{smallmatrix}a\\b\end{smallmatrix}\right)$ and vector $\left(\begin{smallmatrix}1\\1\end{smallmatrix}\right)$ to twice itself, then what are the matrix entries of B?

c) What are the coordinates of vectors $\left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}0\\1\end{smallmatrix}\right)$ after we rotate them by the angle θ?

d) How do you write a matrix that rotates every vector in the plane by the angle θ?

e) Is the matrix in the previous part orthogonal?

11. *Angles between vectors.* Consider an n-cube C – the set of points in $\mathbb{R}^n$ all of whose coordinates vary from 0 to 1.

a) In a two-dimensional cube, find the angle between the diagonal and an edge.

b) In a three-dimensional cube, find the angle between the long diagonal and an edge.

c) In a three-dimensional cube, find the angle between the long diagonal and a face.

d) In an n-dimensional cube, find the angle between the long diagonal and an edge.

Recitation 2. February 26

Focus: QR decomposition, SVD, least squares.

1. Say that a square matrix A is factored as a product $A = BC^{-1}$. Perform the same column operation on both B and C, for example add the first column to the second: $B \mapsto B'$, $C \mapsto C'$. Show that this operation does not change the result, that is $A = BC^{-1} = B'(C')^{-1}$.

Solution:

2. *Finding a QR decomposition.* Write the following matrix A as a product $A = QR$ for some orthogonal Q and upper-triangular R:

$$A = \begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix}.$$

Solution:

3. *Finding an SVD.* Consider matrix A:

$$A = \begin{pmatrix} 1 & 3 & -1 \\ -2 & -6 & 2 \\ 3 & 9 & -3 \end{pmatrix}.$$

   a) Describe its column space.
   b) Express A as an outer product of two vectors. Is this decomposition unique?
   c) Find a compact and a full form SVD for this matrix.

Solution:

Solution:

4. *Least squares approximation.* Consider the set of functions of the form $f(x) = ae^x + be^{-x}$, where a and b vary over real numbers. In this space of functions, use the least squares algorithm to approximate the unknown function U that takes the following values:

$$U\left(\ln\frac{1}{2}\right) = 1,\ U(0) = 1,\ U(\ln 2) = 2.$$

You will get the following picture:

pictures/wk4.jpeg

a) Write the sum $S(a, b)$ of squared errors.
b) Write the condition of finding local minimum using partial derivatives with respect to a and b.
c) Write the condition above as a matrix equation and solve this equation.

Solution:

Recitation 2. Solution

Focus: QR decomposition, SVD, least squares.

1. Say that a square matrix A is factored as a product $A = BC^{-1}$. Perform the same column operation on both B and C, for example add the first column to the second: $B \mapsto B'$, $C \mapsto C'$. Show that this operation does not change the result, that is $A = BC^{-1} = B'(C')^{-1}$.

Solution: A column operation on B corresponds to multiplying B on the right by an appropriate invertible matrix, say M. Assuming this rule for a moment, we observe that then $B' = BM$ and $C' = CM$, because we perform the same column operation on both matrices. Therefore, plugging these formulas in, we get $B'(C')^{-1} = (BM)(CM)^{-1}$, and since taking inverses switches the order of factors, we can continue the string of equalities as $B'(C')^{-1} = BMM^{-1}C^{-1} = BC^{-1} = A$. So we achieved the desired result.

In order to understand why multiplying by M performs column operations on B, first write B as a block matrix $B = \begin{pmatrix} b_1 & \cdots & b_n \end{pmatrix}$, where b_i are columns of B. Then we can multiply B by M as block matrices (M is considered to have only trivial 1×1 blocks), and each column of BM would be expressed as a linear combination of columns of B with coefficients from some column of M. For example, if $M = (m_{ij})$, then the first column of BM is:

$$\begin{pmatrix} b_1 & \cdots & b_n \end{pmatrix} \begin{pmatrix} m_{11} \\ \vdots \\ m_{n1} \end{pmatrix} = m_{11}b_1 + m_{21}b_2 + \cdots + m_{n1}b_n.$$

This should have concluded explanation, but it might still look confusing, so let's consider several examples. First, take a 3×3 matrix B written in its block form $B = \begin{pmatrix} b_1 & b_2 & b_3 \end{pmatrix}$ and try to find such a matrix M that multiplication by M on the right would add twice the first column to the third. I claim that the following matrix works (blank spaces are zeroes):

$$M = \begin{pmatrix} 1 & 0 & 2 \\ & 1 & 0 \\ & & 1 \end{pmatrix}.$$

Let's check it (don't forget that the b_i are all column 3-vectors):

$$BM = \begin{pmatrix} b_1 & b_2 & b_3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 2 \\ & 1 & 0 \\ & & 1 \end{pmatrix} = \begin{pmatrix} b_1 & b_2 & 2b_1 + b_3 \end{pmatrix}.$$

For the second example, I want to scale up the second column of B by a factor of $\sqrt{5}$. Then I need to take the following matrix:

$$M = \begin{pmatrix} 1 & & \\ & \sqrt{5} & \\ & & 1 \end{pmatrix}.$$

You can check that BM looks like B with its second column scaled by direct block multiplication as above.

Finally, let's say that I want to switch the second and the third column (although this operation will not be used in the remainder of the worksheet). Then I need to take an appropriate *permutation matrix*:

$$M = \begin{pmatrix} 1 & & \\ & 0 & 1 \\ & 1 & 0 \end{pmatrix}.$$

2. *Finding a QR decomposition.* Write the following matrix A as a product $A = QR$ for some orthogonal Q and upper-triangular R:

$$A = \begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix}.$$

Solution: In this problem, I want to apply the previous one by writing A as a product AI^{-1}, where I is the identity matrix, and then perform column operations on both A and I simultaneously until the first matrix becomes orthogonal and we get $A = QR$. In order to orthogonalize the first factor, I will use a modified version of Gram-Schmidt process.

Recall that Gram-Schmidt algorithm goes as follows:

- Normalize the first vector.
- Subtract a suitable scalar multiple of the first vector from all the rest vectors so that they become orthogonal to the first vector.
- Normalize the second vector.
- Subtract a suitable scalar multiple of the first vector from the third and later vectors so that they become orthogonal to the second vector. Note that they stay orthogonal to the first vector.
- Etc.

For the purpose of hand calculation, I find it bothersome to normalize vectors before orthogonalizing, because I wouldn't like to carry a lot of square roots around. So I will use the following algorithm for the given 2×2 matrix:

- Subtract a suitable scalar multiple of the first vector from the second so that the result becomes orthogonal to the first vector.
- Normalize the first vector.
- Normalize the second vector.
- Now we have Q explicitly. Compute R.

Note that since we are subtracting multiples of the first column to the second, the second factor becomes upper-triangular as required. More generally, since Gram-Schmidt algorithm subtracts multiples of one column from those to the right, the second factor stays upper-triangular. See the first example in the first problem.

First step. Find a scalar λ such that $\begin{pmatrix} 4 \\ 3 \end{pmatrix} - \lambda \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ is orthogonal to $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$. I claim that $\lambda = \dfrac{\begin{pmatrix} 1 \\ 2 \end{pmatrix} \cdot \begin{pmatrix} 4 \\ 3 \end{pmatrix}}{\begin{pmatrix} 1 \\ 2 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix}} = \frac{4+6}{1+4} = 2$ works. Note that dots denote dot product here. Then subtracting the twice the first column from the second in both factor gives us to the following equality (do not forget that the second matrix is inverted!):

$$A = \begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}^{-1}.$$

Now columns of the first factor are orthogonal.

Second step. The magnitude of the first column is $\sqrt{1 + 2^2} = \sqrt{5}$, so we divide the first column of both matrices by $\sqrt{5}$:

$$A = \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{\sqrt{5}} & 2 \\ \frac{2}{\sqrt{5}} & -1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & -2 \\ 0 & 1 \end{pmatrix}^{-1}.$$

Third step. The magnitude of the second column is also $\sqrt{5}$, so we divide the second column of both matrices by $\sqrt{5}$:

$$A = \begin{pmatrix} \frac{1}{\sqrt{5}} & 2 \\ \frac{2}{\sqrt{5}} & -1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & -2 \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{5}} \\ 0 & \frac{1}{\sqrt{5}} \end{pmatrix}^{-1}.$$

Fourth step. Compute the inverse. For that, we first compute the quatity $ad-bc$, because we will need to divide by it. So $ad - bc = \frac{1}{\sqrt{5}} \cdot \frac{1}{\sqrt{5}} - 0 = \frac{1}{5}$. Dividing by it means multiplying by 5. So apply our formula for inverses of 2×2 matrices:

$$A = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{5}} \\ 0 & \frac{1}{\sqrt{5}} \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \end{pmatrix} 5 \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ 0 & \frac{1}{\sqrt{5}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} \sqrt{5} & 2\sqrt{5} \\ 0 & \sqrt{5} \end{pmatrix}.$$

3. *Finding an SVD.* Consider matrix A:

$$A = \begin{pmatrix} 1 & 3 & -1 \\ -2 & -6 & 2 \\ 3 & 9 & -3 \end{pmatrix}.$$

a) Describe its column space.

b) Express A as an outer product of two vectors. Is this decomposition unique?

c) Find a compact and a full form SVD for this matrix.

Solution:

a) $\text{Col}(A) = \text{Span} \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}$, because all other columns are scalar multiples of the first one.

b) $A = \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix} \begin{pmatrix} 1 & 3 & -1 \end{pmatrix}$. The decomposition is not unique, because for example, we can multiply the column by 5 and divide the row by 5, and the result will not change.

c) A compact form SVD follows almost immediately from the column-row decomposition – what remains is to normalize the row and teh column. Note that (1) denotes a 1×1 matrix as opposed to a scalar 1.

$$A = \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix} (1) \begin{pmatrix} 1 & 3 & -1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{14}} \\ \frac{-2}{\sqrt{14}} \\ \frac{3}{\sqrt{14}} \end{pmatrix} (\sqrt{154}) \begin{pmatrix} \frac{1}{\sqrt{11}} & \frac{3}{\sqrt{11}} & \frac{-1}{\sqrt{11}} \end{pmatrix}.$$

A full form SVD can be found by picking two complementary orthogonal vector to the column, and then to the row. Here is an example for the column:

$$A = \begin{pmatrix} \frac{1}{\sqrt{14}} & \frac{1}{\sqrt{3}} & \frac{5}{\sqrt{42}} \\ \frac{-2}{\sqrt{14}} & \frac{-1}{\sqrt{3}} & \frac{4}{\sqrt{42}} \\ \frac{3}{\sqrt{14}} & \frac{-1}{\sqrt{3}} & \frac{1}{\sqrt{42}} \end{pmatrix} \begin{pmatrix} \sqrt{154} & & \\ & 0 & \\ & & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{11}} & \frac{3}{\sqrt{11}} & \frac{-1}{\sqrt{11}} \\ * & * & * \\ * & * & * \end{pmatrix}.$$

4. *Least squares approximation.* Consider the set of functions of the form $f(x) = ae^x + be^{-x}$, where a and b vary over real numbers. In this space of functions, use the least squares algorithm to approximate the unknown function U that takes the following values:

$$U\left(\ln \frac{1}{2}\right) = 1,\ U(0) = 1,\ U(\ln 2) = 2.$$

You will get the following picture:

pictures/wk4.jpeg

a) Write the sum $S(a,b)$ of squared errors.

b) Write the condition of finding local minimum using partial derivatives with respect to a and b.

c) Write the condition above as a matrix equation and solve this equation.

Solution:

a) $S(a,b) = \sum_{i=1}^{n} (f(x_i) - y_i)^2$. In our case, $n = 3$ and the datapoints are given by the values of U, so we plug this in: $S(a,b) = \left(\frac{1}{2}a + 2b - 1\right)^2 + (a+b-1)^2 \left(2a + \frac{1}{2}b - 2\right)^2$.

b) First compute partial derivative with respect to a:

$$\frac{\partial S}{\partial a}(a,b) = 2\sum_{i=1}^{n} \left(ae^{x_i} + be^{-x_i} - y_i\right) e^{x_i} = 2\sum_{i=1}^{n} \left(ae^{2x_i} + b - y_i e^{x_i}\right).$$

Combine similar summands:

$$\frac{\partial S}{\partial a}(a,b) = 2\left(\sum_{i=1}^{n} e^{2x_i}\right) a + 2\left(\sum_{i=1}^{n} 1\right) b - 2\left(\sum_{i=1}^{n} y_i e^{x_i}\right).$$

Now we can plug in our datapoints:

$$\frac{\partial S}{\partial a}(a,b) = \frac{21}{4}a + 3b - \frac{11}{2}.$$

Similarly for b:

$$\frac{\partial S}{\partial b}(a,b) = 2\left(\sum_{i=1}^{n} 1\right) a + 2\left(\sum_{i=1}^{n} e^{-2x_i}\right) b - 2\left(\sum_{i=1}^{n} y_i e^{-x_i}\right).$$

And now with the given data:

$$\frac{\partial S}{\partial b}(a,b) = 3a + \frac{21}{4}b - 4.$$

The condition for finding an extremum is for all the partial derivatives to vanish, so:

$$\begin{cases} \frac{21}{4}a + \ 3b = \frac{11}{2}, \\ 3a + \frac{21}{4}b = 4. \end{cases}$$

c) $\begin{pmatrix} \frac{21}{4} & 3 \\ 3 & \frac{21}{4} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \frac{11}{2} \\ 4 \end{pmatrix}$. Solution can be obtained by multiplying both sides by the inverse of $\begin{pmatrix} \frac{21}{4} & 3 \\ 3 & \frac{21}{4} \end{pmatrix}$ on the left. The resulting coefficients are $a = \frac{30}{33}$ and $b = \frac{8}{33}$. You can find the plot of the corresponding function above.

Recitation 3. March 5

Focus: exam comments, linear independence, projections.

Definition. Recall that if A is a matrix with independent columns, then the *projection matrix* P on the column space of A can be written as $P = A(A^TA)^{-1}A^T$.

1. *Recognizing vector spaces.* Is the following set a vector space?

    a) The set of all vectors in $\mathbb{R}^3$ except those of the form $\begin{pmatrix} x & 0 & 0 \end{pmatrix}^T$ with $x > 0$.

    b) The set of 2×3 matrices whose six elements sum to 6.

    c) The set of rank one 3×3 matrices together with the zero matrix.

Solution:

2. *Vector spaces and bases.* Let V be the space of homogeneous quadratic polynomials in two variables, i.e. polynomials of the form $f(x, y) = ax^2 + bxy + cy^2$.

    a) Are elements x^2, xy, y^2 linearly independent?

    b) What about x^2, $x^2 + xy + y^2$, $xy + y^2$?

    c) And x^2, $x^2 + xy$, $x^2 + xy + y^2$?

    d) What is the dimension of this vector space V?

Solution:

3. *Subspaces.* Let V be the space from the previous problem. Consider the subset of V that consists of functions $f \in V$ such that $f(1, 1) = 0$. Denote this subset by W – it is a vector subspace of V.

    a) Prove that W is a vector space.

    b) Find a basis of this vector space.

    c) What is the dimension of W?

Solution:

4. *Projection onto a subspace.* Consider the following matrix A written as a full SVD $A = U\Sigma V^T$:

$$A = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 10 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}^T.$$

a) What is the rank of A?

b) Mark U_1, U_2, V_1 and V_2.

c) Circle columns of U that span col A.

d) Compute the projection matrix on col A.

Solution:

5. *Bonus. Might be useful for PSet 4. (Or not.)* If A is decomposed into a product $A = BC$ with C being square invertible, then col A = col B.

6. *Bonus.* We declare x^2, xy and y^2 to be the standard orthonormal basis in V, that is we write them as follows:

$$x^2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \; xy = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \; y^2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Compute the projection matrix on W.

Recitation 3. Solution

Focus: exam comments, linear independence, projections.

Definition. Recall that if A is a matrix with independent columns, then the *projection matrix* P on the column space of A can be written as $P = A(A^TA)^{-1}A^T$.

1. *Recognizing vector spaces.* Is the following set a vector space?

   a) The set of all vectors in $\mathbb{R}^3$ except those of the form $\begin{pmatrix} x & 0 & 0 \end{pmatrix}^T$ with $x > 0$.

   b) The set of 2×3 matrices whose six elements sum to 6.

   c) The set of rank one 3×3 matrices together with the zero matrix.

2. *Vector spaces and bases.* Let V be the space of homogeneous quadratic polynomials in two variables, i.e. polynomials of the form $f(x,y) = ax^2 + bxy + cy^2$.

   a) Are elements x^2, xy, y^2 linearly independent?

   b) What about x^2, $x^2 + xy + y^2$, $xy + y^2$?

   c) And x^2, $x^2 + xy$, $x^2 + xy + y^2$?

   d) What is the dimension of this vector space V?

Solution:

a) Yes. Check the definition: assume that we have a linear combination $f(x,y) = ax^2 + bxy + cy^2 = 0$. Then for every pair of real numbers (x_0, y_0), we have that $f(x_0, y_0)$. In particular, $f(1,0) = a = 0$, hence the linear combination can be simplified to $f(x,y) = bxy + cy^2$. Similarly, $f(0,1) = c = 0$ and $f(x,y) = bxy$. And finally, $f(1,1) = b = 0$, hence all coefficients in the linear combination are zero: $a = b = c = 0$. So by definition, monomials x^2, xy, y^2 are linearly independent.

b) No, because $x^2 - (x^2 + xy + y^2) + (xy + y^2) = 0$.

c) Yes.

d) $\dim V = 3$, because monomials x^2, xy, y^2 generate V and by part (a) are linearly independent, hence form a basis.

3. *Subspaces.* Let V be the space from the previous problem. Consider the subset of V that consists of functions $f \in V$ such that $f(1,1) = 0$. Denote this subset by W – it is a vector subspace of V.

   a) Prove that W is a vector space.

   b) Find a basis of this vector space.

   c) What is the dimension of W?

Solution:

a)

b) $x^2 - xy$, $y^2 - xy$.

c) $\dim W = 2$.

4. *Projection onto a subspace.* Consider the following matrix A written as a full SVD $A = U\Sigma V^T$:

$$A = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 10 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}^T.$$

a) What is the rank of A?

b) Mark U_1, U_2, V_1 and V_2.

c) Circle columns of U that span $\operatorname{col} A$.

d) Compute the projection matrix on $\operatorname{col} A$.

Solution:

a) $A = 2$.

b) $U_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \end{pmatrix}$; $U_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}$; $V_1 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{pmatrix}$; $V_2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$.

c) $\operatorname{col} A = \operatorname{col} U_1$, so we should circle $\begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \end{pmatrix}$.

d) $\operatorname{col} A = \operatorname{col} U_1$, so we can compute the projection matrix P on $\operatorname{col} U_1$. But U_1 is a matrix with linearly independent columns, so we can use one of the formulas for computing projections. In this case, we can use the formula for the QR decomposition, because in the QR decomposition of $U_1 = QR$, we have $U_1 = Q$ already, because U_1 is tall skinny orthogonal from SVD. Therefore, we can use the formula $P = QQ^T = U_1 U_1^T = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

5. *Bonus. Might be useful for PSet 4. (Or not.)* If A is decomposed into a product $A = BC$ with C being square invertible, then $\operatorname{col} A = \operatorname{col} B$.

6. *Bonus.* We declare x^2, xy and y^2 to be the standard orthonormal basis in V, that is we write them as follows:

$$x^2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \; xy = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \; y^2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Compute the projection matrix on W.

Solution: In the given basis, coordinate vectors of the basis of W found in 3(b) is $x^2 - xy = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$, $y^2 - xy = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}$.

Then $W = \operatorname{col} \begin{pmatrix} 1 & 0 \\ -1 & -1 \\ 0 & 1 \end{pmatrix}$. Denote this matrix by A, and the projection matrix by P. Now we can either use the formula $P = A(A^T A)^{-1} A^T$ or compute QR decomposition of A and use $P = QQ^T$.

Recitation 4. March 12

Focus: bases, four fundamental subspaces, fitting everything together.

Notation. Let V and W denote two real vector spaces.

Definition (reminder). Vectors $v_1, \ldots, v_k$ are said to be *linearly independent* if the only way to write a zero linear combination $c_1 v_1 + \cdots + c_k v_k = 0$ is to let all the scalars be zero: $c_1 = \cdots = c_k = 0$.

Definition (reminder). The *span*, or *linear span*, of some set of vectors $S \subset V$ is the set of all possible finite linear combinations of vectors from S, or mathematically:

$$\operatorname{Span} S = \{c_1 v_1 + \cdots + c_l v_l \mid l \in \mathbb{Z};\ v_1, \ldots, v_l \in V;\ c_1, \ldots, c_l \in \mathbb{R}\}.$$

The set S can be finite or infinite, and it can be linearly independent or linearly dependent. If $\operatorname{Span} S = V$, then we say that S *generates*, or *spans*, the vector space V.

Definition (reminder). A set of vectors $v_1, \ldots, v_n$ is called a *basis* of V if it these vectors are linearly independent and span V. In this case, we say that V is n-dimensional. All bases in the same vector space have equal number of elements.

Definition. A *linear operator*, or a *linear transformation*, between vector spaces V amd W is a set function $A : V \to W$ that is linear, which means that $A(v + v') = Av + Av'$ for vectors v and v' in V, and $A(\lambda v) = \lambda A v$ for a vector $v \in V$ and a scalar $\lambda \in \mathbb{R}$.

Definition. The *image* of a linear operator $A : V \to W$ is a subset of W that consists of all vectors of the form Av for $v \in V$, or mathematically: $\operatorname{Im} A = \{Av \mid v \in V\}$.

Definition. The *kernel* of a linear operator $A : V \to W$ is a subset of V that consists of all vectors that are sent to zero, or mathematically: $\operatorname{Ker} A = \{v \in V \mid Av = 0\}$.

Definition. The *rank* of a linear operator $A : V \to W$ is the dimension of its image $\dim \operatorname{Im} A$.

1. Prove that $\operatorname{Im} A$ and $\operatorname{Ker} A$ are vector subspaces of W and V, respectively.

   Solution:

2. How can an $m \times n$ matrix be viewed as a linear transformation? What are the dimensions of the two vector spaces?

   Solution:

3. Let us consider a matrix A as a linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$. Let $e_1, \ldots, e_n$ be the *standard basis vectors* of $\mathbb{R}^n$, that is: $e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, $e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}$, $\ldots$, $e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$. Describe $Ae_1, \ldots, Ae_n$ in terms of A. Conclusion: we can define a linear operator $A : V \to W$ by its action on a basis of V.

   Solution:

4. Let V be the space of polynomials in two variables of the form $f(x, y) = a + bx + cy + dx^2$, and let W be the space of degree one polynomials in two variables.

   a) Find (the simplest) bases of V and W. What are the dimensions of these spaces?
   b) Consider a linear operator $A = \frac{d}{dx}$ from V to W. Write A as a matrix in the bases that we found in part (a).
   c) What are the nullspace and column space of A? What are the kernel and image of $\frac{d}{dx}$? What is the conclusion?
   d) What is the rank of A?
   e) *Bonus.* Let us add twice the second column of A to the first, and denote the new matrix (linear transformation) by A'. How did the transformation change?
   f) *Bonus.* Write A' as a composition of A and some other linear transformation M. What are the vector spaces that M operates between?

Hint: recall column operations and how they are related to matrix multiplication on the right.

Solution:

5. Fix a linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$ of rank r. Describe the relations between the four fundamental subspaces in terms of kernel and image. *Tricky question:* Would you be able to do that if we said that $A : V \to W$ with the same rank and dimensions of the spaces?

Solution:

6. Fix a linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$. Understand that if $b \in \mathbb{R}^m$ is in the image of A, then the system $Ax = b$ has a solution, say x_0. In this case, show in addition that the space of all solutions is $x_0 + \operatorname{Ker} A$. Now conclude that in the case of nonzero kernel (nullspace), the system $Ax = b$ has either infinitely many solutions or no solutions at all, depending on whether $b \in \operatorname{Im} A$ or not.

Recitation 4. Solution

Focus: bases, four fundamental subspaces, fitting everything together.

Notation. Let V and W denote two real vector spaces.

Definition (reminder). Vectors $v_1, \ldots, v_k$ are said to be *linearly independent* if the only way to write a zero linear combination $c_1v_1 + \cdots + c_kv_k = 0$ is to let all the scalars be zero: $c_1 = \cdots = c_k = 0$.

Definition (reminder). The *span*, or *linear span*, of some set of vectors $S \subset V$ is the set of all possible finite linear combinations of vectors from S, or mathematically:

$$\operatorname{Span} S = \{c_1v_1 + \cdots + c_lv_l \mid l \in \mathbb{Z};\ v_1, \ldots, v_l \in V;\ c_1, \ldots, c_l \in \mathbb{R}\}.$$

The set S can be finite or infinite, and it can be linearly independent or linearly dependent. If $\operatorname{Span} S = V$, then we say that S *generates*, or *spans*, the vector space V.

Definition (reminder). A set of vectors $v_1, \ldots, v_n$ is called a *basis* of V if it these vectors are linearly independent and span V. In this case, we say that V is n-dimensional. All bases in the same vector space have equal number of elements.

Definition. A *linear operator*, or a *linear transformation*, between vector spaces V amd W is a set function $A : V \to W$ that is linear, which means that $A(v + v') = Av + Av'$ for vectors v and v' in V, and $A(\lambda v) = \lambda Av$ for a vector $v \in V$ and a scalar $\lambda \in \mathbb{R}$.

Definition. The *image* of a linear operator $A : V \to W$ is a subset of W that consists of all vectors of the form Av for $v \in V$, or mathematically: $\operatorname{Im} A = \{Av \mid v \in V\}$.

Definition. The *kernel* of a linear operator $A : V \to W$ is a subset of V that consists of all vectors that are sent to zero, or mathematically: $\operatorname{Ker} A = \{v \in V \mid Av = 0\}$.

Definition. The *rank* of a linear operator $A : V \to W$ is the dimension of its image $\dim \operatorname{Im} A$.

1. Prove that $\operatorname{Im} A$ and $\operatorname{Ker} A$ are vector subspaces of W and V, respectively.

   Solution: Need to check that both are closed under addition, multiplication by a scalar and contain the zero vector.

2. How can an $m \times n$ matrix be viewed as a linear transformation? What are the dimensions of the two vector spaces?

   Solution: Denote this $m \times n$ matrix by A. Then we can define a map $\mathbb{R}^n \to \mathbb{R}^m$ which we also denote by A as follows: whenever we have a vector $v \in \mathbb{R}^n$, we send it to Av as defined by matrix multiplication. So we can use the words "matrix" and "linear transformation" (almost) as synonyms.

3. Let us consider a matrix A as a linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$. Let $e_1, \ldots, e_n$ be the *standard basis vectors* of $\mathbb{R}^n$, that is: $e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, $e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}$, $\ldots$, $e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$. Describe $Ae_1, \ldots, Ae_n$ in terms of A. Conclusion: we can define a linear operator $A : V \to W$ by its action on a basis of V.

   Solution: Ae_i is the m-vector that is equal to the ith column of A.

4. Let V be the space of polynomials in two variables of the form $f(x, y) = a + bx + cy + dx^2$, and let W be the space of degree one polynomials in two variables.

   a) Find (the simplest) bases of V and W. What are the dimensions of these spaces?

   b) Consider a linear operator $A = \frac{d}{dx}$ from V to W. Write A as a matrix in the bases that we found in part (a).

c) What are the nullspace and column space of A? What are the kernel and image of $\frac{d}{dx}$? What is the conclusion?

d) What is the rank of A?

e) *Bonus.* Let us add twice the second column of A to the first, and denote the new matrix (linear transformation) by A'. How did the transformation change?

f) *Bonus.* Write A' as a composition of A and some other linear transformation M. What are the vector spaces that M operates between?

Hint: recall column operations and how they are related to matrix multiplication on the right.

g) *Added during recitation.* Compute the projection matrix on the image of $\frac{d}{dx}$ in W.

Solution:

a) $V = \text{Span}\left(e_1 = 1, e_2 = x, e_3 = y, e_4 = x^2\right)$; $W = \text{Span}\,(1, x, y)$. $\dim V = 4$; $\dim W = 3$.

b) First use problem 3 to compute columns of A:

- The first column of A is $Ae_1 = \frac{d}{dx}1 = 0 = 0 \cdot 1 + 0x + 0y = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$;
- The second column of A is $Ae_2 = \frac{d}{dx}x = 1 = 1 \cdot 1 + 0x + 0y = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$;
- The third column of A is $Ae_3 = \frac{d}{dx}y = 0 = 0 \cdot 1 + 0x + 0y = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$;
- The fourth column of A is $Ae_4 = \frac{d}{dx}x^2 = 2x = 0 \cdot 1 + 2 \cdot x + 0y = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}$.

Since we now know all the columns of A, we can write the matrix: $A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}$.

c) $\text{Ker}\,\frac{d}{dx} = \text{nul}\,A = \text{Span}\,(e_1, e_3)$; $\text{Im}\,\frac{d}{dx} = \text{col}\,A = \text{Span}\,(1, x)$. Conclusion: kernel is a coordinate-independent (read: fancy) word for the familiar nullspace, and image is a coordinate-independent incarnation of column space.

d) $A = \dim \text{col}\,A = \dim \text{Im}\,A = 2$.

e)

f)

g) $\text{Im}\,\frac{d}{dx} = \text{Span}\,(1, x) = \text{col}\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$, and the matrix that appeared is tall skinny orthogonal. Denote the matrix by Q. Then, using that Q is tall skinny orthogonal, the projection matrix is equal to $QQ^T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

5. Fix a linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$ of rank r. Describe the relations between the four fundamental subspaces in terms of kernel and image. *Tricky question:* Would you be able to do that if we said that $A : V \to W$ with the same rank and dimensions of the spaces?

Solution:

- $\text{col}\,A = \text{Im}\,A$;

- $\operatorname{nul} A = \operatorname{Ker} A$;
- $\operatorname{row} A^T = \operatorname{Im} A^T$;
- $\operatorname{nul} A^T = \operatorname{Im} A^T$.

We cannot speak of row space and left nullspace of a general linear operator as of subspaces in V and W, respectively, because we cannot define a transpose of a linear transformation. We only know how to transpose matrices, not linear operators.

6. Fix a linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$. Understand that if $b \in \mathbb{R}^m$ is in the image of A, then the system $Ax = b$ has a solution, say x_0. In this case, show in addition that the space of all solutions is $x_0 + \operatorname{Ker} A$. Now conclude that in the case of nonzero kernel (nullspace), the system $Ax = b$ has either infinitely many solutions or no solutions at all, depending on whether $b \in \operatorname{Im} A$ or not.

Recitation 6. April 2

Focus: matrix calculus, determinants.

Three axioms of determinants:

(i) Normalization: $\det I = 1$.

(ii) Sign reversal (antisymmetry): when we exchange any two rows, the determinant flips sign.

(iii) The determinant is a linear function of any row, if all other rows are fixed.

Computing Determinant: Suppose A is a square matrix.

(i) Check if the $(1,1)$ entry is nonzero. If not, flip two rows so that this is the case. If this is impossible, then the first column is 0 and $\det(A) = 0$.

(ii) Using row operations, add the first row to the other rows until all the entries below the $(1,1)$ entry in the first column are 0.

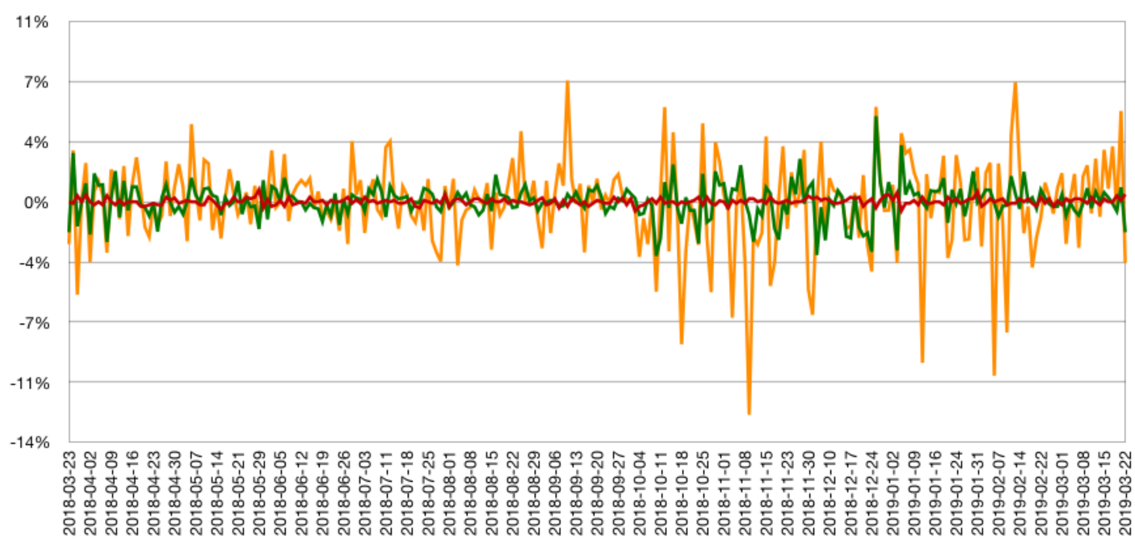
(iii) Repeat from the beginning ignoring the first column and row. If this iterates until the matrix is upper triangular, then

$$\det(A) = (-1)^{\text{number of row flips}} \times \text{product of diagonal entries of the upper triangular matrix.}$$

Note. As a consequence, we have that the same axioms are satisfied by columns as well. Don't forget that determinants are defined only for square matrices.

1. *Portfolio optimization.* In this problem, we will consider three financial instruments: Dow Jones Industrial Average index (first coordinate), Activision Blizzard equities (second coordinate) and Fidelity US Bond Index fund (third coordinate) – and will try to figure out what would be the optimal investment in those. Note that this cannot be used to devise your investment strategy, because the expectation of future returns is very hard to estimate. For the purpose of this problem, we will assume that expected returns are proportional to standard deviation of these instruments.

   a) We won't define what standard deviation means, but we should intuitively understand it as a measure of how far values tend to spread from the mean. Study the following plots of daily percentage gains/losses of the three instruments. (So value 2 on day D means that the price of the instrument rose by 2% compared to day $D-1$.)



   Knowing that standard deviation of gains of Dow Jones is $\sigma_1 = 1$, of Blizzard stock – $\sigma_2 = 2.6$, of US bonds – $\sigma_3 = 0.2$, make a guess which plot corresponds to which instrument.

   b) We write how much we invest in each instrument as a vector $w = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$, so here we invest k\$$w_1$ in Dow Jones, k\$$w_2$ in Activision Blizzard and k\$$w_3$ in US bonds. We call this vector w the *investment portfolio*.

c) Covariance is a measure of the joint variability of two random variables. The (approximate) covariance matrix C for the given instruments, and its inverse are the following:

$$C = \begin{pmatrix} 1 & 1.2 & 0 \\ 1.2 & 6.8 & -0.1 \\ 0 & -0.1 & 0.04 \end{pmatrix}, \quad C^{-1} = \begin{pmatrix} 1.3 & -0.23 & -0.6 \\ -0.23 & 0.2 & 0.5 \\ -0.6 & 0.5 & 26.2 \end{pmatrix}.$$

The importance of the covariance matrix is that we can compute variance of a linear combination of stocks (portfolio) w easily via the formula $w^T C w$. Note that variance is the square of standard deviation.
Compute the variance of the investment of k\$1 in Blizzard and k\$1 in US bonds. Explain why the diagonal entries of C are exactly the squares of standard deviations: $C_{ii} = \sigma_i^2$.

d) Modern theory of portfolio optimization says that we want to maximize "Sharpe ratio" – the ratio of expected return to the standard deviation of the portfolio. Assume that the expected return is given by vector $\mu = \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \end{pmatrix}$.
Verify that the Sharpe ratio of the portfolio w is given by the following formula:

$$\mathrm{SR} = \frac{\mu^T w}{\sqrt{w^T C w}}.$$

e) If $f(w)$ is a scalar function of w, i.e. its values are 1×1 matrices – real numbers, compute $\mathrm{d}\left(f(w)^{-1}\right)$ using the product formula for $0 = \mathrm{d}1 = \mathrm{d}\left(f(w)f(w)^{-1}\right)$. Differentials are taken with respect to w.

f) Compute the differential $\mathrm{d}(\mathrm{SR})$ with respect to w.

g) Write the condition for finding a local extremum of SR.

h) Conclude that w should be parallel to $C^{-1}\mu$, and argue that scaling w does not affect Sharpe ratio, therefore we can take $w = C^{-1}\mu$.

i) Use the above formula and the given values of μ and C^{-1} to obtain the weights of the optimal portfolio.

2. Using row operations, show that $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$.

3. *(Parts of problems 5.1.14-15 from Strang.)* Use row operations to simplify and compute:

a) $\begin{vmatrix} 1 & 2 & 3 & 0 \\ 2 & 6 & 6 & 1 \\ -1 & 0 & 0 & 3 \\ 0 & 2 & 0 & 7 \end{vmatrix} =$

b) $\begin{vmatrix} 1 & t & t^2 \\ t & 1 & t \\ t^2 & t & 1 \end{vmatrix} =$

4. Prove that $\det\left(A^{-1}\right) = (\det A)^{-1}$.

Recitation 6. Solution

Focus: matrix calculus, determinants.

Three axioms of determinants:

(i) Normalization: $\det I = 1$.

(ii) Sign reversal (antisymmetry): when we exchange any two rows, the determinant flips sign.

(iii) The determinant is a linear function of any row, if all other rows are fixed.

Computing Determinant: Suppose A is a square matrix.

(i) Check if the $(1,1)$ entry is nonzero. If not, flip two rows so that this is the case. If this is impossible, then the first column is 0 and $\det(A) = 0$.

(ii) Using row operations, add the first row to the other rows until all the entries below the $(1,1)$ entry in the first column are 0.

(iii) Repeat from the beginning ignoring the first column and row. If this iterates until the matrix is upper triangular, then

$$\det(A) = (-1)^{\text{number of row flips}} \times \text{product of diagonal entries of the upper triangular matrix.}$$

Note. As a consequence, we have that the same axioms are satisfied by columns as well. Don't forget that determinants are defined only for square matrices.

1. *Portfolio optimization.* In this problem, we will consider three financial instruments: Dow Jones Industrial Average index (first coordinate), Activision Blizzard equities (second coordinate) and Fidelity US Bond Index fund (third coordinate) – and will try to figure out what would be the optimal investment in those. Note that this cannot be used to devise your investment strategy, because the expectation of future returns is very hard to estimate. For the purpose of this problem, we will assume that expected returns are proportional to standard deviation of these instruments.

   a) Knowing that standard deviation of gains of Dow Jones is $\sigma_1 = 1$, of Blizzard stock – $\sigma_2 = 2.6$, of US bonds – $\sigma_3 = 0.2$, make a guess which plot corresponds to which instrument.

   b) We write how much we invest in each instrument as a vector $w = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$, so here we invest k\$$w_2$ in Activision Blizzard. Note that there is a way to "invest" a negative amount of money in a certain instrument. Interpret this vector w as an investment into a linear combination of the three instruments. We call this vector w the *investment portfolio.*

   c) Covariance is a measure of the joint variability of two random variables. The (approximate) covariance matrix C for the given instruments, and its inverse are the following:

   $$C = \begin{pmatrix} 1 & 1.2 & 0 \\ 1.2 & 6.8 & -0.1 \\ 0 & -0.1 & 0.04 \end{pmatrix}, \ C^{-1} = \begin{pmatrix} 1.3 & -0.23 & -0.6 \\ -0.23 & 0.2 & 0.5 \\ -0.6 & 0.5 & 26.2 \end{pmatrix}.$$

   The importance of the covariance matrix is that we can compute variance of a linear combination of stocks (portfolio) w easily via the formula $w^T C w$. Note that variance is the square of standard deviation.

   Compute the variance of the investment of k\$1 in Blizzard and k\$1 in US bonds. Explain why the diagonal entries of C are exactly the squares of standard deviations: $C_{ii} = \sigma_i^2$.

   d) Modern theory of portfolio optimization says that we want to maximize "Sharpe ratio" – the ratio of expected return to the standard deviation of the portfolio. Assume that the expected return is given by vector $\mu = \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \end{pmatrix}$.

   Verify that the Sharpe ratio of the portfolio w is given by the following formula:

   $$\text{SR} = \frac{\mu^T w}{\sqrt{w^T C w}}.$$

   e) If $f(w)$ is a scalar function of w, i.e. its values are 1×1 matrices – real numbers, compute $\mathrm{d}\left(f(w)^{-1}\right)$ using the product formula for $0 = \mathrm{d}1 = \mathrm{d}\left(f(w)f(w)^{-1}\right)$. Differentials are taken with respect to w.

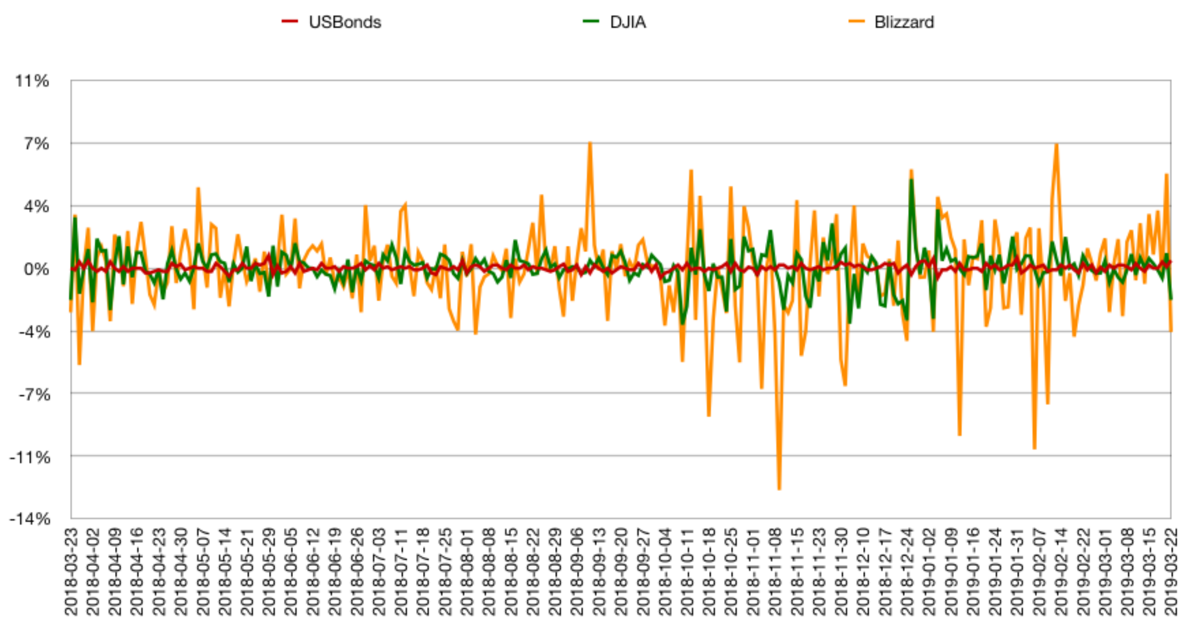
f) Compute the differential d(SR) with respect to w.

g) Write the condition for finding a local extremum of SR.

h) Conclude that w should be parallel to $C^{-1}\mu$, and argue that scaling w does not affect Sharpe ratio, therefore we can take $w = C^{-1}\mu$.

i) Use the above formula and the given values of μ and C^{-1} to obtain the weights of the optimal portfolio.

Solution:



a)

b)

c) For the first question, our portfolio is $w = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$, so variance is $w^T C w = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1.2 & 0 \\ 1.2 & 6.8 & -0.1 \\ 0 & -0.1 & 0.04 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} =$ $\begin{pmatrix} 1.2 & 6.7 & -0.06 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = 6.64.$

d)

e)

f) The computation runs as follows:

$$\begin{aligned} \mathrm{d}(\mathrm{SR}) = \mathrm{d}\frac{\mu^T w}{\sqrt{w^T C w}} &= \frac{\mathrm{d}\left(\mu^T w\right)\sqrt{w^T C w} - \mu^T w \cdot \mathrm{d}\left(\sqrt{w^T C w}\right)}{w^T C w} = \\ &= \frac{\mu^T \mathrm{d}w \cdot w^T C w - \mu^T w \cdot \frac{1}{2}\left(\mathrm{d}w^T C w + w^T C \mathrm{d}w\right)}{(w^T C w)^{\frac{3}{2}}} = \frac{\mu^T \mathrm{d}w \cdot w^T C w - \mu^T w \cdot \frac{1}{2}\left(w^T C \mathrm{d}w + w^T C \mathrm{d}w\right)}{(w^T C w)^{\frac{3}{2}}} = \\ &= \frac{\mu^T \mathrm{d}w \cdot w^T C w - \mu^T w \cdot w^T C \mathrm{d}w}{(w^T C w)^{\frac{3}{2}}}. \end{aligned}$$

g) The condition is $\mu^T w \cdot w^T C = w^T C w \cdot \mu^T$. We can transpose both sides (noting that $w^T C w$ and $\mu^T w$ are scalars) and get $\mu^T w \cdot C w = w^T C w \cdot \mu$. After multiplying both sides by C^{-1} we get $\mu^T w \cdot w = w^T C w \cdot C^{-1}\mu$.

h) Note again that $w^T C w$ and $\mu^T w$ are scalars, so from the previous part we get that w and $C^{-1}\mu$ are parallel.

i) $w = C^{-1}\mu = \begin{pmatrix} 1.3 & -0.23 & -0.6 \\ -0.23 & 0.2 & 0.5 \\ -0.6 & 0.5 & 26.2 \end{pmatrix} \begin{pmatrix} 1 \\ 2.6 \\ 0.2 \end{pmatrix} = \begin{pmatrix} 0.582 \\ 0.39 \\ 5.94 \end{pmatrix}$

2. Using row operations, show that $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$.

3. *(Part of problem 5.1.15 from Strang.)* Use row operations to simplify and compute:
$$\begin{vmatrix} 1 & t & t^2 \\ t & 1 & t \\ t^2 & t & 1 \end{vmatrix}$$

4. Prove that $\det\left(A^{-1}\right) = (\det A)^{-1}$.

Recitation 8. April 22

Focus: eigenvectors, eigenvalues and eigendecomposition.

Notation. For the rest of this worksheet, let A be an $n \times n$ matrix operating on an n-dimensional vector space V, so $A : V \to V$.

Definition. A nonzero vector $v \in V$ is called an *eigenvector* for the matrix A if for some real or complex scalar λ we have $Av = \lambda v$.

Definition. The value λ is then called the *eigenvalue* corresponding to this eigenvector v.

Remark. Since for the eigenvector v we have $(A - \lambda)v = 0$, the matrix $A - \lambda I$ is not invertible, and so an eigenvalue is necessarily a root of the polynomial $\chi_A(\lambda) = \det(A - \lambda I)$.

Definition. If all roots of $\chi_A(\lambda)$ are different, then A is *diagonalizable*, which means that we can write $A = X\Lambda X^{-1}$ for some diagonal matrix Λ and invertible matrix X. This representation of A as $X\Lambda X^{-1}$ is called *eigendecomposition.*

1. Suppose we have $B = XAX^{-1}$.
   a) Prove that $\chi_B(\lambda) = \chi_A(\lambda)$.
   b) How are eigenvalues of B related to those of A?
   c) How are eigenvectors of B related to those of A?
   d) Suppose that one of the eigenvalues of A is zero. Does it mean that A is singular? Does it mean that B is singular?

Solution:

2. Give an example of a diagonalizable matrix with a pair of equal eigenvalues.

Solution:

3. Prove that if V is odd-dimensional, then $A : V \to V$ has at least one real eigenvalue.

Solution:

4. *Closed formula for Fibonacci numbers.* Let F_i denote the ith element in the Fibonacii sequence, defined by setting $F_0 = 0$, $F_1 = 1$ and $F_{i+2} = F_{i+1} + F_i$ for all natural values of i (including zero).

a) Find a matrix A such that $A \begin{pmatrix} F_{i+1} \\ F_i \end{pmatrix} = \begin{pmatrix} F_{i+2} \\ F_{i+1} \end{pmatrix}$.

b) Find the eigenvalues of A. Let φ denote the largest eigenvalue.

c) Find the eigenvectors of A.

d) Compute A^{50} up to nine decimal points. You can only use simple calculators (e.g. Google engine), no matrix calculators are needed.

e) Using the result of part (c), explain why $\frac{F_{51}}{F_{50}}$ is very close to φ.

Recitation 8. Solution

Focus: eigenvectors, eigenvalues and eigendecomposition.

Notation. For the rest of this worksheet, let A be an $n \times n$ matrix operating on an n-dimensional vector space V, so $A : V \to V$.

Definition. A nonzero vector $v \in V$ is called an *eigenvector* for the matrix A if for some real or complex scalar λ we have $Av = \lambda v$.

Definition. The value λ is then called the *eigenvalue* corresponding to this eigenvector v.

Remark. Since for the eigenvector v we have $(A - \lambda)v = 0$, the matrix $A - \lambda I$ is not invertible, and so an eigenvalue is necessarily a root of the polynomial $\chi_A(\lambda) = \det(A - \lambda I)$.

Definition. If all roots of $\chi_A(\lambda)$ are different, then A is *diagonalizable*, which means that we can write $A = X\Lambda X^{-1}$ for some diagonal matrix Λ and invertible matrix X. This representation of A as $X\Lambda X^{-1}$ is called *eigendecomposition*.

1. Suppose we have $B = XAX^{-1}$.
   a) Prove that $\chi_B(\lambda) = \chi_A(\lambda)$.
   b) How are eigenvalues of B related to those of A?
   c) How are eigenvectors of B related to those of A?
   d) Suppose that one of the eigenvalues of A is zero. Does it mean that A is singular? Does it mean that B is singular?

Solution:

a) $\chi_B(\lambda) = \det(B - \lambda I) = \det(XAX^{-1} - \lambda XX^{-1}) = \det\left(X(A - \lambda I)X^{-1}\right) = \det X \cdot \det(A - \lambda I) \cdot \det X^{-1} = \chi_A(\lambda)$.

b) Since eigenvalues correspond to roots of the characteristic polynomial, and those are equal for A and B, as follows from part (a), we can conclude that eigenvalues of B coincide with those of A, counted with multiplicities.

c) If v is and eigevector of A with eigenvalue λ_0, then Xv is an eigenvector of B with eigenvalue λ_0: $B(Xv) = XAX^{-1}Xv = XAv = \lambda_0 Xv$.

d) Yes, because then $\det(A - 0 \cdot I) = \det A = 0$, so A is singular. And since B is a matrix similar to A, it is also singular.

2. Give an example of a diagonalizable matrix with a pair of equal eigenvalues.

Solution: For example, $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

3. Prove that if V is odd-dimensional, then $A : V \to V$ has at least one real eigenvalue.

Solution: First way. All complex eigenvalues come in pairs, so if the dimension is odd, then one eigenvalue will have to be real.

Second way. Since V is odd-dimensional, the degree of $\chi_A(\lambda)$ is odd. Also note that the leading coefficient is -1. So $\chi_A(\lambda)$, considered as a function of single variable λ, is positive when λ is large negative and negative when λ is sufficiently big positive. Therefore, it must have a zero.

4. *Closed formula for Fibonacci numbers.* Let F_i denote the ith element in the Fibonacii sequence, defined by setting $F_0 = 0$, $F_1 = 1$ and $F_{i+2} = F_{i+1} + F_i$ for all natural values of i (including zero).

a) Find a matrix A such that $A\begin{pmatrix} F_{i+1} \\ F_i \end{pmatrix} = \begin{pmatrix} F_{i+2} \\ F_{i+1} \end{pmatrix}$.

Solution. $A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$.

b) Find the eigenvalues of A. Let φ denote the largest eigenvalue.

Solution. First compute the characteristic polynomial: $\chi_A(\lambda) = \det\begin{pmatrix} 1-\lambda & 1 \\ 1 & -\lambda \end{pmatrix} = \lambda^2 - \lambda - 1$.

Now compute the discriminant $D = 1 + 4 = 5$.

Then the eigenvalues are $\varphi = \frac{1+\sqrt{5}}{2} \approx 1.618$ and $\overline{\varphi} = \frac{1-\sqrt{5}}{2} \approx -0.618$.

Note that they are related as follows: $\varphi + \overline{\varphi} = 1$, $\varphi\overline{\varphi} = -1$ and $\varphi - \overline{\varphi} = \sqrt{5}$

c) Find the eigenvectors of A and the eigendecomposition $A = X\Lambda X^{-1}$.

Solution. Since there are two distinct eigenvalues, each of the matrices $A - \varphi I$ and $A - \overline{\varphi}I$ has exactly one-dimensional kernel (nullspace).

First find eigenvector v_1 for eigenvalue φ. It should satisfy $\begin{pmatrix} 1-\varphi & 1 \\ 1 & -\varphi \end{pmatrix} v_1 = 0$. Since we know that the matrix is of rank one, we can look for a vector from the nullspace of the second row, and we see that $v_1 = \begin{pmatrix} \varphi \\ 1 \end{pmatrix}$.

Similarly, the vector $v_2 = \begin{pmatrix} \overline{\varphi} \\ 1 \end{pmatrix}$ is an eigenvector of A with eigenvalue $\overline{\varphi}$.

For the eigendecomposition, we know that we can write $X = \begin{pmatrix} v_1 & v_2 \end{pmatrix}$, then:

$$A = \begin{pmatrix} \varphi & \overline{\varphi} \\ 1 & 1 \end{pmatrix}\begin{pmatrix} \varphi & 0 \\ 0 & \overline{\varphi} \end{pmatrix}\begin{pmatrix} \varphi & \overline{\varphi} \\ 1 & 1 \end{pmatrix}^{-1} = \frac{1}{\varphi - \overline{\varphi}}\begin{pmatrix} \varphi & \overline{\varphi} \\ 1 & 1 \end{pmatrix}\begin{pmatrix} \varphi & 0 \\ 0 & \overline{\varphi} \end{pmatrix}\begin{pmatrix} 1 & -\overline{\varphi} \\ -1 & \varphi \end{pmatrix} =$$
$$= \frac{1}{\sqrt{5}}\begin{pmatrix} \varphi & \overline{\varphi} \\ 1 & 1 \end{pmatrix}\begin{pmatrix} \varphi & 0 \\ 0 & \overline{\varphi} \end{pmatrix}\begin{pmatrix} 1 & -\overline{\varphi} \\ -1 & \varphi \end{pmatrix}.$$

d) Compute A^{50} up to 9 decimal points. You can only use simple calculators (e.g. Google engine), no matrix calculators are needed.

Solution.

$$A^{50} = \left(X\Lambda X^{-1}\right)^{50} = X\Lambda^{50}X^{-1} = \frac{1}{\sqrt{5}}\begin{pmatrix} \varphi & \overline{\varphi} \\ 1 & 1 \end{pmatrix}\begin{pmatrix} \varphi & 0 \\ 0 & \overline{\varphi} \end{pmatrix}^{50}\begin{pmatrix} 1 & -\overline{\varphi} \\ -1 & \varphi \end{pmatrix} \approx$$
$$\approx \frac{1}{\sqrt{5}}\begin{pmatrix} \varphi & \overline{\varphi} \\ 1 & 1 \end{pmatrix}\begin{pmatrix} \varphi^{50} & 0 \\ 0 & 4\cdot 10^{-11} \end{pmatrix}\begin{pmatrix} 1 & -\overline{\varphi} \\ -1 & \varphi \end{pmatrix}.$$

Now note that all quantities that $4 \cdot 10^{-11}$ gets multiplied with are smaller than 10 in absolute value, so we can approximate this number with 0:

$$A^{50} \approx \frac{1}{\sqrt{5}}\begin{pmatrix} \varphi & \overline{\varphi} \\ 1 & 1 \end{pmatrix}\begin{pmatrix} \varphi^{50} & 0 \\ 0 & 4\cdot 10^{-11} \end{pmatrix}\begin{pmatrix} 1 & -\overline{\varphi} \\ -1 & \varphi \end{pmatrix} \approx \frac{1}{\sqrt{5}}\begin{pmatrix} \varphi & \overline{\varphi} \\ 1 & 1 \end{pmatrix}\begin{pmatrix} \varphi^{50} & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & -\overline{\varphi} \\ -1 & \varphi \end{pmatrix} =$$
$$= \frac{1}{\sqrt{5}}\begin{pmatrix} \varphi & \overline{\varphi} \\ 1 & 1 \end{pmatrix}\begin{pmatrix} \varphi^{50} & -\varphi^{50}\cdot\overline{\varphi} \\ 0 & 0 \end{pmatrix} = \frac{1}{\sqrt{5}}\begin{pmatrix} \varphi^{51} & -\varphi^{50}\varphi\overline{\varphi} \\ \varphi^{50} & -\varphi^{49}\varphi\overline{\varphi} \end{pmatrix} = \frac{1}{\sqrt{5}}\begin{pmatrix} \varphi^{51} & \varphi^{50} \\ \varphi^{50} & \varphi^{49} \end{pmatrix}.$$

e) Using the result of part (c), explain why $\frac{F_{51}}{F_{50}}$ is very close to φ.

Solution. We will compute the approximation of the vector $\begin{pmatrix} F_{51} \\ F_{50} \end{pmatrix}$:

$$\begin{pmatrix} F_{51} \\ F_{50} \end{pmatrix} = A^{50}\begin{pmatrix} F_1 \\ F_0 \end{pmatrix} \approx \frac{1}{\sqrt{5}}\begin{pmatrix} \varphi^{51} & \varphi^{50} \\ \varphi^{50} & \varphi^{49} \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{5}}\begin{pmatrix} \varphi^{51} \\ \varphi^{50} \end{pmatrix}.$$

Therefore, $\frac{F_{51}}{F_{50}} \approx \frac{\varphi^{51}}{\varphi^{50}} = \varphi$.

Recitation 10. May 7

Focus: positive definite matrices, Markov matrices.

Definition. A symmetric matrix S is called *positive definite* if all of its eigenvalues are positive. It is *positive semidefinite* all of its eigenvalues are nonnegative, that is we allow zeroes.

Definition. A matrix A is called a *Markov matrix* if all of its entries are nonnegative and the elements in each column sum up to one. It is called a *positive Markov matrix* if in addition we require all matrix entries to be positive.

Fact. A Markov matrix A always has an eigenvalue equal to one, because columns of the matrix $A-I$ lie in the hyperplane $x_1 + \cdots + x_n = 0$. A nonpositive Markov matrix can have more than one largest eigenvalue, take for example I.

Definition. A *steady state* of a positive Markov matrix A is the unique vector v which is an eigenvector of A with eigenvalue one and whose coordinates sum up to one. It is called "steady vector", because any positive vector x whose coordinates sum up to one converges to v as we iteratively apply A, that is $\lim_n A^n x = v$.

1. Let S be a positive definite matrix. Show that then for any nonzero vector v, we have $v^T S v > 0$.

Solution:

2. *(Strang, problem 6.5.30.)* The graph of $z = x^2 + y^2$ is a bowl opening upward, or *convex*. The graph of $z = -x^2 - y^2$ is a downward bowl, which means that it is *concave*. The graph of $z = x^2 - y^2$ is a saddle. What is a condition on a, b, c for $z = F(x, y) = ax^2 + 2bxy + cy^2$ to have a saddle point at $(0, 0)$?

Solution:

3. If A and B are two Markov matrices, then show that their product AB is Markov as well. Further, derive that then any power A^k, $k > 0$, of a Markov matrix is Markov.

Solution:

4. *Weather prediciton. (Inspired by en.wikipedia.org/wiki/Examples_of_Markov_chains.)*

Last May in Boston, there were 10 rainy days and 21 days without precipitation. There were 3 occasions where a rainy day followed a rainy day, 7 occasions where a dry day followed a rainy day. After a dry day, a rainy day followed on 7 occasions and another dry day happened on 13 occasions. (Note than since there are 31 days in May, there are 30 pairs of consecutive days.)

a) With the first coordinate corresponding to a rainy day and the second – to a dry day, write the vector a_1 of probabilities for what happens after a rainy day and the vector a_2 – for after a dry day.

b) Using the results from part (a), write the Markov matrix A corresponding to this setup.

c) Find a steady vector v for A.

d) Normalize v so that the sum of its coordinates equals to 1 – this will be the steady state.

e) Compare probability of having a rainy day in May 2018 with the first coordinate in the steady state vector.

Solution:

5. If a Markov matrix A has the steady state $(1, \ldots, 1)^T$, then what can you say about the rows of this matrix?

Solution:

Recitation 10. May 7

Focus: positive definite matrices, Markov matrices.

Definition. A symmetric matrix S is called *positive definite* if all of its eigenvalues are positive. It is *positive semidefinite* all of its eigenvalues are nonnegative, that is we allow zeroes.

Definition. A matrix A is called a *Markov matrix* if all of its entries are nonnegative and the elements in each column sum up to one. It is called a *positive Markov matrix* if in addition we require all matrix entries to be positive.

Fact. A Markov matrix A always has an eigenvalue equal to one, because columns of the matrix $A-I$ lie in the hyperplane $x_1+\cdots+x_n=0$. A nonpositive Markov matrix can have more than one largest eigenvalue, take for example I.

Definition. A *steady state* of a positive Markov matrix A is the unique vector v which is an eigenvector of A with eigenvalue one and whose coordinates sum up to one. It is called "steady vector", because any positive vector x whose coordinates sum up to one converges to v as we iteratively apply A, that is $\lim_n A^n x = v$.

1. Let S be a positive definite matrix. Show that then for any nonzero vector v, we have $v^T S v > 0$.

Solution: We know that a symmetric matrix can be diagonalized: $S = Q\Sigma Q^{-1}$ with an orthogonal Q. Since S is positive definite, all diagonal elements of Σ are positive, call them $\sigma_1, \ldots, \sigma_n$. Let q_i denote the ith column of Q. Then take any vector v and consider its decomposition with respect to the basis of eigenvectors: $v = v_1 q_1 + \cdots + v_n q_n$. Now plug this into the formula, and use e_i to denote the ith standard basis vector:

$$v^T SV = \left(\sum_i v_i q_i^T\right) Q\Sigma Q^T \left(\sum_i v_i q_i\right) = \left(\sum_i v_i e_i^T\right) \Sigma \left(\sum_i v_i e_i\right) =$$
$$= \left(\sum_i v_i e_i^T\right)\left(\sum_i \sigma_i v_i e_i\right) = \sum_i v_i^2 \sigma_i > 0.$$

2. *(Strang, problem 6.5.30.)* The graph of $z = x^2+y^2$ is a bowl opening upward, or *convex*. The graph of $z = -x^2-y^2$ is a downward bowl, which means that it is *concave*. The graph of $z = x^2 - y^2$ is a saddle. What is a condition on a, b, c for $z = F(x,y) = ax^2+2bxy+cy^2$ to have a saddle point at $(0,0)$?

Solution: Note that we can rewrite the function $F(x,y)$ as follows:

$$F(x,y) = ax^2+2bxy+cy^2 = \begin{pmatrix} x & y\end{pmatrix}\begin{pmatrix} a & b \\ b & c\end{pmatrix}\begin{pmatrix} x \\ y\end{pmatrix}.$$

Now we can take the derivative of this function with respect to $w = \begin{pmatrix} x \\ y\end{pmatrix}$:

$$\mathrm{d}F(x,y) = \mathrm{d}w^T \begin{pmatrix} a & b \\ b & c\end{pmatrix} w + w^T \begin{pmatrix} a & b \\ b & c\end{pmatrix} \mathrm{d}w = 2w^T \begin{pmatrix} a & b \\ b & c\end{pmatrix}\mathrm{d}w.$$

Note that the last transition was possible because the 2×2 matrix is symmetric, and this does not work in general. So the differential can be written as the following row vector (also called a *covector*):

$$\frac{\mathrm{d}F(x,y)}{\mathrm{d}w} = 2w^T \begin{pmatrix} a & b \\ b & c\end{pmatrix}.$$

Taking the second derivative with respect to w^T will now look like this:

$$\mathrm{d}\frac{\mathrm{d}F(x,y)}{\mathrm{d}w} = 2\mathrm{d}w^T \begin{pmatrix} a & b \\ b & c\end{pmatrix}.$$

Therefore, we get the following formula: $\frac{\mathrm{d}^2F(x,y)}{\mathrm{d}w^2} = 2\begin{pmatrix} a & b \\ b & c\end{pmatrix}$.

We get a saddle point when this matrix is indefinite.

3. If A and B are two Markov matrices, then show that their product AB is Markov as well. Further, derive that then any power A^k, $k > 0$, of a Markov matrix is Markov.

Solution: Write A as a column of rows $\begin{pmatrix} a^1 \\ \vdots \\ a^n \end{pmatrix}$ and write each row in terms of its elements $a^i = \begin{pmatrix} a^i_1 & \cdots a^i_n \end{pmatrix}$. Write B as a row of its columns: $B = \begin{pmatrix} b_1 & \cdots & b_n \end{pmatrix}$ and use a similar notation b^k_j for the matrix element in row k and column j.

The element in position (i,j) in AB is $(AB)_{ij} = a^i b_j$. And now we want to check the Markov property that $\sum_i (AB)_{ij} = 1$. So let us check it. For the check, note that we can formulate the Markov property of A as follows: when

$$\sum_i (AB)_{ij} = \sum_i a^i b_j = \sum_i \sum_k a^i_k b^k_j = \sum_k \left(\sum_i a^i_k \right) b^k_j = \sum_k 1 \cdot b^k_j = 1.$$

4. *Weather prediciton. (Inspired by en.wikipedia.org/wiki/Examples_of_Markov_chains.)*

Last May in Boston, there were 10 rainy days and 21 days without precipitation. There were 3 occasions where a rainy day followed a rainy day, 7 occasions where a dry day followed a rainy day. After a dry day, a rainy day followed on 7 occasions and another dry day happened on 13 occasions. (Note than since there are 31 days in May, there are 30 pairs of consecutive days.)

a) With the first coordinate corresponding to a rainy day and the second – to a dry day, write the vector a_1 of probabilities for what happens after a rainy day and the vector a_2 – for after a dry day.

b) Using the results from part (a), write the Markov matrix A corresponding to this setup.

c) Find a steady vector v for A.

d) Normalize v so that the sum of its coordinates equals to 1 – this will be the steady state.

e) Compare probability of having a rainy day in May 2018 with the first coordinate in the steady state vector.

Solution:

a) $a_1 = \begin{pmatrix} 30\% \\ 70\% \end{pmatrix} = \begin{pmatrix} 0.3 \\ 0.7 \end{pmatrix}$; $a_2 = \begin{pmatrix} \frac{7}{20} \\ \frac{13}{20} \end{pmatrix} = \begin{pmatrix} 0.35 \\ 0.65 \end{pmatrix}$.

b) $A = \begin{pmatrix} a_1 & a_2 \end{pmatrix} = \begin{pmatrix} 0.3 & 0.35 \\ 0.7 & 0.65 \end{pmatrix}$.

c) We know that a steady vector is an eigenvector with eigenvalue one, so it lies in the kernel of the matrix $A - I = \begin{pmatrix} -0.7 & 0.35 \\ 0.7 & -0.35 \end{pmatrix}$, and we can find one vector with this property: $v = \begin{pmatrix} 0.35 & 0.7 \end{pmatrix}$.

d) $v_{st} = \frac{1}{0.35+0.7} \begin{pmatrix} 0.35 & 0.7 \end{pmatrix} = \begin{pmatrix} \frac{0.35}{1.05} \\ \frac{0.7}{1.05} \end{pmatrix} = \begin{pmatrix} 1/3 \\ 2/3 \end{pmatrix}$.

e) The probability of having a rainy day in May 2018 is $\frac{10}{31} \approx 32\%$, which is very close to the first coordinate in v_{st}, which is $\frac{1}{3} \approx 32\%$. The fact that these quantities are similar can be explained as follows: Markov matrix gives a good simple model of weather prediction. Then, whatever day we start with, applying Markov matrix 30 times to this day will give as something close to v_{st}. So in the long run, we should expect a rainy day in May with probability approximately 33%.

5. If a Markov matrix A has the steady state $(1, \ldots, 1)^T$, then what can you say about the rows of this matrix?

Solution: In each row, the sum of the entries must be equal to one.