1 Review of matrix algebra

1.1 Vectors and rules of operations

An \( p \)-dimensional vector is \( p \) numbers put together. Written as

\[
x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}.
\]

When \( p = 1 \), this represents a point in the line. When \( p = 2 \) it represents a point in a plane; when \( p = 3 \), it represents a point in the 3-d space. For general \( p \), this represents a point in a \( p \)-dimension Euclidean space, written as \( \mathbb{R}^p \). Alternatively, it can also be interpreted as an arrow, or vector, starting from the origin, pointing to that point.

The transpose operation changes a column vector to a row vector; and is written as \( x' \) or \( x^T \); that is

\[
x' = (x_1, \ldots, x_p).
\]

A number is also called a scalar. We can define the product between a vector \( x \in \mathbb{R}^p \) and a scalar, \( c \in \mathbb{R} \), as follows:

\[
cx = c \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} = \begin{pmatrix} cx_1 \\ \vdots \\ cx_p \end{pmatrix}.
\]

Draw a picture here. We define addition of two vectors, \( x \) and \( y \), as follows

\[
x + y = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_p + y_p \end{pmatrix}
\]

Draw a picture here for interpretation.
1.2 Length and angle

Suppose $p = 2$, then by Pythagoras theorem we know that the length of $x = (x_1, x_2)'$ is simply

$$\sqrt{x_1^2 + x_2^2}.$$  

We can extend this concept to arbitrary $p$. The length of a vector $x \in \mathbb{R}^p$ is the length of the arrow from origin to that point:

$$\|x\| = \sqrt{x_1^2 + \cdots + x_p^2}.$$  

Note that, for any scalar $c \in \mathbb{R}^p$,

$$\|cx\| = |c|\|x\|.$$  

Now let us define the angle between two vectors. Again start with $p = 2$. Let $\theta$ be the angle between $x$ and $y$. Let $\theta_1$ be the angle between $x$ and the horizontal axis and $\theta_2$ be the angle between $y$ and the horizontal axis. Then

$$\cos(\theta_1) = \frac{x_1}{\|x\|}, \quad \cos(\theta_2) = \frac{y_1}{\|x\|}, \quad \sin(\theta_1) = \frac{x_2}{\|x\|}, \quad \sin(\theta_2) = \frac{y_2}{\|x\|}.$$  

Now we can express the cosine of the angle $\theta$ in terms of $x$ and $y$, as follows:

$$\cos(\theta_2 - \theta_1) = \cos(\theta_2) \cos(\theta_1) + \sin(\theta_2) \sin(\theta_1) = \frac{x_1 y_1 + x_2 y_2}{\|x\|\|y\|}.$$  

This formula is generalized to the $p$-dimensional case as

$$\cos(\theta) = \frac{x_1 y_1 + \cdots + x_p y_p}{\|x\|\|y\|}.$$  

We call $x_1 y_1 + \cdots + x_p y_p$ the inner product, or scalar product, between $x$ and $y$, and write it as $x^T y$; that is,

$$x^T y = x_1 y_1 + \cdots + x_p y_p.$$  

In this notation,

$$\cos(\theta) = \frac{x^T y}{\|x\|\|y\|}.$$
The number $\theta$:

$$\cos^{-1}\left(\frac{x^T y}{\|x\|\|y\|}\right)$$

is defined as the angle between two vectors $x$ and $y$ in $\mathbb{R}^p$.

**Example:** Let

$$x = (1, 3, 2)', \quad y = (-2, 1, -1)'$$

1. Find $3x$,
2. Find $x + y$,
3. Find the lengths of $x$ and $y$,
4. Find the cosine of the angle between $x$ and $y$,
5. Find the angle between $x$ and $y$.
6. Check that $\|3x\| = 3\|x\|$. 

**Solution:**

1. $3x = 3(1, 3, 2)' = (3 \times 1, 3 \times 3, 3 \times 2)' = (3, 9, 6)'$.
2. $x + y = (1, 3, 2)' + (-2, 1, -1)' = (1 - 2, 3 + 1, 2 - 1)' = (-1, 4, 1)'$.
3. 

$$\|x\| = \sqrt{1^2 + 3^2 + 2^2} = \sqrt{14} = 3.742,$$

$$\|y\| = \sqrt{(-2)^2 + 1^2 + (-1)^2} = \sqrt{6} = 2.449.$$ 

4. Let $\theta$ be the angle between $x$ and $y$. Then

$$\cos(\theta) = \frac{(1)(-2) + (3)(1) + (2)(-1)}{\sqrt{14}\sqrt{6}} = -0.109$$

5. Now $\theta = \cos^{-1}(-0.109) = 1.680 = 96.26^\circ$.

6. 

$$3\|x\| = 3\sqrt{14}, \quad \|3x\| = \sqrt{3^2(14)}.$$

Obviously, they are the same. $\Box$
1.3 Projection of a vector onto the direction of another vector

Suppose we have two vectors, \( \mathbf{x}, \mathbf{y} \in \mathbb{R}^p \). The vector \( \mathbf{x}'(\mathbf{y}'\mathbf{y})^{-1}\mathbf{y} \) is called the projection of \( \mathbf{x} \) onto \( \mathbf{y} \). We write this as \( P_y(\mathbf{x}) \). Note that this vector is proportional to \( \mathbf{y} \). Then length of this projection is

\[
\|P_y(\mathbf{x})\| = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{y}\|^2} \mathbf{y} = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} \|\mathbf{x}\| = \|\mathbf{x}\| \cos(\theta).
\]

1.4 Linear independence

A set of vectors \( \mathbf{x}_1, \ldots, \mathbf{x}_k \) are said to be linearly dependent, if there exist scalars \( c_1, \ldots, c_k \), not all zero, such that

\[
c_1\mathbf{x}_1 + \cdots + c_k\mathbf{x}_k = 0.
\]

Interpretation: if \( k = 2 \), then this means two vectors are proportional to each other; that is, they are on the same line. When \( k = 3 \), this means the \( k \) points are in the same plane. The set of vectors \( \{\mathbf{x}_1, \ldots, \mathbf{x}_k\} \) are linearly dependent if they are not linear independent. That is, if there does not exist \( c_1, \ldots, c_k \), not all zero, such that the above holds. In other words, whenever the above holds, \( c_1 = \cdots = c_k = 0 \). Interpretation, they do not reside in any lower dimensional hyperplane.

Example: Are the following groups of vectors linearly dependent?

1. \( \mathbf{x}_1 = (1, 2, 3)', \quad \mathbf{x}_2 = (1, 1, 1)', \quad \mathbf{x}_3 = (0, 0, 1)' \)

2. \( \mathbf{x}_1 = (1, 2, 3)', \quad \mathbf{x}_2 = (1, 1, 1)', \quad \mathbf{x}_3 = (1, 3, 5)' \)

3. Illustrate your conclusions by vectors in a 3-dimensional space.

Solution:

1. Suppose

\[
c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + c_3\mathbf{x}_3 = 0.
\]

Then

\[
c_1 \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + c_2 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + c_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = 0
\]
This is equivalent to

\begin{align*}
  c_1 + c_2 &= 0 \quad (2) \\
  2c_1 + c_2 &= 0 \quad (3) \\
  3c_1 + c_2 + c_3 &= 0 \quad (4)
\end{align*}

From equation (2) we see that \( c_2 = -c_1 \). Substitute this into equation (3) to obtain \( 2c_1 - c_1 = c_1 = 0 \). Substitute this into (2) to show that \( c_2 = 0 \). Substitute \( c_1 = c_2 = 0 \) into (4) to show that \( c_3 = 0 \). Thus we have shown that whenever (1) holds we have that \( c_1 = c_2 = c_3 = 0 \). That is, \( x_1, x_2, x_3 \) are linearly independent.

2. Similarly, equation (1) implies that

\begin{align*}
  c_1 + 2c_2 + 3c_3 &= 0 \quad (5) \\
  c_1 + c_2 + c_3 &= 0 \quad (6) \\
  c_2 + 2c_3 &= 0 \quad (7)
\end{align*}

From (7) we see that \( c_2 = -2c_3 \). Substitute this into (6) and (5) to obtain

\begin{align*}
  c_1 - c_3 &= 0 \\
  c_1 - c_3 &= 0
\end{align*}

Thus equation (1) will be satisfied if \( c_1 = c_3 \) and \( c_2 = -2c_3 \). For example, if \((c_1, c_2, c_3) = (1, -2, 1)\) the equation (1) is satisfied. Hence \( x_1, x_2, x_3 \) are linearly dependent.

3. Draw two pictures to illustrate whether the three points are on the lower dimensional space or not.
2 Matrices

2.1 Notations

A matrix is a collection of numbers ordered by rows and columns. It is customary to enclose the elements of a matrix in parentheses, brackets, or braces. For example, the following is a matrix:

$$X = \begin{pmatrix} 5 & 8 & 2 \\ -1 & 0 & 7 \end{pmatrix}.$$ 

This matrix has two rows and three columns, so it is referred to as a “2 by 3” matrix. The elements of a matrix are numbered in the following way:

$$X = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{pmatrix}.$$

The first subscript in a matrix refers to the row and the second subscript refers to the column. $x_{ij}$ denotes the element on the $i$th row and $j$th column of matrix $X$.

A square matrix has as many rows as it has columns.

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

A symmetric matrix is a square matrix in which $x_{ij} = x_{ji}$ for all possible $i$ and $j$.

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

A diagonal matrix is a square matrix where all the off diagonal elements are 0. A diagonal matrix is always symmetric.

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

An identity matrix is a diagonal matrix with 1s and only 1s on the diagonal. The $p \times p$ identity matrix is almost always denoted as $I_p$.

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$
2.2 Basic operations

Matrix addition: to add two matrices, they both must have the same number of rows and they both must have the same number of columns. The elements of the two matrices are simply added together, element by element, to produce the results. Let $A = \{a_{ij}\}$ and $B = \{b_{ij}\}$ be two $m \times n$ matrices, then

$$(A + B)_{ij} = a_{ij} + b_{ij} \text{ for } 1 \leq i \leq m, 1 \leq j \leq n.$$ 

$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 2 & 3 \\ 0 & 2 & 0 \end{pmatrix}.$$

Matrix subtraction works in the same way.

Scalar multiplication: each element in the product matrix is simply the scalar multiplied by the element in the matrix. Let $c$ be a scalar and $A$ be a $m \times n$ matrix, then

$$(cA)_{ij} = ca_{ij} \text{ for } 1 \leq i \leq m, 1 \leq j \leq n.$$ 

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

Scalar multiplication is commutative, or $cA = Ac$.

Multiplication of a row vector and a column vector: the row vector must have as many columns as the column vector has rows. The product is a scalar. Let $b = (b_1 \cdots b_p), \ a = \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix},$ then

$\mathbf{b} \cdot \mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_p \end{pmatrix} = \sum_{i=1}^{p} b_i a_i.$

Multiplication of two matrices: multiplication of two matrices is defined only if the number of columns of the left matrix is the same as the number of rows of the right matrix. Let $A$ be $m \times p$ and $B$ be $p \times n$, then

$$(AB)_{ij} = \sum_{l=1}^{p} a_{il}b_{lj}.$$
Let \( \mathbf{a}_i \) be the \( i \)th row vector of \( \mathbf{A} \). Let \( \mathbf{b}_j \) be the \( j \)th column vector of \( \mathbf{B} \). Then

\[
(\mathbf{AB})_{ij} = \mathbf{a}_i \cdot \mathbf{b}_j \\
(\mathbf{AB})_i = \mathbf{a}_i \cdot \mathbf{B} \\
(\mathbf{AB})_j = \mathbf{A} \cdot \mathbf{b}_j
\]

Some properties:

1. \( \mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A} \) (commutativity).
2. \( (c\mathbf{A})\mathbf{B} = \mathbf{A}(c\mathbf{B}) \) (scalar associativity).
3. \( c(\mathbf{A} + \mathbf{B}) = c\mathbf{A} + c\mathbf{B} \) (scalar distributivity).
4. \( (\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}) \) (matrix associativity).
5. \( (\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{A}\mathbf{C} + \mathbf{B}\mathbf{C} \) (matrix left distributivity).
6. \( \mathbf{C}(\mathbf{A} + \mathbf{B}) = \mathbf{CA} + \mathbf{CB} \) (matrix right distributivity).
7. \( \mathbf{AI} = \mathbf{A}, \mathbf{IB} = \mathbf{B} \).

Generally, \( \mathbf{AB} \neq \mathbf{BA} \), or the commutative law doesn’t apply to matrix multiplication. That is, if \( \mathbf{A} \) is \( k \times k \) matrix and \( \mathbf{B} \) is a \( k \times k \) matrix. Then both \( \mathbf{AB} \) and \( \mathbf{BA} \) are defined. But it may not be true that \( \mathbf{AB} = \mathbf{BA} \). If this does hold then we say \( \mathbf{A} \) and \( \mathbf{B} \) commute.

**Matrix transpose:** the transpose of a matrix \( \mathbf{A} \) is denoted by \( \mathbf{A}' \) or \( \mathbf{A}^T \).

The first row of a matrix becomes the first column of the transpose matrix, the second row of the matrix becomes the second column of the transpose, etc. \( \mathbf{A}'_{ij} = \mathbf{A}_{ji} \). The transpose of a row vector will be a column vector, and the transpose of a column vector will be a row vector. The transpose of a symmetric matrix is simply the original matrix.

**Matrix inverse:** the inverse of a matrix \( \mathbf{A} \) is denoted by \( \mathbf{A}^{-1} \).

\[
\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}
\]

A matrix must be square to have an inverse, but not all square matrices have an inverse. In some cases, the inverse does not exist.
We say that \( A \) is nonsingular if it is invertible. We say \( A \) is singular if it is not invertible. A \( k \times k \) matrix \( A \) is invertible if and only if its columns, \( a_1, \ldots, a_k \), are linearly independent.

**Trace of matrix:** denoted as \( tr(A) \), the trace is used only for square matrices and equals the sum of the diagonal elements of the matrix. Let \( A \) be \( p \times p \), then \( tr(A) = \sum_{i=1}^{p} a_{ii} \).

### 2.3 Determinant

The determinant of \( A \) is denoted by \( det(A) \) or \( |A| \).

**Some properties:**

1. A matrix is invertible if and only if its determinant is nonzero.

2. The determinant of a product of square matrices equals the product of their determinants: \( det(AB) = det(A)det(B) \).

3. \( det(A^{-1}) = [det(A)]^{-1} \), \( det(A') = det(A) \).

4. Adding a multiple of any row to another row, or adding a multiple of any column to another column, does not change the determinant.

5. Interchanging two rows or two columns affects the determinant by multiplying it by \(-1\).

6. The \( 2 \times 2 \) matrix \( A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \) has determinant \( det(A) = ad - bc \).

7. The determinant \( det(A) \) can be viewed as the oriented area of the parallelogram with vertices at \((0,0)\), \((a,b)\), \((a+c,b+d)\), and \((c,d)\). The oriented area is the same as the usual area, except that it is negative when the vertices are listed in clockwise order.

Let \( A \) be a \( k \times k \) matrix. Its determinant is a number defined recursively as follows. Let \( A_{-i,-j} \) be the \((k-1) \times (k-1)\) matrix constructed by removing the \( i \)th row, \( j \)th column of the matrix \( A \).

1. If \( k = 1 \); that is, if \( A \) is a number \( a \), then \( det(A) = a \);
2. For general \( k \), suppose the \((i, j)\)th element of \( A \) is \( a_{ij} \). Then

\[
\det(A) = \sum_{j=1}^{k} a_{1j} \det(A_{-1,-j})(-1)^{1+j}.
\]

In fact, 1 can be replaced by any row label \( i \), and the resulting numbers are all the same. Here \( \det(A_{-i,-j}) \) is called the minor of \( a_{ij} \), and \((-1)^{i+j} \det(A_{-i,-j})\) is called the cofactor of \( a_{ij} \), and is written as \( C_{ij} \). So the determinant of a matrix is the sum of each element of the first row of \( A \) multiplied by its cofactor. It is helpful to first see what \((-1)^{i+j}\) means. This means whenever \( i + j \) is even, take + for the cofactor, whenever \( (i + j) \) is odd, take - as for that cofactor. Thus, if we have a 2 × 2 matrix, the pattern is

\[
\begin{pmatrix}
+ & - \\
- & +
\end{pmatrix}
\]

If we have a 3 × 3 matrix, then the sign pattern is

\[
\begin{pmatrix}
+ & - & + \\
- & + & - \\
+ & - & +
\end{pmatrix}
\]

It is like a checker board with the first grid always colored white (meaning +).

**Example:** Suppose \( A = (-3) \). Then \( \det(A) = -3 \). Suppose

\[
A = \begin{pmatrix}
a_{11} & a_{12} \\
a_{21} & a_{22}
\end{pmatrix}
\]

Then

\[
C_{11} = a_{22}, \quad C_{12} = -a_{21}.
\]

So

\[
\det(A) = a_{11}a_{22} + a_{12}(-a_{21}).
\]

If \( A \) is a 3 × 3 matrix, then

\[
A = \begin{pmatrix}
a_{11} & a_{12} & a_{13} \\
a_{21} & a_{22} & a_{23} \\
a_{31} & a_{32} & a_{33}
\end{pmatrix}
\]
Then
\[ C_{11} = + \det \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix}, \quad C_{12} = - \det \begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}, \quad C_{13} = + \det \begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}. \]

and

\[
\det(A) = a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13}.
\]

More properties of determinant:

1. If \( c \) is a number and \( A \) is \( k \times k \) matrix, then \( \det(cA) = c^k \det(A) \).
2. If \( A \) is a square, diagonal matrix, then its determinant is the product of the diagonal elements.
3. The determinant of a square, symmetric matrix is the product of all eigenvalues of \( A \).
3 Matrices (Day 3)

3.1 Rank

The column rank of a matrix \( A \) is the maximum number of linearly independent column vectors of \( A \). The row rank of a matrix \( A \) is the maximum number of linearly independent row vectors of \( A \).

The column rank and the row rank are always equal, and are called the rank of matrix \( A \), denoted as \( \text{rank}(A) \).

Some Properties:
1. Only a zero matrix has rank zero.
2. For \( A \in \mathbb{R}^{m \times n} \), \( \text{rank}(A) \leq \min(m, n) \).
3. For \( A \in \mathbb{R}^{m \times n} \), \( A \) is invertible if and only if \( \text{rank}(A) = n \).
4. For \( A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times k} \), \( \text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\} \).
5. For \( A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times k} \) with \( \text{rank}(B) = n \), then \( \text{rank}(AB) = \text{rank}(A) \).
6. \( \text{rank}(A + B) \leq \text{rank}(A) + \text{rank}(B) \).
7. \( \text{rank}(A) = \text{rank}(A^T) = \text{rank}(AA^T) = \text{rank}(A^TA) \).

3.2 Orthogonal matrix

A square matrix \( Q \) is called an orthogonal matrix if \( QQ' = Q'Q = I \).

In other words, \( Q^{-1} = Q' \). In this case, if we write the columns of \( Q \) as \( q_1, \ldots, q_p \), then we have

\[
Q'Q = \begin{pmatrix}
q_1' \\
\vdots \\
q_p'
\end{pmatrix}
\begin{pmatrix}
q_1 & \cdots & q_p
\end{pmatrix} = \begin{pmatrix}
q_1q_1 & \cdots & q_1q_p \\
\vdots & \ddots & \vdots \\
q_pq_1 & \cdots & q_pq_p
\end{pmatrix} = I
\]

That is,

\[
q'_i q_j = \begin{cases} 
0 & \text{if } i \neq j \\
1 & \text{if } i = j 
\end{cases}
\]

We say that two vectors, \( a, b \) are orthogonal if \( a'b = 0 \). This is because the cosine of their angle \( a'b/\|a\|\|b\| = 0 \). Thus the columns of an orthogonal matrix are orthogonal with length 1. Similarly, the rows of an orthogonal matrix are orthogonal with length 1.
3.3 Eigenvalues and Eigenvectors

We say that a square matrix $A_{p \times p}$ has eigenvalue $\lambda$ if there is a vector $x_{p \times 1} \neq 0$ such that

$$Ax = \lambda x.$$  

The vector $x$ is called an eigenvector of $A$ associated with $\lambda$. Note that the above is unchanged if we multiply both sides by a non-zero constant. So we can multiply both sides of the equality by $1/||x||$, so that, the convention of eigenvector is that it has length 1.

To find eigenvalues of $A$ such that $Ax = \lambda x$, we set the characteristic function equal to 0, and solve

$$det(A - \lambda I) = 0.$$  

A $k \times k$ symmetric matrix $A$ has $k$ pairs of eigenvalues and eigenvectors:

$$(\lambda_1, e_1), \ldots, (\lambda_k, e_k).$$  

Furthermore, $e_1, \ldots, e_k$ can always be chosen so that

$$e'_i e_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

That is, they are orthogonal and have length 1. Such a set of vectors is called an orthonormal set of vectors.

If $(\lambda_1, e_1), \ldots, (\lambda_k, e_k)$ are the $k$ eigenpairs of the matrix $A_{k \times k}$, then the following equality holds:

$$A = \lambda_1 e_1 e'_1 + \cdots + \lambda_k e_k e'_k.$$  

This is called the spectral decomposition of $A$, or eigen decomposition of $A$.

Denote $A$ as a diagonal matrix with diagonal elements $A_{ii} = \lambda_i$. Denote $E$ with the $i$th column being $e_i$. Then we have eigenvalue decomposition of $A$ as

$$A = E \Lambda E^T.$$
3.4 Power of a matrix

Let $A$ be a $k \times k$ square, symmetric matrix, and let $(\lambda_1, e_1), \ldots, (\lambda_k, e_k)$ be the eigen pairs of $A$. Let $\alpha$ be a number. The following $k \times k$ matrix

$$\lambda_1^\alpha e_1 e_1' + \cdots + \lambda_k^\alpha e_k e_k'$$

is called the $\alpha$th power of $A$. Notice that when $0 \leq \alpha < 1$, the above is defined only when $\lambda_1, \ldots, \lambda_k$ are nonnegative. This is written as $A^\alpha$. In particular, if $\lambda_1 \geq 0, \ldots, \lambda_k \geq 0$ and $\alpha = 1/2$, then $A^{1/2}$ is called the square root of the matrix $A$. The matrix $A$ raised to the power $-1$ is the same as the inverse of a matrix.

3.5 System of equations

Consider solving for $x$ and $y$ in

$$\begin{cases} 3x - y = 7 \\ 2x + 3y = 1 \end{cases}$$

Denote

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

Then

$$A \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 7 \\ 1 \end{pmatrix}$$

Pre-multiply both sides by $A^{-1}$, we get

$$\begin{pmatrix} x \\ y \end{pmatrix} = A^{-1} \begin{pmatrix} 7 \\ 1 \end{pmatrix} = \begin{pmatrix} 3/11 & 1/11 \\ -2/11 & 3/11 \end{pmatrix} \begin{pmatrix} 7 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$