Linear Algebra-Part I

Peter Minary, office: Room 367
email: peter.minary@cs.ox.ac.uk

- Chapter 1: Vectors and vector spaces

  (Week 1, Lectures 1-3)

- Chapter 2: Independence and orthogonality

  (Week 2, Lectures 4-6)

- Chapter 3: Matrices

  (Week 3, Lectures 7-9)

- Chapter4: Systems of linear equations

  (Week 4, Lectures 10-12)

Chapter 3

Matrices

3.1. Matrix Notation and Basic Operations

Recall that a matrix is a rectangular list of elements, in our case real numbers e.g.

$$\begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{2} \\ \mathbf{0} & \mathbf{2} & \mathbf{1} \end{bmatrix},$$

this particular matrix has two rows and three columns, we say it is a two by three matrix. Alternatively, we say that

$$\begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{2} \\ \mathbf{0} & \mathbf{2} & \mathbf{1} \end{bmatrix} \in \mathbb{R}^{\mathbf{2}\times\mathbf{3}}.$$

Note, that using the elements along rows we can compose two row vectors and elements along columns can form three column vectors.

Matrix Addition

Two matrices that have m rows and n columns can be added together as follows

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & & a_{1n} \\ a_{21} & a_{22} & a_{23} & & a_{2n} \\ & & & & \\ & & & & \\ a_{m1} & a_{m2} & a_{m3} & & a_{mn} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} & & b_{1n} \\ b_{21} & b_{22} & b_{23} & & b_{2n} \\ & & & & \\ & & & & \\ b_{m1} & b_{m2} & b_{m3} & & b_{mn} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & a_{13}+b_{13} & & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & a_{23}+b_{23} & & a_{2n}+b_{2n} \\ & & & & \\ & & & & \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & a_{m3}+b_{m3} & & a_{mn}+b_{mn} \end{bmatrix}.$$

Scalar Matrix Multiplication

If $\alpha \in \mathbb{R}$ and A is an m by n matrix then αA produces the m by n matrix B where every entry of B is α multiplied by the corresponding entry of A.

$$\alpha \begin{bmatrix} a_{11} & a_{12} & a_{13} & & a_{1n} \\ a_{21} & a_{22} & a_{23} & & a_{2n} \\ & & & & \\ & & & & \\ a_{m1} & a_{m2} & a_{m3} & & a_{mn} \end{bmatrix} = \begin{bmatrix} \alpha a_{11} & \alpha a_{12} & \alpha a_{13} & & \alpha a_{1n} \\ \alpha a_{21} & \alpha a_{22} & \alpha a_{23} & & \alpha a_{2n} \\ & & & & \\ & & & & \\ \alpha a_{m1} & \alpha a_{m2} & \alpha a_{m3} & & \alpha a_{mn} \end{bmatrix}$$

Remark 3.1.1 *Given the properties of matrix addition and scalar multiplication the set of all* m *by* n *matrices* ($m, n \geq 1$ *and integer), denoted* M_{mn}*, is a vector space. The space of symmetric matrices of dimension* n *by* n *is a subspace of* M_{nn}.

Matrix Matrix Multiplication

If matrix $\boldsymbol{A}$ has $\boldsymbol{m}$ rows and $\boldsymbol{n}$ columns and matrix $\boldsymbol{B}$ has $\boldsymbol{n}$ rows and $\boldsymbol{p}$ columns then we may pre-multiply $\boldsymbol{B}$ by $\boldsymbol{A}$ and obtain the matrix $\boldsymbol{AB}$ with $\boldsymbol{m}$ rows and $\boldsymbol{p}$ columns as follows

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & & a_{1n} \\ a_{21} & a_{22} & a_{23} & & a_{2n} \\ & & & & \\ & & & & \\ a_{m1} & a_{m2} & a_{m3} & & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} & & b_{1p} \\ b_{21} & b_{22} & b_{23} & & b_{2p} \\ & & & & \\ & & & & \\ b_{n1} & b_{n2} & b_{n3} & & b_{np} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} a_{1i}b_{i1} & \sum_{i=1}^{n} a_{1i}b_{i2} & \sum_{i=1}^{n} a_{1i}b_{i3} & & \sum_{i=1}^{n} a_{1i}b_{ip} \\ \sum_{i=1}^{n} a_{2i}b_{i1} & \sum_{i=1}^{n} a_{2i}b_{i2} & \sum_{i=1}^{n} a_{2i}b_{i3} & & \sum_{i=1}^{n} a_{2i}b_{ip} \\ & & & & \\ & & & & \\ \sum_{i=1}^{n} a_{mi}b_{i1} & \sum_{i=1}^{n} a_{mi}b_{i2} & \sum_{i=1}^{n} a_{mi}b_{i3} & & \sum_{i=1}^{n} a_{mi}b_{ip} \end{bmatrix}.$$

The matrix vector multiplication can be viewed as the multiplication of an m by n matrix with an n by 1 matrix.

$$
\begin{bmatrix}
a_{11} & a_{12} & a_{13} & & a_{1n} \\
a_{21} & a_{22} & a_{23} & & a_{2n} \\
 & & & & \\
 & & & & \\
a_{m1} & a_{m2} & a_{m3} & & a_{mn}
\end{bmatrix}
\begin{bmatrix}
x_1 \\ x_2 \\ \\ \\ x_n
\end{bmatrix}
=
\begin{bmatrix}
b_1 \\ b_2 \\ \\ \\ b_m
\end{bmatrix}
=
\begin{bmatrix}
\sum_{i=1}^{n} a_{1i} x_i \\ \sum_{i=1}^{n} a_{2i} x_i \\ \\ \\ \sum_{i=1}^{n} a_{mi} x_i
\end{bmatrix}
$$

$$
A \qquad \mathbf{x} \qquad = \qquad \mathbf{b}
$$

Finding $\mathbf{x}$ given A and b is one of the classical problems of linear algebra.

For matrix addition and multiplication we have

1. Associative

$$(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC}) = \boldsymbol{ABC}.$$

2. Distributive

$$\boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C}) = \boldsymbol{AB}+\boldsymbol{AC}, \quad (\boldsymbol{A}+\boldsymbol{B})\boldsymbol{C} = \boldsymbol{AC}+\boldsymbol{BC}.$$

3. Matrix multiplication is NOT commutative.

The transpose of a matrix A, denoted A^T, is such that the $i-$th row of A is the $i-$th column of A^T.

Example 3.1.2

$$\begin{bmatrix} 1 & 2 & -1 \\ 3 & 1 & -2 \end{bmatrix}^T = \begin{bmatrix} 1 & 3 \\ 2 & 1 \\ -1 & -2 \end{bmatrix}.$$

Clearly, $(A^T)^T = A$.

Exercise 3.1.3 *Let A and B n by n matrices. Show that the transpose of AB is $B^T A^T$.*

A matrix is symmetric if $\boldsymbol{A} = \boldsymbol{A}^T$.

Example 3.1.4 *Suppose $\boldsymbol{R}$ is an $\boldsymbol{m}$ by $\boldsymbol{n}$ matrix. Then $\boldsymbol{R}^T\boldsymbol{R}$ is symmetric.*

$$(\boldsymbol{R}^T\boldsymbol{R})^T = \boldsymbol{R}^T(\boldsymbol{R}^T)^T = \boldsymbol{R}^T\boldsymbol{R}$$

Remark 3.1.5 *An $\boldsymbol{m}$ by $\boldsymbol{n}$ orthonormal matrix satisfies $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}_n$.*

The $\boldsymbol{n}$ by $\boldsymbol{n}$ identity matrix, denoted $\boldsymbol{I}_n$ is given by

$$\begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \cdots & \mathbf{0} \\ & & & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1} \end{bmatrix}$$

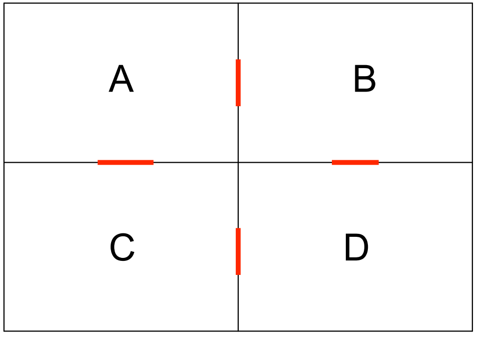
and is such that

$$\boldsymbol{A}\boldsymbol{I}_n = \boldsymbol{I}_m\boldsymbol{A} = \boldsymbol{A},$$

where $\boldsymbol{A}$ is an $\boldsymbol{m}$ by $\boldsymbol{n}$ matrix.

3.2. Applications

3.2.1. Markov Chains: Moving Rooms



P = 1/2

We may write a matrix $\boldsymbol{P}$ that describes the probability of your next move. This is known as the transition matrix and is given by

$$\boldsymbol{P} = \begin{bmatrix} \mathbf{0} & \mathbf{1/2} & \mathbf{1/2} & \mathbf{0} \\ \mathbf{1/2} & \mathbf{0} & \mathbf{0} & \mathbf{1/2} \\ \mathbf{1/2} & \mathbf{0} & \mathbf{0} & \mathbf{1/2} \\ \mathbf{0} & \mathbf{1/2} & \mathbf{1/2} & \mathbf{0} \end{bmatrix}.$$

Using the vector $\mathbf{x}^{\boldsymbol{k}} \in \mathbb{R}^4$ to denote the probability vector of being in a particular room on the $\boldsymbol{k}-$transition and starting in room $\boldsymbol{A}$,

$$\mathbf{x}^0 = \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \text{ and } \mathbf{x}^1 = \boldsymbol{P}\mathbf{x}^0 = \begin{bmatrix} \mathbf{0} \\ \mathbf{1/2} \\ \mathbf{1/2} \\ \mathbf{0} \end{bmatrix}.$$

Carry on this chain of events

$$\mathbf{x}^2 = \boldsymbol{P}\mathbf{x}^1 = \begin{bmatrix} \mathbf{1/2} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{1/2} \end{bmatrix}.$$

3.2.2. Population growth: Ducks in the Parks Pond

We will break the duck population within the Parks duck pond into four categories, eggs, yearlings, two year olds and three year olds, we assume no duck is over three.

Using these categories we may represent the duck population for year $\boldsymbol{n}$ with the population vector

$$\boldsymbol{p}_n = \begin{bmatrix} y_0^n \\ y_1^n \\ y_2^n \\ y_3^n \end{bmatrix}.$$

In a given year a yearling produces four eggs, a two year old twenty eggs and a three and over produces sixty eggs. Each year one in twenty eggs will survive, three in ten yearlings will survive, six in ten two year olds will survive. Using this information we may represent the transition from one year $\boldsymbol{p}_n$ to the next $\boldsymbol{p}_{n+1}$ by a matrix-vector product, using the following Leslie Matrix

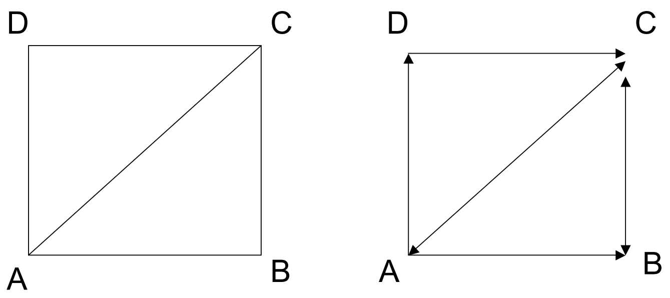
$$\boldsymbol{L} = \begin{array}{c} \boldsymbol{eggs} \\ \boldsymbol{yearlings} \\ \boldsymbol{two\ year\ olds} \\ \boldsymbol{three\ and\ over} \end{array} \begin{bmatrix} \mathbf{0} & \mathbf{4} & \mathbf{20} & \mathbf{60} \\ \mathbf{0.05} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0.3} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0.6} & \mathbf{0} \end{bmatrix}.$$

This leads to

$$\boldsymbol{p}_{n+1} = \boldsymbol{L}\boldsymbol{p}_n.$$

3.2.3. Adjacency Matrices: Spreading of roumors

The figure illustrates two types of graphs showing how $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ are connected. In the first graph we have node $\boldsymbol{A}$ linked to all the other three nodes, node $\boldsymbol{B}$ linked to nodes $\boldsymbol{A}$ and $\boldsymbol{C}$, node $\boldsymbol{C}$ linked to all other nodes and node $\boldsymbol{D}$ linked to $\boldsymbol{A}$ and $\boldsymbol{C}$.



Graph Digraph

We may express this information in the form of a matrix

$$M_1 = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix},$$

where we have expressed a link by the value 1 and no link by 0. If the links can possibly be only made in one direction, see the digraph, then we may also express this information in the form of a matrix

$$M_2 = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

This could be the traffic flow on a combination of one way and two way streets. The matrices are known as Adjacency Matrices.
The values of the squared matrices represent interesting properties about the graph.

For example, if we calculate M_1^2 and M_2^2,

$$M_1^2 = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 1 & 2 & 1 \\ 1 & 2 & 1 & 2 \\ 2 & 1 & 3 & 1 \\ 1 & 2 & 1 & 2 \end{bmatrix}$$

and

$$M_2^2 = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 2 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix},$$

then the elements of these squared matrices represent the number of different two paths between two nodes represented in the (i, j) entry. For example $M_1^2(2, 4) = 2$ and this is how many different ways you may get from B to D using two paths only, these are B to C and then C to D and B to A and then A to D. In the other case of $M_2^2(2, 4) = 0$ and so there are no two paths between B and D.

Exercise 3.2.1 *Gossip between a group of five friends, Anne, Bert, Carla, David, and Eric is usually received by email. But not all friends email all the others immediately. It is known that any gossip received by Ann will be sent to Carla and Eric, Bert will send it to Carla and David, Carla will send it to Eric, David will send it to Anne and Carla and Eric will send it to Bert. Draw a digraph for this email gossip spreading and the corresponding adjacency matrix. If Anne hears a rumour how many steps will it take for everyone else to hear it?*

3.3. Column, row, null space and rank of a matrix

3.3.1. Columns and rows in matrices and matrix operations revisited

Recall that elements in an $m \times n$ matrix can be arranged into column vectors $\mathbf{a}_i \in \mathbb{R}^{m\times 1}, i = 1, ..., n,$

$$\boldsymbol{A} = \left[\begin{array}{c|c|c} & & \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ & & \end{array}\right] \in \mathbb{R}^{m\times n}.$$

Similarly elements of an $n \times m$ matrix can be arranged into row vectors

$$\boldsymbol{A}^T = \left[\begin{array}{c} \mathbf{a}_1^T \\ \hline \cdot \\ \hline \mathbf{a}_n^T \end{array}\right] \in \mathbb{R}^{n\times m}.$$

Using this level of description of matrices, we can revisit matrix operations.

Matrix-Vector Multiplication

Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{x} \in \mathbb{R}^{n\times 1}$, column vectors. Then,

$$\boldsymbol{A}\mathbf{x} = \left[\begin{array}{c|c|c} \mathbf{a}_1 & \cdot & \mathbf{a}_n \end{array}\right]\left[\begin{array}{c} \mathbf{x} \end{array}\right] = x_1\left[\begin{array}{c} \mathbf{a}_1 \end{array}\right] + \cdots + x_n\left[\begin{array}{c} \mathbf{a}_n \end{array}\right].$$

The matrix-vector product, $\boldsymbol{A}\mathbf{x}$ can be expressed as a linear combination of the columns of $\boldsymbol{A}$! Furthermore, we have

$$\mathbf{x}^T\boldsymbol{A}^T = [\quad \mathbf{x}^T \quad]\left[\begin{array}{c} \mathbf{a}_1^T \\ \hline \cdot \\ \hline \mathbf{a}_n^T \end{array}\right] =$$

$$x_1[\quad \mathbf{a}_1^T \quad] + \cdots + x_n[\quad \mathbf{a}_n^T \quad].$$

Matrix-Matrix Multiplication

Let $\boldsymbol{A} \in \mathbb{R}^{l \times m}$, $\boldsymbol{B} \in \mathbb{R}^{m \times n}$ are matrices expressed in terms of column vectors. Then the the matrix product $\boldsymbol{AB}$ can be written as

$$\left[\begin{array}{c|c|c} & & \\ \mathbf{a}_1 & \cdot & \mathbf{a}_m \\ & & \end{array}\right] \left[\begin{array}{c|c|c} & & \\ \mathbf{b}_1 & \cdots & \mathbf{b}_n \\ & & \end{array}\right] = \left[\begin{array}{c|c|c} & & \\ \boldsymbol{A}\mathbf{b}_1 & \cdots & \boldsymbol{A}\mathbf{b}_n \\ & & \end{array}\right],$$

where $\boldsymbol{C} = \boldsymbol{AB} \in \mathbb{R}^{l \times n}$ is an $\boldsymbol{l}$ by $\boldsymbol{n}$ matrix expressed in terms of column vectors,

$$\mathbf{c}_j = \boldsymbol{A}\mathbf{b}_j,\ \boldsymbol{j} = 1, \ldots, \boldsymbol{n}.$$

3.3.2. Column space, row space and null space

Definition 3.3.1 *Let $\boldsymbol{A}$ be an $\boldsymbol{m}$ by $\boldsymbol{n}$ matrix, the column space, denoted $\boldsymbol{\mathcal{C}}(\boldsymbol{A})$, contains all linear combinations of the columns of $\boldsymbol{A}$ and it is a subspace of $\mathbb{R}^m$. $\boldsymbol{\mathcal{C}}(\boldsymbol{A})$ is sometimes also called the range of a matrix.*

Using this space we have the following theorem:

Theorem 3.3.2 *Let $\boldsymbol{A}$ be an $\boldsymbol{m}$ by $\boldsymbol{n}$ matrix. The system $\boldsymbol{A}\mathbf{x} = \mathbf{b}$, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$ is solvable if and only if $\mathbf{b}$ can be expressed as a linear combination of the columns of $\boldsymbol{A}$, $\mathbf{b} \in \boldsymbol{\mathcal{C}}(\boldsymbol{A})$.*

Example 3.3.3 *Let*

$$\boldsymbol{A} = \begin{bmatrix} \mathbf{1} & \mathbf{2} \\ \mathbf{0} & \mathbf{1} \\ \mathbf{3} & \mathbf{2} \end{bmatrix}.$$

Consider the problem: Given $\mathrm{b} = [\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3]^T$ *find* $\mathrm{x} \in \mathbb{R}^2$ *such that* $\boldsymbol{A}\mathrm{x} = \mathrm{b}$.

This may be written in the equivalent form: Find $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}$ *such that*

$$\boldsymbol{x}_1 \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \\ \mathbf{3} \end{bmatrix} + \boldsymbol{x}_2 \begin{bmatrix} \mathbf{2} \\ \mathbf{1} \\ \mathbf{2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \\ \boldsymbol{b}_3 \end{bmatrix}.$$

Hence, b *has to be expressed as a linear combination of the columns of* $\boldsymbol{A}$.

For the above example for solvability we require that b lies in the plane defined by the two column vectors of $\boldsymbol{A}$.

Definition 3.3.4 *Let $\boldsymbol{A}$ be an $\boldsymbol{m}$ by $\boldsymbol{n}$ matrix, the row space, denoted $\boldsymbol{\mathcal{R}}(\boldsymbol{A})$, contains all linear combinations of the rows of $\boldsymbol{A}$ and it is a subspace of $\mathbb{R}^{\boldsymbol{n}}$.*

Definition 3.3.5 *The nullspace of an m by n matrix A, consists of all vectors* $\mathbf{x}$ *such that* $A\mathbf{x} = \mathbf{0}$. *It is denoted by* $\mathcal{N}(A)$. *It is subspace of* $\mathbb{R}^n$.

If we have a particular solution $\mathbf{x}_{\boldsymbol{p}}$ of $\boldsymbol{A}\mathbf{x} = \mathbf{b}$, $\boldsymbol{A}\mathbf{x}_{\boldsymbol{p}} = \mathbf{b}$, then any vector $\mathbf{x}_{\boldsymbol{n}} \in \boldsymbol{\mathcal{N}}(\boldsymbol{A})$ may be added to $\mathbf{x}_{\boldsymbol{p}}$ and still satisfy

$$\boldsymbol{A}(\mathbf{x}_{\boldsymbol{p}} + \mathbf{x}_{\boldsymbol{n}}) = \mathbf{b}.$$

We have seen that if $\boldsymbol{\mathcal{C}}(\boldsymbol{A})$ is not the whole space $\mathbb{R}^m$ then not every $\mathbf{b} \in \mathbb{R}^m$ will give a solution to $\boldsymbol{A}\mathbf{x} = \mathbf{b}$.

Consider the case $\boldsymbol{n} = \mathbf{1}$. We have

1. $\mathbf{0}\boldsymbol{x} = \boldsymbol{b}$ this only has solution if $\boldsymbol{b} = \mathbf{0}$. In this case $\boldsymbol{\mathcal{C}}(\mathbf{0}) = \mathbf{0}$.
2. $\mathbf{0}\boldsymbol{x} = \mathbf{0}$ this has infinitely many solutions. In this case we have $\boldsymbol{\mathcal{N}}(\mathbf{0}) = \mathbb{R}$. Choosing $\boldsymbol{x_p} = \mathbf{0}$, the complete set of solutions $\boldsymbol{x} = \boldsymbol{x_p} + \boldsymbol{x_n}$ for any $\boldsymbol{x_n} \in \boldsymbol{\mathcal{N}}(\mathbf{0}) = \mathbb{R}$.

Consider the problem $\boldsymbol{A}\mathbf{x} = \mathbf{b}$ for the case $\boldsymbol{n} = \mathbf{2}$ and the matrix

$$\boldsymbol{A} = \begin{bmatrix} \mathbf{1} & \mathbf{1} \\ \mathbf{2} & \mathbf{2} \end{bmatrix}.$$

For any $\mathbf{b} \in \mathbb{R}^2$ this can be written in the form:
Find $\mathbf{x} = [\boldsymbol{x}, \boldsymbol{y}]^T$, $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}$ such that

$$\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{b}_1 \quad \mathbf{2}\boldsymbol{x} + \mathbf{2}\boldsymbol{y} = \boldsymbol{b}_2.$$

Clearly this only has a solution when $\boldsymbol{b}_2 = \mathbf{2}\boldsymbol{b}_1$. In this case $\mathcal{C}(\boldsymbol{A})$ is the set of vectors that are multiples of $[\mathbf{1}, \mathbf{2}]^T$.

In the case $\boldsymbol{b}_2 = \mathbf{2}\boldsymbol{b}_1$ we have infinitely many solutions. A particular solution is $\mathbf{x}_p = [\mathbf{1}, \mathbf{1}]^T$. $\mathcal{N}(\boldsymbol{A})$ is the set of all vectors of the form $\boldsymbol{\alpha}[-\mathbf{1}, \mathbf{1}]^T$, $\boldsymbol{\alpha} \in \mathbb{R}$. Hence, the complete set of solutions is given by

$$\mathbf{x}_p + \mathbf{x}_n = \begin{bmatrix} \mathbf{1} - \boldsymbol{\alpha} \\ \mathbf{1} + \boldsymbol{\alpha} \end{bmatrix}.$$

Theorem 3.3.6 *The columns of* $\boldsymbol{A}$ *are linearly independent exactly when* $\boldsymbol{\mathcal{N}}(\boldsymbol{A}) = \{\boldsymbol{0}\}$.

Theorem 3.3.7 *The row space of a matrix $\boldsymbol{A}$ is orthogonal to the nullspace, $\mathcal{N}(\boldsymbol{A})$. The column space $\mathcal{C}(\boldsymbol{A})$ is orthogonal to the left nullspace, $\mathcal{N}(\boldsymbol{A}^T)$.*

proof 3.3.8 *Suppose $\mathbf{x} \in \mathcal{N}(\boldsymbol{A}) => \boldsymbol{A}\mathbf{x} = \mathbf{0}$. Denote $\mathbf{a}_i^T$ to be the $\boldsymbol{i}-$th row of $\boldsymbol{A}$. Then*

$$(\boldsymbol{A}\mathbf{x})_i = \mathbf{a}_i^T \mathbf{x} = \mathbf{0},$$

as required.
Now suppose $\mathbf{y}^T \boldsymbol{A} = \mathbf{0}$. Denote $\mathbf{a}_i$ to be the $\boldsymbol{i}-$th column of $\boldsymbol{A}$. Then

$$(\mathbf{y}^T \boldsymbol{A})_i = \mathbf{y}^T \mathbf{a}_i = \mathbf{0},$$

as required.

We know that the vector $\boldsymbol{A}\mathbf{x} \in \boldsymbol{\mathcal{C}}(\boldsymbol{A})$ and nothing is carried to the left nullspace. For $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ we may write $\mathbf{x} \in \mathbb{R}^n$ in the form $\mathbf{x} = \mathbf{x}_r + \mathbf{x}_n$, where $\mathbf{x}_r$ is in the row space and $\mathbf{x}_n$ is in the null space.

$$\boldsymbol{A}\mathbf{x}_n = \mathbf{0} \quad \boldsymbol{A}\mathbf{x} = \boldsymbol{A}\mathbf{x}_r$$

3.3.3. The rank of a matrix

The column rank of a matrix is the dimension of its column space and the row rank of a matrix is the dimension spanned by its rows. Column rank and row rank are always equal.

Theorem 3.3.9 *The dimension of $\mathcal{C}(A)$ equals the dimension of the space spanned by the rows. The number of independent columns equals the number of independent rows,*

$$\boldsymbol{dim}(\boldsymbol{\mathcal{C}}(\boldsymbol{A})) = \boldsymbol{dim}(\boldsymbol{\mathcal{R}}(\boldsymbol{A})).$$

For interest

proof 3.3.10 *Let A be an m by n matrix*

$$A = \left[\begin{array}{c} \mathbf{a}_1 \\ \hline \mathbf{a}_2 \\ \hline \cdot \\ \hline \mathbf{a}_m \end{array}\right], \quad dim(\mathcal{R}(A)) = k.$$

Hence, there exists k linearly independent row vectors $\mathbf{r}^l$, $l = 1, \ldots, k$ such that any $\mathbf{a}_i$, $i = 1, \ldots, m$ may be written in the form

$$\mathbf{a}_i = c_{i1}\mathbf{r}^1 + c_{i2}\mathbf{r}^2 + \cdots + c_{ik}\mathbf{r}^k.$$

Using the notation $\boldsymbol{A} = \{\boldsymbol{a}_{ij}\}$ *we obtain (expanding the above vector equations component-wise)*

$$\boldsymbol{a}_{1j} = \boldsymbol{c}_{11}\boldsymbol{r}_j^1 + \boldsymbol{c}_{12}\boldsymbol{r}_j^2 + \cdots + \boldsymbol{c}_{1k}\boldsymbol{r}_j^k$$

$$\boldsymbol{a}_{2j} = \boldsymbol{c}_{21}\boldsymbol{r}_j^1 + \boldsymbol{c}_{22}\boldsymbol{r}_j^2 + \cdots + \boldsymbol{c}_{2k}\boldsymbol{r}_j^k$$

$$\begin{array}{cccccc} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{array}$$

$$\boldsymbol{a}_{mj} = \boldsymbol{c}_{m1}\boldsymbol{r}_j^1 + \boldsymbol{c}_{m2}\boldsymbol{r}_j^2 + \cdots + \boldsymbol{c}_{mk}\boldsymbol{r}_j^k$$

Which we may write in the matrix form

$$\begin{bmatrix} \boldsymbol{a_{1j}} \\ \boldsymbol{a_{2j}} \\ \cdot \\ \cdot \\ \boldsymbol{a_{mj}} \end{bmatrix} = \boldsymbol{r_j^1} \begin{bmatrix} \boldsymbol{c_{11}} \\ \boldsymbol{c_{21}} \\ \cdot \\ \cdot \\ \boldsymbol{c_{m1}} \end{bmatrix} + \boldsymbol{r_j^2} \begin{bmatrix} \boldsymbol{c_{12}} \\ \boldsymbol{c_{22}} \\ \cdot \\ \cdot \\ \boldsymbol{c_{m2}} \end{bmatrix} + \cdots + \boldsymbol{r_j^k} \begin{bmatrix} \boldsymbol{c_{1k}} \\ \boldsymbol{c_{2k}} \\ \cdot \\ \cdot \\ \boldsymbol{c_{mk}} \end{bmatrix}.$$

In other words, we may express the $\boldsymbol{j}-$*th column of* $\boldsymbol{A}$ *by a set of no more than* $\boldsymbol{k}$ *vectors. Thus the dimension of the column space is no more than* $\boldsymbol{k}$*.*

Hence,

$$\boldsymbol{dim(\mathcal{C}(A)) \leq dim(\mathcal{R}(A))}.$$

Furthermore, applying this result to $\boldsymbol{A^T}$

$$\boldsymbol{dim(\mathcal{C}(A^T)) \leq dim(\mathcal{R}(A^T))},$$

or

$$\boldsymbol{dim(\mathcal{R}(A)) \leq dim(\mathcal{C}(A))},$$

and we conclude that

$$\boldsymbol{dim(\mathcal{C}(A)) = dim(\mathcal{R}(A))},$$

and we refer to this number as the rank of a $\boldsymbol{A}$,

$$\boldsymbol{rank(A)}.$$

Exercise 3.3.11 *Let* $\boldsymbol{A, B} \in \mathbb{R}^{n\times n}$ *matrices.*
Prove that $\boldsymbol{rank(AB) \leq \min(rank(A), rank(B))}$.

3.4. Inverse

Definition 3.4.1 *The inverse of an* $\boldsymbol{n}$ *by* $\boldsymbol{n}$ *matrix* $\boldsymbol{A}$*, denoted* $\boldsymbol{A}^{-1}$*, satisfies*

$$\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}_n.$$

Remark 3.4.2 *If* $\boldsymbol{Q}$ *is an* $\boldsymbol{n}$ *by* $\boldsymbol{n}$ *orthonormal matrix, then* $\boldsymbol{Q}^T = \boldsymbol{Q}^{-1}$*, since* $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{Q}^T = \boldsymbol{I}_n$.

The definition of inverse implies that if you multiply $\mathbf{x}$ by $\boldsymbol{A}$ and then multiply by $\boldsymbol{A}^{-1}$ you return to $\mathbf{x}$.

Example 3.4.3 *Consider the following problem of finding* $\mathbf{x}$ *in*

$$\boldsymbol{A}\mathbf{x} = \mathbf{b}.$$

Multiplying this equation by $\boldsymbol{A}^{-1}$ *gives*

$$\boldsymbol{A}^{-1}\boldsymbol{A}\mathbf{x} = \mathbf{x} = \boldsymbol{A}^{-1}\mathbf{b}.$$

NOT ALL MATRICES HAVE INVERSES

Example 3.4.4 *Consider a non-zero* $\mathbf{x}$ *and a matrix* $\boldsymbol{A}$ *such that* $\boldsymbol{A}\mathbf{x} = \mathbf{0}$*. Then if this matrix had an inverse it would have to return* $\mathbf{0}$ *back to* $\mathbf{x}$ *and no matrix can do this, e.g.*

$$\boldsymbol{A} = \begin{bmatrix} \mathbf{1} & \mathbf{-1} \\ \mathbf{-1} & \mathbf{1} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \end{bmatrix}.$$

Theorem 3.4.5 *If the inverse exists it is unique.*

proof 3.4.6 *When showing uniqueness it is common to suppose there exists at least two solutions, in this case* $\boldsymbol{B}$ *and* $\boldsymbol{C}$*, and then show they must be the same.*

Suppose $\boldsymbol{BA} = \boldsymbol{I} = \boldsymbol{AB}$ *and* $\boldsymbol{CA} = \boldsymbol{I} = \boldsymbol{AC}$*.*
Therefore,

$$\boldsymbol{B} = \boldsymbol{BI} = \boldsymbol{B}(\boldsymbol{AC}) = (\boldsymbol{BA})\boldsymbol{C} = \boldsymbol{IC} = \boldsymbol{C}.$$

Example 3.4.7 *A 2 by 2 matrix is invertible if and only if $ad \neq bc$. In this case the inverse is given as follows*

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

A product of two invertible matrices, AB, is invertible with inverse

$$(AB)^{-1} = B^{-1}A^{-1}.$$

Exercise 3.4.8 *Prove that $(AB)^{-1} = B^{-1}A^{-1}$.*

Exercise 3.4.9 *Show that the transpose of A^{-1} is $(A^{-1})^T = (A^T)^{-1}$.*

3.5. Elementary Matrices

3.5.1. Elementary Matrices and Elementary Matrix Transformations

Definition 3.5.1 *An n by n elementary matrix is any matrix that can be obtained by performing an elementary row operation on the n by n identity matrix.*

The three types of elementary matrices are a scaling of a row of $\boldsymbol{I}$, an interchange of two rows of $\boldsymbol{I}$ and adding a multiple of one row to another row. For example, assuming that $\boldsymbol{s}$,$\boldsymbol{c} \in \mathbb{R}$, we have

$$\boldsymbol{E}_{r_2 \leftarrow s r_2} = \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{s} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix}$$

$$\boldsymbol{E}_{r_1 \leftrightarrow r_3} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{1} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{1} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

$$\boldsymbol{E}_{r_3 \leftarrow r_3 + c r_2} = \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{c} & \mathbf{1} \end{bmatrix}.$$

Let $\boldsymbol{A}$ be an $\mathbf{3}$ by $\mathbf{4}$ matrix

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_{11} & \boldsymbol{a}_{12} & \boldsymbol{a}_{13} & \boldsymbol{a}_{14} \\ \boldsymbol{a}_{21} & \boldsymbol{a}_{22} & \boldsymbol{a}_{23} & \boldsymbol{a}_{24} \\ \boldsymbol{a}_{31} & \boldsymbol{a}_{32} & \boldsymbol{a}_{33} & \boldsymbol{a}_{34} \end{bmatrix}$$

An elementary matrix represents an elementary transformation realized by pre-multiplying $\boldsymbol{A}$ with the elementary matrix. For example, applying the above elementary matrices leads to

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ sa_{21} & sa_{22} & sa_{23} & sa_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix}$$

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} = \begin{bmatrix} a_{31} & a_{32} & a_{33} & a_{34} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{11} & a_{12} & a_{13} & a_{14} \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & c & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} =$$

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ ca_{21} + a_{31} & ca_{22} + a_{32} & ca_{23} + a_{33} & ca_{24} + a_{34} \end{bmatrix}.$$

Given that all these elementary transformations are invertible, each elementary matrix is invertible. For example,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & c & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & c & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & c & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -c & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & s^{-1} & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Properties of elementary matrices:

1. Let $\boldsymbol{E}$ be an elementary matrix obtained by performing an elementary row operation on $\boldsymbol{I}$. If the same elementary row operation is performed on an $\boldsymbol{n}$ by $\boldsymbol{r}$ matrix A, the resulting matrix is $\boldsymbol{EA}$.

2. Each elementary matrix is invertible, and its inverse is an elementary matrix of the same type. We can see this since $\boldsymbol{E}$ is an elementary row operation on $\boldsymbol{I}$ and so this operation can just be reversed by another elementary row operation to get back to $\boldsymbol{I}$.

3. An elementary matrix performing the subtraction of a multiple of one row from another is always of the form $\boldsymbol{I} + \boldsymbol{B}$, where $\boldsymbol{B}$ has only the one non-zero entry $\boldsymbol{b_{ij}}$, $\boldsymbol{i} \neq \boldsymbol{j}$ and this value represents the multiple of the $\boldsymbol{j}-$th row that is to be added to the $\boldsymbol{i}-$th.

3.5.2. Row Reduction by Elementary Matrix Transformations

By successively applying elementary row operations, every matrix can be transformed into the following forms discussed below.

A matrix is in row echelon form (REF) if:

- all nonzero rows are above any rows of all zero.
- the leading coefficient (the first nonzero number from the left, also called the pivot) of a nonzero row is always strictly to the right of the leading coefficient of the row above it.

A matrix is in reduced row echelon form (RREF) if:

- it is in row echelon form and
- every leading coefficient is 1 and is the only nonzero entry in its column.

Example 3.5.2

$$\boldsymbol{A} = \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix}$$

We start applying the elementary row operations until reaching the row echelon and reduced row echelon forms.

$$\boldsymbol{E}_1\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ -2 & 7 & 2 \end{bmatrix}$$

$$\boldsymbol{E}_2(\boldsymbol{E}_1\boldsymbol{A}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ -2 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 8 & 3 \end{bmatrix}$$

$$\boldsymbol{E}_3(\boldsymbol{E}_2\boldsymbol{E}_1\boldsymbol{A}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 8 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 0 & 1 \end{bmatrix} \textit{REF}$$

$$\boldsymbol{E_4(E_3E_2E_1A)} = \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{2} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{2} & \mathbf{1} & \mathbf{1} \\ \mathbf{0} & \mathbf{-8} & \mathbf{-2} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} = \begin{bmatrix} \mathbf{2} & \mathbf{1} & \mathbf{1} \\ \mathbf{0} & \mathbf{-8} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix}$$

$$\boldsymbol{E_5(E_4E_3E_2E_1A)} = \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{-1} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{2} & \mathbf{1} & \mathbf{1} \\ \mathbf{0} & \mathbf{-8} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} = \begin{bmatrix} \mathbf{2} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{-8} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix}$$

$$\boldsymbol{E_6(E_5E_4E_3E_2E_1A)} = \begin{bmatrix} \mathbf{1} & \mathbf{\frac{1}{8}} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{2} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{-8} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} = \begin{bmatrix} \mathbf{2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{-8} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix}$$

$$\boldsymbol{E_7(E_6E_5E_4E_3E_2E_1A)} = \begin{bmatrix} \mathbf{\frac{1}{2}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{-8} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{-8} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix}$$

$$\boldsymbol{E_8(E_7E_6E_5E_4E_3E_2E_1A)} = \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{-\frac{1}{8}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{-8} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \; \textit{RREF}$$

The process of transforming the matrix into Row Echelon Form is called Gaussian elimination, while the one that leads to Reduced Row Echelon Form is often called Gauss-Jordan elimination.

3.5.3. Elementary Matrices and Inverse

We saw that the application of elementary row operations, which are represented by the matrices $\{\boldsymbol{E}_i\}_{i=1}^{k}$ (with $\boldsymbol{k} = \mathbf{8}$ in previous example), leads to the transformation of a matrix into its Reduced Row Echelon Form by the successive evaluation of matrices, $\boldsymbol{A}_1 = \boldsymbol{E}_1\boldsymbol{A}$, $\boldsymbol{A}_2 = \boldsymbol{E}_2(\boldsymbol{E}_1\boldsymbol{A})$, $\boldsymbol{A}_3 = \boldsymbol{E}_3(\boldsymbol{E}_2(\boldsymbol{E}_1\boldsymbol{A}))$ and so on.

Given that matrix multiplication is associative, we have

$$\boldsymbol{A}_{RREF} = \boldsymbol{A}_k = \boldsymbol{E}_k\boldsymbol{E}_{k-1}\cdots\boldsymbol{E}_1\boldsymbol{A},$$

where $\boldsymbol{A}_{RREF}$ is the Reduced Row Echelon Form of $\boldsymbol{A}$.

Let be $\boldsymbol{A}$ an $\boldsymbol{n}$ by $\boldsymbol{n}$ matrix. Since all $\{\boldsymbol{E_i}\}_{i=1}^{k}$ is invertible, the equation

$$\boldsymbol{A_{RREF}} = \boldsymbol{E_k E_{k-1} \cdots E_1 A}$$

implies that $\boldsymbol{A}$ is invertible if and only if $\boldsymbol{A_{RREF}}$ is invertible.

However, $\boldsymbol{A_{RREF}}$ is in Reduced Row Echelon Form, so it either has an all zero row or it is the identity matrix, $\boldsymbol{I_n}$.

To sum up, $\boldsymbol{A}$ is invertible if and only if $\boldsymbol{A_{RREF}}$ is invertible if and only if $\boldsymbol{A_{RREF}} = \boldsymbol{I_n}$. With these consideration, we can state the following theorem.

Theorem 3.5.3 *An n by n matrix is invertible if and only if its Reduced Row Echelon Form is the identity matrix, I_n.*

Remark 3.5.4 *A will be not invertible if and only if A_{RREF} is not invertible if and only if A_{RREF} has an all zero row.*

Assuming that $\boldsymbol{A}$ is invertible we have both $\boldsymbol{A}_{\boldsymbol{RREF}} = \boldsymbol{I}_{\boldsymbol{n}}$ and $\boldsymbol{A}_{\boldsymbol{RREF}} = \boldsymbol{E}_k \boldsymbol{E}_{k-1} \cdots \boldsymbol{E}_1 \boldsymbol{A}$, implying that

$$\boldsymbol{E}_k \boldsymbol{E}_{k-1} \cdots \boldsymbol{E}_1 \boldsymbol{A} = \boldsymbol{I}_{\boldsymbol{n}}.$$

Corollary 3.5.5 *If $\boldsymbol{A}$ is invertible then $\boldsymbol{A}^{-1} = \boldsymbol{E}_k \boldsymbol{E}_{k-1} \cdots \boldsymbol{E}_1$, where $\{\boldsymbol{E}_i\}_{i=1}^{k}$ are the matrices representing the elementary row operations that are applied sequentially until the RREF of $\boldsymbol{A}$ is obtained. Furthermore, the matrix*

$$\boldsymbol{E} = \boldsymbol{E}_k \boldsymbol{E}_{k-1} \cdots \boldsymbol{E}_1$$

represents this entire sequence of elementary row transformations.

Following up on example 3.5.2 we can evaluate, $\boldsymbol{E} = \boldsymbol{E}_8\boldsymbol{E}_7 \cdots \boldsymbol{E}_1$.

$$\boldsymbol{E}_2\boldsymbol{E}_1 = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

$$\boldsymbol{E}_3(\boldsymbol{E}_2\boldsymbol{E}_1) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix}$$

$$\boldsymbol{E}_4(\boldsymbol{E}_3\boldsymbol{E}_2\boldsymbol{E}_1) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -4 & 3 & 2 \\ -1 & 1 & 1 \end{bmatrix}$$

$$\boldsymbol{E}_5(\boldsymbol{E}_4\boldsymbol{E}_3\boldsymbol{E}_2\boldsymbol{E}_1) = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -4 & 3 & 2 \\ -1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -1 & -1 \\ -4 & 3 & 2 \\ -1 & 1 & 1 \end{bmatrix}$$

$$\boldsymbol{E_6(E_5E_4E_3E_2E_1)} = \begin{bmatrix} \mathbf{1} & \frac{\mathbf{1}}{\mathbf{8}} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{2} & \mathbf{-1} & \mathbf{-1} \\ \mathbf{-4} & \mathbf{3} & \mathbf{2} \\ \mathbf{-1} & \mathbf{1} & \mathbf{1} \end{bmatrix} = \begin{bmatrix} \frac{\mathbf{12}}{\mathbf{8}} & -\frac{\mathbf{5}}{\mathbf{8}} & -\frac{\mathbf{6}}{\mathbf{8}} \\ \mathbf{-4} & \mathbf{3} & \mathbf{2} \\ \mathbf{-1} & \mathbf{1} & \mathbf{1} \end{bmatrix}$$

$$\boldsymbol{E_7(E_6E_5E_4E_3E_2E_1)} = \begin{bmatrix} \frac{\mathbf{1}}{\mathbf{2}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \frac{\mathbf{12}}{\mathbf{8}} & -\frac{\mathbf{5}}{\mathbf{8}} & -\frac{\mathbf{6}}{\mathbf{8}} \\ \mathbf{-4} & \mathbf{3} & \mathbf{2} \\ \mathbf{-1} & \mathbf{1} & \mathbf{1} \end{bmatrix} = \begin{bmatrix} \frac{\mathbf{12}}{\mathbf{16}} & -\frac{\mathbf{5}}{\mathbf{16}} & -\frac{\mathbf{6}}{\mathbf{16}} \\ \mathbf{-4} & \mathbf{3} & \mathbf{2} \\ \mathbf{-1} & \mathbf{1} & \mathbf{1} \end{bmatrix}$$

$$\begin{aligned} \boldsymbol{E_8(E_7E_6E_5E_4E_3E_2E_1)} &= \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\frac{\mathbf{1}}{\mathbf{8}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \frac{\mathbf{12}}{\mathbf{16}} & -\frac{\mathbf{5}}{\mathbf{16}} & -\frac{\mathbf{6}}{\mathbf{16}} \\ \mathbf{-4} & \mathbf{3} & \mathbf{2} \\ \mathbf{-1} & \mathbf{1} & \mathbf{1} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\mathbf{12}}{\mathbf{16}} & -\frac{\mathbf{5}}{\mathbf{16}} & -\frac{\mathbf{6}}{\mathbf{16}} \\ \frac{\mathbf{4}}{\mathbf{8}} & -\frac{\mathbf{3}}{\mathbf{8}} & -\frac{\mathbf{2}}{\mathbf{8}} \\ \mathbf{-1} & \mathbf{1} & \mathbf{1} \end{bmatrix} \end{aligned}$$

Therefore,

$$\boldsymbol{E} = \begin{bmatrix} \frac{12}{16} & -\frac{5}{16} & -\frac{6}{16} \\ \frac{4}{8} & -\frac{3}{8} & -\frac{2}{8} \\ -1 & 1 & 1 \end{bmatrix}.$$

It is clear from corollary 3.5.5 that $\boldsymbol{E}$ is the inverse of

$$\boldsymbol{A} = \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix}$$

in Example 3.5.2.

To see we could obtain $\boldsymbol{EA}$,

$$\begin{bmatrix} \frac{12}{16} & -\frac{5}{16} & -\frac{6}{16} \\ \frac{4}{8} & -\frac{3}{8} & -\frac{2}{8} \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

3.6. Finding the inverse

Suppose that $\boldsymbol{A}$ is an invertible $\boldsymbol{n}$ by $\boldsymbol{n}$ matrix. Therefore its RREF is $\boldsymbol{I}_n$ and

$$\boldsymbol{A}^{-1} = \boldsymbol{E}_k \boldsymbol{E}_{k-1} \cdots \boldsymbol{E}_1,$$

where $\{\boldsymbol{E}_i\}_{i=1}^{k}$ is the sequence of elementary row operations applied on $\boldsymbol{A}$ to produce its RREF ($\boldsymbol{I}_n$). Therefore,

$$\boldsymbol{E}_k \boldsymbol{E}_{k-1} \cdots \boldsymbol{E}_1 \boldsymbol{A} = \boldsymbol{I}_n$$

and

$$\boldsymbol{E}_k \boldsymbol{E}_{k-1} \cdots \boldsymbol{E}_1 \boldsymbol{I}_n = \boldsymbol{A}^{-1}.$$

This suggests an algorithm of finding the inverse of $\boldsymbol{A}$.

The Gauss-Jordan method

First we compose an $\boldsymbol{n}$ by $\boldsymbol{2n}$ augmented matrix

$$[\boldsymbol{A} \mid \boldsymbol{I}_n].$$

Next, we apply the sequence of elementary row operations $\{\boldsymbol{E}_i\}_{i=1}^k$ on the augmented matrix until it is in RREF.

At that stage,

$$\boldsymbol{E}_k \boldsymbol{E}_{k-1} \cdots \boldsymbol{E}_1 [\boldsymbol{A} \mid \boldsymbol{I}_n] = [\boldsymbol{I}_n \mid \boldsymbol{A}^{-1}]$$

and the resulted augmented matrix

$$[\boldsymbol{I}_n \mid \boldsymbol{A}^{-1}]$$

contains the inverse of $\boldsymbol{A}$.

Example 3.6.1 *Firstly we augment the matrix $\boldsymbol{A}$ with the identity matrix $\boldsymbol{I_n}$. For the previous example 3.5.2 we obtain*

$$
\begin{aligned}
[\boldsymbol{A} \mid \boldsymbol{I_n}] \quad &= \left[\begin{array}{rrr|rrr} \mathbf{2} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{4} & \mathbf{-6} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{-2} & \mathbf{7} & \mathbf{2} & \mathbf{0} & \mathbf{0} & \mathbf{1} \end{array}\right] \\
\textit{First Column} \quad &=> \left[\begin{array}{rrr|rrr} \mathbf{2} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{-8} & \mathbf{-2} & \mathbf{-2} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{8} & \mathbf{3} & \mathbf{1} & \mathbf{0} & \mathbf{1} \end{array}\right] \\
\textit{Second Column} \quad &=> \left[\begin{array}{rrr|rrr} \mathbf{2} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{-8} & \mathbf{-2} & \mathbf{-2} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{-1} & \mathbf{1} & \mathbf{1} \end{array}\right]
\end{aligned}
$$

We must now place zeros above the diagonal and finally scale the diagonal entries to one.

$$\text{\textit{Third column}} \quad => \begin{bmatrix} 2 & 1 & 0 & | & 2 & -1 & -1 \\ 0 & -8 & 0 & | & -4 & 3 & 2 \\ 0 & 0 & 1 & | & -1 & 1 & 1 \end{bmatrix}$$

$$\text{\textit{Second column}} \quad => \begin{bmatrix} 2 & 0 & 0 & | & \frac{12}{8} & -\frac{5}{8} & -\frac{6}{8} \\ 0 & -8 & 0 & | & -4 & 3 & 2 \\ 0 & 0 & 1 & | & -1 & 1 & 1 \end{bmatrix}$$

$$\text{\textit{Scaling}} \quad => \begin{bmatrix} 1 & 0 & 0 & | & \frac{12}{16} & -\frac{5}{16} & -\frac{6}{16} \\ 0 & 1 & 0 & | & \frac{4}{8} & -\frac{3}{8} & -\frac{2}{8} \\ 0 & 0 & 1 & | & -1 & 1 & 1 \end{bmatrix}$$

In the previous example 3.5.2 we obtained the same results by explicitly forming the elementary matrices representing the elementary row operations. We can observe that the present approach is a more practical procedure leading to the same result.

We can always apply the elementary row operations on $[\boldsymbol{A} \mid \boldsymbol{I}_n]$ but we may not always be able to obtain the identity on the left hand side because Gaussian elimination will produce a REF with an all zero row.

Example 3.6.2

$$[\boldsymbol{A} \mid \boldsymbol{I}_n] = \left[\begin{array}{ccc|ccc} \mathbf{2} & \mathbf{1} & \mathbf{-4} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{-4} & \mathbf{-1} & \mathbf{6} & \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{-2} & \mathbf{2} & \mathbf{-2} & \mathbf{0} & \mathbf{0} & \mathbf{1} \end{array}\right]$$

$$\textit{First Column} \quad => \left[\begin{array}{ccc|ccc} \mathbf{2} & \mathbf{1} & \mathbf{-4} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{-2} & \mathbf{2} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{3} & \mathbf{-6} & \mathbf{1} & \mathbf{0} & \mathbf{1} \end{array}\right]$$

$$\textit{Second Column} \quad => \left[\begin{array}{ccc|ccc} \mathbf{2} & \mathbf{1} & \mathbf{-4} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{-2} & \mathbf{2} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{-5} & \mathbf{-3} & \mathbf{1} \end{array}\right].$$

It is not possible to reduce $\boldsymbol{A}$ to $\boldsymbol{I}$ as this is the case in the latter example having a row of zeros on the left hand side of the augmented matrix. Consequently $\boldsymbol{A}$ is not invertible. This will form part of a main theorem in the next section.

Given $\boldsymbol{A}\mathrm{x} = \mathrm{b}$ problem it can be solved by using a similar approach. First, form an $[\boldsymbol{A} \mid \mathrm{b}]$, an $\boldsymbol{n}$ by $\boldsymbol{n+1}$ augmented matrix. Then apply elementary row operations until the augmented matrix is in RREF.
If $\boldsymbol{A}$ has an inverse, than the same set of elementary matrices, $\{\boldsymbol{E}_i\}_{i=1}^{k}$ that transform $[\boldsymbol{A} \mid \boldsymbol{I}_n]$ into $[\boldsymbol{I}_n \mid \boldsymbol{A}^{-1}]$ transforms $[\boldsymbol{A} \mid \mathrm{b}]$ into $[\boldsymbol{I}_n \mid \mathrm{x}]$.
This relation

$$\boldsymbol{E}_k \cdots \boldsymbol{E}_1 [\boldsymbol{A} \mid \mathrm{b}] = [\boldsymbol{I}_n \mid \mathrm{x}]$$

holds because

$$\boldsymbol{E}_k \cdots \boldsymbol{E}_1 \boldsymbol{A} = \boldsymbol{I}_n$$

and applying $\boldsymbol{E}_k \cdots \boldsymbol{E}_1$ on both sides of $\boldsymbol{A}\mathrm{x} = \mathrm{b}$, we get

$$\boldsymbol{I}_n \mathrm{x} = \boldsymbol{E}_k \cdots \boldsymbol{E}_1 \mathrm{b}.$$

Thus,

$$\boldsymbol{E}_k \cdots \boldsymbol{E}_1 \mathrm{b} = \mathrm{x}.$$

Exercise 3.6.3 *Suppose, we have vectors,*

$$\mathbf{v}_1 = \begin{bmatrix} \mathbf{1} \\ \mathbf{2} \\ \mathbf{4} \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} \mathbf{1} \\ \mathbf{3} \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} \mathbf{1} \\ \mathbf{5} \\ \mathbf{5} \end{bmatrix}.$$

Determine, whether you can express vector

$$\mathbf{u} = \begin{bmatrix} \mathbf{5} \\ \mathbf{8} \\ \mathbf{2} \end{bmatrix}$$

as a linear combination of vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$*?*

Remark 3.6.4 *Using elementary transformations on* $[\boldsymbol{A} \mid \mathbf{b}]$ *will not always lead to* $[\boldsymbol{I_n} \mid \mathbf{x}]$ *as we may not be able to obtain the identity on the left hand side because Gaussian elimination can also produce a REF with an all zero row.*

3.6.1. An Application: Finite Linear Games

Suppose we have five light bulbs, L_1, L_2, L_3, L_4, L_5, in a row and five switches directly below the light bulbs. Each switch changes the state (on or off) of the light bulb above it and that of the two adjacent ones.

For example, given that initially all the lights are off pushing switch L_3 will turn on lights L_2, L_3 and L_4. Then pushing switch L_5 will produce the state L_1 off, L_2 on, L_3 on, L_4 off and L_5 on.

Due to the on/off nature of this problem it seems sensible to work with $\mathbf{x} \in \mathbb{Z}_2^5$, where $x_i = 0$ says L_i is off and $x_i = 1$ says L_i is on. Hence, the initial state is

$$\mathbf{y} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

and using this notation we have the following five vectors, $\mathbf{s}_i$ that represent the action of pressing the i—th switch.

$$\mathbf{s}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \mathbf{s}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \mathbf{s}_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} \mathbf{s}_4 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} \mathbf{s}_5 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

Since $2 = 0$ in $\mathbb{Z}_2$ it is clear that pressing any switch twice results in no change.

Exercise 3.6.5 *Given that all the lights are initially off is it possible (1) to press the switches in some order to give only lights* $\boldsymbol{L_1}$,$\boldsymbol{L_3}$ *and* $\boldsymbol{L_5}$ *on (2) to press the switches in a certain order to leave only* $\boldsymbol{L_1}$ *on?*

3.6.2. The fundamental theorem of invertible matrices

We may now state one of the main results in linear algebra:

Theorem 3.6.6 *The fundamental theorem of Invertible matrices: part I*
Let A be an n by n matrix. The following statements are equivalent:
(a) A is invertible.
(b) $A\mathrm{x} = \mathrm{b}$ has a unique solution for every $\mathrm{b} \in \mathbb{R}^n$.
(c) $A\mathrm{x} = \mathbf{0}$ has only the trivial solution.
(d) Using Gauss-Jordan elimination we may reduce the matrix down to the identity.
(e) A is the product of elementary matrices.

proof 3.6.7 *For equivalence we must show that*

$$(\boldsymbol{a}) \Rightarrow (\boldsymbol{b}) \Rightarrow (\boldsymbol{c}) \Rightarrow (\boldsymbol{d}) \Rightarrow (\boldsymbol{e}) \Rightarrow (\boldsymbol{a}).$$

1. $(\boldsymbol{a}) \Rightarrow (\boldsymbol{b})$
   Since $\boldsymbol{A}$ *has an inverse* $\boldsymbol{A}^{-1}$ *then*

   $$\boldsymbol{A}\mathbf{x} = \mathbf{b} \;\Rightarrow\; \boldsymbol{A}^{-1}\boldsymbol{A}\mathbf{x} = \boldsymbol{A}^{-1}\mathbf{b} \;\Rightarrow\; \mathbf{x} = \boldsymbol{A}^{-1}\mathbf{b},$$

   and so $\boldsymbol{A}\mathbf{x} = \mathbf{b}$ *has a solution (existence). Suppose that both* $\mathbf{x}, \mathbf{y}$ *are such that* $\boldsymbol{A}\mathbf{x} = \mathbf{b}$ *and* $\boldsymbol{A}\mathbf{y} = \mathbf{b}$ *then*

   $$\boldsymbol{A}^{-1}(\boldsymbol{A}\mathbf{y}) = \boldsymbol{A}^{-1}\mathbf{b} \;\Rightarrow\; \mathbf{y} = \boldsymbol{A}^{-1}\mathbf{b} = \mathbf{x}.$$

   (uniqueness).

2. $\boldsymbol{(b) => (c)}$

Assume $\boldsymbol{A}\mathbf{x} = \mathbf{b}$ *has a unique solution for any* $\mathbf{b} \in \mathbb{R}^{\boldsymbol{n}}$. *Hence,* $\boldsymbol{A}\mathbf{x} = \mathbf{0}$ *has only one solution and we know that* $\boldsymbol{A}\mathbf{0} = \mathbf{0}$. *So* $\mathbf{x} = \mathbf{0}$ *must be the only solution.*

3. $(\boldsymbol{c}) => (\boldsymbol{d})$

Suppose $\boldsymbol{A}\mathbf{x} = \mathbf{0}$ *has only a trivial solution. Putting this in augmented (matrix-vector) form*

$$\left[\begin{array}{cccc|c} \boldsymbol{a_{11}} & \boldsymbol{a_{12}} & \cdots & \boldsymbol{a_{1n}} & \mathbf{0} \\ \boldsymbol{a_{21}} & \boldsymbol{a_{22}} & \cdots & \boldsymbol{a_{2n}} & \mathbf{0} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \boldsymbol{a_{n1}} & \boldsymbol{a_{n2}} & \cdots & \boldsymbol{a_{nn}} & \mathbf{0} \end{array}\right].$$

Due to the unique solution $\mathbf{x} = \mathbf{0}$ *using Gauss-Jordan we must be able to reduce the original matrix to the identity, note the right hand side will always be* $\mathbf{0}$*,*

$$\left[\begin{array}{cccc|c} \mathbf{1} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \cdots & \mathbf{0} & \mathbf{0} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1} & \mathbf{0} \end{array}\right].$$

If this was not the case, the Gauss-Jordan process at some stage would produce a row of zeros on the LHS. This says that there are less equations than unknowns. Hence, the solution to $\boldsymbol{A}\mathbf{x} = \mathbf{0}$ *would not have a unique solution.*

4. $(d) => (e)$

Assume that Gauss-Jordan elimination on A *produces* I_n*. Due to the properties of elementary matrices, there exists a finite sequence of elementary matrixs* $\{E_i\}_{i=1}^{k}$ *such that*

$$E_k E_{k-1} \cdots E_1 A = I_n$$

$$=> \;\; A = (E_k E_{k-1} \cdots E_1)^{-1} = E_1^{-1} E_2^{-1} \cdots E_k^{-1}.$$

Finally, the inverse of an elementary matrix is an elementary matrix and so

$$A = E_1^{'} E_2^{'} \cdots E_k^{'}.$$

5. $(e) => (a)$
Assuming $A = E_1 E_2 \cdots E_k$ *we immediately see that* A *is invertible, since each* E_i *is.*

Exercise 3.6.8 *Prove that an* n *by* n *matrix is invertible if and only if* $rank(A) = n$

We can summarise this section with the extension of the fundamental theorem of invertible matrices, Theorem 3.6.6:

Theorem 3.6.9 *The fundamental theorem of Invertible matrices: part II*
Let $\boldsymbol{A}$ be an $\boldsymbol{n}$ by $\boldsymbol{n}$ matrix. The following statements are equivalent:
(a) $\boldsymbol{A}$ is invertible.
(b) $\boldsymbol{A}\mathrm{x} = \mathrm{b}$ has a unique solution for every $\mathrm{b} \in \mathbb{R}^{\boldsymbol{n}}$.
(c) $\boldsymbol{A}\mathrm{x} = \mathbf{0}$ has only the trivial solution.
(d) Using Gauss-Jordan elimination we may reduce the matrix down to the identity.
(e) $\boldsymbol{A}$ is the product of elementary matrix.
(f) $\boldsymbol{rank}(\boldsymbol{A}) = \boldsymbol{n}$.
(g) $\boldsymbol{\mathcal{N}}(\boldsymbol{A}) = \{0\}$.
(h) The column vectors of $\boldsymbol{A}$ are linearly independent and so $\boldsymbol{span}$ $\mathbb{R}^{\boldsymbol{n}}$.
(i)The row vectors of $\boldsymbol{A}$ are linearly independent and so $\boldsymbol{span}$ $\mathbb{R}^{\boldsymbol{n}}$.