Chapter 1

Mathematical Preliminaries

1.1 Sets

A **set** (or **collection**) is specified by the **elements** (or **members**) that **belong** to it. If element x belongs to the set (does not belong to the set) X, we write $x \in X$ ($x \notin X$). Two sets are equal iff they have the same members. The set with no elements is called the empty set and is denoted by $\emptyset$. A set is **finite** if it has a finite number of elements. Otherwise it is **infinite**. A set is often specified by actually listing its members, e.g. $\{e_1, e_2, e_3\}$ is the set with members e_1, e_2, e_3. More usually it is specified by a property, e.g. the set of even numbers is specified as $\{x : x \text{ is an integer and } x \text{ is even }\}$ or as $\{x, x \text{ is an integer and } x \text{ is even }\}$. The symbols $\forall$ and $\exists$ are used to denote 'forall' and 'there exists'. Thus, '$\forall x$' or '$\forall\ x$' should be read as 'forall x' and '$\exists x$' should be read as 'there exists x'. A singleton set is one that has only one element. The singleton set with the element x as its only member, is denoted by $\{x\}$. In this book, very often, we abuse this notation and write x in place of $\{x\}$, if we feel that the context makes the intended object unambiguous.

We say that set X is **contained** in Y (**properly contained in** Y), if every element of X is also a member of Y (every element of X is a member of Y and $X \neq Y$) and denote it by $X \subseteq Y (X \subset Y)$.
The **union** of two sets X and Y denoted by $X \cup Y$, is the set whose members are either in X or in Y (or in both). The **intersection** of X and Y denoted by $X \cap Y$, is the set whose members belong both to X and to Y. When X and Y do not have common elements, they are said to be **disjoint**. Union of disjoint sets X and Y is often denoted by $X \uplus Y$. Union of sets $X_1, \cdots, X_n$ is denoted by $\bigcup_{i=1}^n X_i$ or simply by $\bigcup X_i$. When the X_i are pairwise disjoint, their union is denoted by $\biguplus_{i=1}^n X_i$ or $\biguplus X_i$.
The **difference** of X relative to Y, denoted by $X - Y$, is the set of all elements in X but not in Y. Let $X \subseteq S$. Then the **complement** of X **relative to** S is the set $S - X$ and is denoted by $\bar{X}$ when the set S is clear from the context.

A **mapping** $f : X \to Y$, denoted by $f(\cdot)$, associates with each element $x \in X$, the element $f(x)$ in Y. The element $f(x)$ is called the image of x under $f(\cdot)$. We say $f(\cdot)$ maps X into Y. The sets X, Y are called, respectively, the **domain** and **codomain** of $f(\cdot)$. We denote by $f(Z), Z \subseteq X$, the subset of Y which has as members, the images of elements in Z. The set $f(X)$ is called the **range** of $f(\cdot)$. The **restriction** of $f(\cdot)$ to $Z \subseteq X$, denoted by $f/Z(\cdot)$ is the mapping from Z to Y defined by $f/Z(x) \equiv f(x), x \in Z$. A mapping that has distinct images for distinct elements in the domain is said to be **one to one** or **injective**. If the range of $f(\cdot)$ equals its codomain, we say that $f(\cdot)$ is **onto** or **surjective**. If the mapping is one to one onto we say it is **bijective**. Let $f : X \to Y, g : Y \to Z$. Then the **composition** of g and f is the map, denoted by $gf(\cdot)$ or $g \circ f(\cdot)$, defined by $gf(x) \equiv g(f(x)) \ \ \forall x \in X$. The **Cartesian product** $X \times Y$ of sets X, Y is the collection of all ordered pairs (x, y), where $x \in X$ and $y \in Y$. The **direct sum** $X \oplus Y$ denotes the union of disjoint sets X, Y. We use '**direct sum**' loosely to indicate that structures on two disjoint sets are 'put together'. We give some examples where we anticipate definitions which would be given later. The **direct sum** of vector spaces $\mathcal{V}_1, \mathcal{V}_2$ on disjoint sets S_1, S_2 is the vector space $\mathcal{V}_1 \oplus \mathcal{V}_2$ on $S_1 \oplus S_2$ whose typical vectors are obtained by taking a vector $\mathbf{x}_1 \equiv (a_1, \cdots, a_k)$ in $\mathcal{V}_1$ and a vector $\mathbf{x}_2 \equiv (b_1, \cdots, b_m)$ in $\mathcal{V}_2$ and putting them together as $\mathbf{x}_1 \oplus \mathbf{x}_2 \equiv (a_1, \cdots, a_k, b_1, \cdots, b_m)$. When we have two graphs $\mathcal{G}_1, \mathcal{G}_2$ on disjoint edge sets E_1, E_2, $\mathcal{G}_1 \oplus \mathcal{G}_2$ would have edge set $E_1 \oplus E_2$ and is obtained by 'putting together' $\mathcal{G}_1$ and $\mathcal{G}_2$. Usually the vertex sets would also be disjoint. However, where the context permits, we may relax the latter assumption and allow 'hinging' of vertices.

We speak of a **family** of subsets as distinct from a **collection** of subsets. The collection $\{\{e_1, e_2\}, \{e_1, e_2\}, \{e_1\}\}$ is identical to $\{\{e_1, e_2\}, \{e_1\}\}$. But often (e.g. in the definition of a hypergraph in Subsection **??**) we have to use copies of the same subset many times and distinguish between copies. This we do by 'indexing' them. A family of subsets of S may be defined to be a mapping from an index set I to the collection of all subsets of S. For the purpose of this book, the index set I can be taken to be $\{1, \cdots, n\}$. So the family $(\{e_1, e_2\}, \{e_1, e_2\}, \{e_1\})$ can be thought of as the mapping $\phi(\cdot)$ with

$$\begin{aligned} \phi(1) &\equiv \{e_1, e_2\} \\ \phi(2) &\equiv \{e_1, e_2\} \\ \phi(3) &\equiv \{e_1\}. \end{aligned}$$

(Note that a family is denoted using ordinary brackets while a set is denoted using curly brackets).

1.2 Vectors and Matrices

In this section we define vectors, matrices and related notions. Most present day books on linear algebra treat vectors as primitive elements in a vector space and leave them undefined. We adopt a more old fashioned approach which is convenient for the applications we have in mind. The reader who wants a more leisurely treatment of the topics in this section is referred to [Hoffman+Kunze72].

Let S be a finite set $\{e_1, e_2, \ldots, e_n\}$ and let $\mathcal{F}$ be a field. We will confine ourselves to the field $\Re$ of real numbers, the field $\mathcal{C}$ of complex numbers and the $GF2$ field on elements $0, 1$ $(0+0=0, 0+1=1, 1+0=1, 1+1=0, 1.1=1, 1.0=0, 0.1=0, 0.0=0)$. For a general definition of a field see for instance [Jacobson74]. By a vector on S over $\mathcal{F}$ we mean a mapping $\mathbf{f}$ of S into $\mathcal{F}$.The field $\mathcal{F}$ is called the **scalar field** and its elements are called **scalars**. The **support** of $\mathbf{f}$ is the subset of S over which it takes nonzero values. The **sum** of two vectors $\mathbf{f}, \mathbf{g}$ on S over $\mathcal{F}$ is defined by $(\mathbf{f}+\mathbf{g})(e_i) \equiv \mathbf{f}(e_i) + \mathbf{g}(e_i)\ \forall e_i \in S$. (For convenience the **sum** of two vectors $\mathbf{f}$ on S,$\mathbf{g}$ on T over $\mathcal{F}$ is defined by $(\mathbf{f}+\mathbf{g})(e_i) \equiv \mathbf{f}(e_i)+\mathbf{g}(e_i)\ \forall e_i \in S \cap T$, as agreeing with $\mathbf{f}$ on $S-T$, and as agreeing with $\mathbf{g}$ on $T-S$). The **scalar product** of $\mathbf{f}$ by a scalar λ is a vector $\lambda\mathbf{f}$ defined by $(\lambda\mathbf{f})(e_i) \equiv \lambda(\mathbf{f}(e_i))\ \forall e_i \in S$. A collection $\mathcal{V}$ of vectors on S over $\mathcal{F}$ is a vector space iff it is closed under addition and scalar product. We henceforth omit mention of underlying set and field unless required.

A set of vectors $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n\}$ is **linearly dependent** iff there exist scalars $\lambda_1, \ldots, \lambda_n$ not all zero such that $\lambda_1\mathbf{f}_1 + \ldots + \lambda_n\mathbf{f}_n = \mathbf{0}$. (Here the $\mathbf{0}$ vector is one which takes value 0 on all elements of S). Vector $\mathbf{f}_n$ is a linear combination of $\mathbf{f}_1, \ldots, \mathbf{f}_{n-1}$ iff $\mathbf{f}_n = \lambda_1\mathbf{f}_1 + \ldots + \lambda_{n-1}\mathbf{f}_{n-1}$ for some $\lambda_1, \ldots, \lambda_{n-1}$.

The set of all vectors linearly dependent on a collection $\mathcal{C}$ of vectors can be shown to form a vector space which is said to be **generated** by or **spanned** by $\mathcal{C}$. Clearly if $\mathcal{V}$ is a vector space and $\mathcal{C} \subseteq \mathcal{V}$, the subset of vectors generated by $\mathcal{C}$ is contained in $\mathcal{V}$. A maximal linearly independent set of vectors of $\mathcal{V}$ is called a **basis** of $\mathcal{V}$.
In general maximal and minimal members of a collection of sets may not be largest and smallest in terms of size.
Example: Consider the collection of sets $\{\{1,2,3\}, \{4\}, \{5,6\}, \{1,2,3,5,6\}\}$. The minimal members of this collection are $\{1,2,3\}, \{4\}, \{5,6\}$, i.e., these do not contain proper subsets which are members of this collection. The maximal members of this collection are $\{4\}, \{1,2,3,5,6\}$, i.e., these are not proper subsets of other sets which are members of this collection.

The following theorem is therefore remarkable.

Theorem 1.2.1 *All bases of a vector space on a finite set have the same cardinality.*

The number of elements in a basis of $\mathcal{V}$ is called the **dimension** of $\mathcal{V}$, denoted by $dim(\mathcal{V})$, or the **rank** of $\mathcal{V}$, denoted by $r(\mathcal{V})$. Using Theorem 1.2.1 one can show that the size of a maximal independent subset contained in a given set $\mathcal{C}$ of vectors is unique. This number is called the **rank** of $\mathcal{C}$. Equivalently, the rank of $\mathcal{C}$ is the dimension of the vector space spanned by $\mathcal{C}$. If $\mathcal{V}_1, \mathcal{V}_2$ are vector spaces and $\mathcal{V}_1 \subseteq \mathcal{V}_2$, we say $\mathcal{V}_1$ is a **subspace** of $\mathcal{V}_2$.

A mapping $\mathbf{A} : \{1, 2, \ldots, m\} \times \{1, 2, \ldots, n\} \longrightarrow \mathcal{F}$ is called a $m \times n$ matrix. It may be thought of as an $m \times n$ array with entries from $\mathcal{F}$. We denote $\mathbf{A}(i,j)$ often by the lower case a_{ij} with i as the row index and j as the column index. We speak of the array $(a_{i1}, \ldots, a_{in})$ as the **ith row** of $\mathbf{A}$ and of the array $(a_{1j}, \ldots, a_{nj})$ as the **jth column** of $\mathbf{A}$. Thus we may think of $\mathbf{A}$ as made up of m row vectors or of n column vectors. Linear dependence, independence and linear combination for row and column vectors are defined the same way as for vectors. We say two matrices are **row equivalent** if the rows of each can be obtained by linearly combining the rows of the other. **Column equivalence** is defined similarly. The vector space spanned by the rows (columns) of $\mathbf{A}$ is called its **row space (column space)** and denoted by $\mathcal{R}(\mathbf{A})(\mathcal{C}(\mathbf{A}))$. The dimension of $\mathcal{R}(\mathbf{A})(\mathcal{C}(\mathbf{A}))$ is called the **row rank (column rank)** of $\mathbf{A}$.

If $\mathbf{A}$ is an $m \times n$ matrix then the **transpose** of $\mathbf{A}$ denoted by $\mathbf{A}^T$ is an $n \times m$ matrix defined by $\mathbf{A}^T(i,j) \equiv \mathbf{A}(j,i)$. Clearly the i^{th} row of $\mathbf{A}$ becomes the i^{th} column of $\mathbf{A}^T$. If $\mathbf{B}$ is also an $m \times n$ matrix the **sum** $\mathbf{A}+\mathbf{B}$ is an $m \times n$ matrix defined by $(\mathbf{A}+\mathbf{B})(i,j) \equiv \mathbf{A}(i,j)+\mathbf{B}(i,j)$. If $\mathbf{D}$ is an $n \times p$ matrix, the product $\mathbf{AD}$ is an $m \times p$ matrix defined by $\mathbf{AD}(i,j) \equiv \sum_{k=1}^{n} a_{ik}d_{kj}$. Clearly if $\mathbf{AD}$ is defined it does not follow that $\mathbf{DA}$ is defined. Even when it is defined, in general $\mathbf{AD} \neq \mathbf{DA}$. The most basic property of this notion of product is that it is **associative** i.e. $\mathbf{A}(\mathbf{DF}) = (\mathbf{AD})\mathbf{F}$.

Matrix operations are often specified by **partitioning**. Here we write a matrix in terms of **submatrices** (i.e., matrices obtained by deleting some rows and columns of the original matrix). A matrix may be partitioned along rows:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} \\ \cdots \\ \mathbf{A}_{21} \end{bmatrix}$$

or along columns:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} | \mathbf{A}_{12} \end{bmatrix}$$

or both:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \cdots & \mathbf{A}_{1k} \\ \vdots & \vdots & \vdots \\ \mathbf{A}_{p1} & \cdots & \mathbf{A}_{pk} \end{bmatrix}.$$

When two partitioned matrices are multiplied we assume that the partitioning is **compatible**, i.e., for each triple (i,j,k) the number of columns of $\mathbf{A}_{ik}$ equals the number of rows of $\mathbf{B}_{kj}$. Clearly this is achieved if the original matrices $\mathbf{A}, \mathbf{B}$ are compatible for product and each block of the column partition of $\mathbf{A}$ has the same size as the corresponding row partition of $\mathbf{B}$. The following **partitioning rules** can then be verified.

i. $\begin{bmatrix} \mathbf{A}_{11} \\ \cdots \\ \mathbf{A}_{21} \end{bmatrix} \mathbf{C} = \begin{bmatrix} \mathbf{A}_{11}\mathbf{C} \\ \cdots \\ \mathbf{A}_{21}\mathbf{C} \end{bmatrix}$

ii. $\mathbf{C} \begin{bmatrix} \mathbf{A}_{11} | \mathbf{A}_{12} \end{bmatrix} = \begin{bmatrix} \mathbf{C}\mathbf{A}_{11} | \mathbf{C}\mathbf{A}_{12} \end{bmatrix}$

iii. $\left[\mathbf{A}_{11} | \mathbf{A}_{12} \right] \begin{bmatrix} \mathbf{C}_{11} \\ \cdots \\ \mathbf{C}_{12} \end{bmatrix} = \left[\mathbf{A}_{11}\mathbf{C}_{11} + \mathbf{A}_{12}\mathbf{C}_{12} \right].$

In general if $\mathbf{A}$ is partitioned into submatrices $\mathbf{A}_{ik}, \mathbf{B}$ into submatrices $\mathbf{B}_{kj}$ then the product $\mathbf{C} = \mathbf{AB}$ would be naturally partitioned into $\mathbf{C}_{ij} \equiv \sum_k \mathbf{A}_{ik}\mathbf{B}_{kj}$.

Matrices arise most naturally in linear equations such as $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ and $\mathbf{b}$ are known and $\mathbf{x}$ is an unknown vector. When $\mathbf{b} = \mathbf{0}$ it is easily verified that the set of all solutions of $\mathbf{Ax} = \mathbf{b}$, i.e.,of $\mathbf{Ax} = \mathbf{0}$, form a vector space. This space will be called the **solution space** of $\mathbf{Ax} = \mathbf{0}$, or the **null space** of $\mathbf{A}$. The **nullity** of $\mathbf{A}$ is the dimension of the null space of $\mathbf{A}$. We have the following theorem.

Theorem 1.2.2 *If two matrices are row equivalent then their null spaces are identical.*

Corollary 1.2.1 *If* $\mathbf{A}, \mathbf{B}$ *are row equivalent matrices then a set of columns of* $\mathbf{A}$ *are independent iff the corresponding set of columns of* $\mathbf{B}$ *are independent.*

The following are **elementary row operations** that can be performed on the rows of a matrix:

i. interchanging rows,

ii. adding a multiple of one row to another,

iii. multiplying a row by a nonzero number.

Each of these operations corresponds to premultiplication by a matrix. Such matrices are called **elementary matrices**. It can be seen that these are the matrices we obtain by performing the corresponding elementary row operations on the unit matrix of the same number of rows as the given matrix. We can define elementary column operations similarly. These would correspond to post multiplication by elementary column matrices.

A matrix is said to be in **Row Reduced Echelon** form (RRE) iff it satisfies the following:
Let r be the largest row index for which $a_{ij} \neq 0$ for some j. Then the columns of the $r \times r$ unit matrix (the matrix with 1s along the diagonal and zero elsewhere) $\mathbf{e_1}, \ldots, \mathbf{e_r}$ appear as columns, say $\mathbf{C}_{i_1}, \ldots, \mathbf{C}_{i_r}$ of $\mathbf{A}$ with $i_1 < \ldots < i_r$. Further if $p < i_k$ then $a_{kp} = 0$. We have the following theorem.

Theorem 1.2.3 *Every matrix can be reduced to a matrix in the RRE form by a sequence of elementary row transformations and is therefore row equivalent to such a matrix.*

It is easily verified that for an RRE matrix row rank equals column rank. Hence using Theorem 1.2.3 and Corollary 1.2.1 we have

Theorem 1.2.4 *For any matrix, row rank equals column rank.*

The **rank** of a matrix $\mathbf{A}$, denoted by $r(\mathbf{A})$, is its row rank (= column rank).
Let the elements of S be ordered as $(e_1, \ldots, e_n)$. Then for any vector $\mathbf{f}$ on S we define $\mathbf{R}_f$, the **representative vector** of $\mathbf{f}$, as the one rowed matrix $(\mathbf{f}(e_1), \ldots, \mathbf{f}(e_n))$. We will not usually distinguish between a vector and its representative vector. When the rows of a matrix $\mathbf{R}$ are representative vectors of some basis of a vector space $\mathcal{V}$ we say that $\mathbf{R}$ is a **representative matrix** of $\mathcal{V}$. When $\mathbf{R}, \mathbf{R}_1$ both represent $\mathcal{V}$ they can be obtained from each other by row operations. Hence by Corollary 1.2.1 their column independence structure is identical. An $r \times n$ representative matrix $\mathbf{R}$, $r \leq n$, is a **standard representative matrix** iff $\mathbf{R}$ has an $r \times r$ submatrix which can be obtained by permutations of the columns of the $r \times r$ unit matrix. For convenience we will write a standard representative matrix in the form $[\mathbf{I}|\mathbf{R}_{12}]$ or $[\mathbf{R}_{11}|\mathbf{I}]$. (Here $\mathbf{I}$ denotes the unit matrix of appropriate size).

The **dot product** of two vectors $\mathbf{f}, \mathbf{g}$ on S denoted by $< \mathbf{f}, \mathbf{g} >$ over $\mathcal{F}$ is defined by $< \mathbf{f}, \mathbf{g} > \equiv \sum_{e \in S} \mathbf{f}(e).\mathbf{g}(e)$. We say $\mathbf{f}, \mathbf{g}$ are **orthogonal** if their dot product is zero. If $\mathcal{C}$ is a collection of vectors on S then $\mathcal{C}^{\perp} \equiv$ set of all vectors orthogonal to every vector in $\mathcal{C}$. It can be verified that $\mathcal{C}^{\perp}$ is a vector space. Let $\mathcal{V}$ be a vector space on S with basis $\mathcal{B}$. Since vectors orthogonal to each vector in $\mathcal{B}$ are also orthogonal to linear combinations of these vectors we have $\mathcal{B}^{\perp} = \mathcal{V}^{\perp}$. If $\mathbf{R}$ is a representative matrix of $\mathcal{V}$, it is clear that $\mathcal{V}^{\perp}$ is its null space. Equivalently $\mathcal{V}^{\perp}$ is the solution space of $\mathbf{Rx} = \mathbf{0}$. If $\mathbf{R}$ is a standard representative matrix with $\mathbf{R} = [\mathbf{I}_{r \times r}|\mathbf{R}_{12}]$, then the solution space of $\mathbf{Rx} = \mathbf{0}$ can be shown to be the vector space generated by the columns of $\begin{bmatrix} -\mathbf{R}_{12} \\ \cdots \\ \mathbf{I}_{n-r \times n-r} \end{bmatrix}$, where $n = |S|$.(Here $\mathbf{I}_{k \times k}$ denotes the unit matrix with k rows). Equivalently $\mathcal{V}^{\perp}$ has the representative matrix $[-\mathbf{R}_{12}^T|\mathbf{I}_{n-r \times n-r}]$. The representative matrix of $(\mathcal{V}^{\perp})^{\perp}$ will then be $\mathbf{R}$. We then have the following

Theorem 1.2.5 *i. if* $[\mathbf{I}_{r \times r}|\mathbf{R}_{12}]$ *is a representative matrix of vector space* $\mathcal{V}$ *on* S *then* $[-\mathbf{R}_{12}^T|\mathbf{I}_{n-r \times n-r}]$ *is a representative matrix of* $\mathcal{V}^{\perp}$.

ii. $r(\mathcal{V}^{\perp}) = |S| - r(\mathcal{V})$

iii. $(\mathcal{V}^{\perp})^{\perp} = \mathcal{V}$. *Hence two matrices are row equivalent iff their null spaces are identical.*

Consider the collection of $n \times n$ matrices over $\mathcal{F}$. We say that $\mathbf{I}$ is an identity for this collection iff for every $n \times n$ matrix $\mathbf{B}$ we have $\mathbf{IB} = \mathbf{BI} = \mathbf{B}$. If $\mathbf{I}_1, \mathbf{I}_2$ are identity matrices we must have $\mathbf{I}_1 = \mathbf{I}_2 = \mathbf{I}$. The unit matrix (with $1s$ along the diagonal and $0s$ elsewhere) is clearly an identity matrix. It is therefore the only identity matrix. Two $n \times n$ matrices $\mathbf{A}, \mathbf{B}$ are **inverses** of each other iff $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$. We say $\mathbf{A}, \mathbf{B}$ are **invertible** or **nonsingular**. If $\mathbf{A}$ has inverses $\mathbf{B}, \mathbf{C}$ we must have $\mathbf{C} = \mathbf{C}(\mathbf{AB}) = (\mathbf{CA})\mathbf{B} = \mathbf{IB} = \mathbf{B}$. Thus the inverse of a matrix $\mathbf{A}$, if it exists, is unique and is denoted by $\mathbf{A}^{-1}$. We then have the following

Theorem 1.2.6 *i.* $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$

ii. If $\mathbf{A}, \mathbf{D}$ *are* $n \times n$ *invertible matrices, then* $(\mathbf{AD})^{-1} = (\mathbf{D}^{-1}\mathbf{A}^{-1})$.

With a square matrix we associate an important number called its determinant. Its definition requires some preparation.

A bijection of a finite set to itself is also called a **permutation**. A permutation that interchanges two elements (i.e. maps each of them to the other) but leaves all others unchanged is a **transposition**. Every permutation can be obtained by repeated application of transpositions. We then have the following

Theorem 1.2.7 *If a permutation σ can be obtained by composition of an even number of transpositions then every decomposition of σ into transpositions will contain an even number of them.*

By Theorem 1.2.7 we can define a permutation to be **even (odd)** iff it can be decomposed into an even (odd) number of transpositions. The **sign** of a permutation σ denoted by $sgn(\sigma)$ is $+1$ if σ is even and -1 if σ is odd. It is easily seen, since the identity ($= \sigma\sigma^{-1}$) permutation is even, that $sgn(\sigma) = sgn(\sigma^{-1})$. The **determinant** of an $n \times n$ matrix is defined by

$$det(\mathbf{A}) \equiv \sum_{\sigma} sgn(\sigma) a_{1\sigma(1)} \dots a_{n\sigma(n)},$$

where the summation is taken over all possible permutations of $\{1, 2, \dots, n\}$. It is easily seen that determinant of the unit matrix is $+1$. We collect some of the important properties of the determinant in the following

Theorem 1.2.8 *i.* $det(\mathbf{A}) = det(\mathbf{A}^T)$

ii. *Let*

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{A}_2 \end{bmatrix}, \mathbf{A}' = \begin{bmatrix} \mathbf{a}'_1 \\ \mathbf{A}_2 \end{bmatrix}, \mathbf{A}" = \begin{bmatrix} \mathbf{a}_1 + \mathbf{a}'_1 \\ \mathbf{A}_2 \end{bmatrix}.$$

Then $det(\mathbf{A}") = det(\mathbf{A}) + det(\mathbf{A}')$.

iii. *If* $\mathbf{A}$ *has two identical rows, or has two identical columns then* $det(\mathbf{A}) = 0$.

iv. *If* $\mathbf{E}$ *is an elementary matrix* $det(\mathbf{EA}) = det(\mathbf{E})det(\mathbf{A})$. *Since every invertible matrix can be factored into elementary matrices, it follows that* $det(\mathbf{AB}) = det(\mathbf{A})det(\mathbf{B})$, *for every pair of* $n \times n$ *matrices* $\mathbf{A}, \mathbf{B}$.

v. $det(\mathbf{A}) \neq 0$ *iff* $\mathbf{A}$ *is invertible.*

Problem 1.1 Size of a basis: *Prove*

i. *Theorem 1.2.1*

ii. *If* $\mathcal{V}_1$ *is a subspace of vector space* $\mathcal{V}_2$, $\dim \mathcal{V}_1 \leq \dim \mathcal{V}_2$.

iii. *If* $\mathcal{V}_1 \subseteq \mathcal{V}_2$ *and* $\dim \mathcal{V}_1 = \dim \mathcal{V}_2$ *then* $\mathcal{V}_1 = \mathcal{V}_2$.

iv. *an* $m \times n$ *matrix with* $m > n$ *cannot have linearly independent rows.*

v. *any vector in a vector space* $\mathcal{V}$ *can be written uniquely as a linear combination of the vectors in a basis of* $\mathcal{V}$.

Problem 1.2 Ways of interpreting the matrix product: *Define product of matrices in the usual way i.e.* $\mathbf{C} = \mathbf{AB}$ *is equivalent to* $\mathbf{C}_{ij} = \sum_k a_{ik} b_{kj}$. *Now show that it can be thought of as follows*

i. *Columns of* $\mathbf{C}$ *are linear combinations of columns of* $\mathbf{A}$ *using entries of columns of* $\mathbf{B}$ *as coefficients.*

ii. *rows of* $\mathbf{C}$ *are linear combinations of rows of* $\mathbf{B}$ *using entries of rows of* $\mathbf{A}$ *as coefficients.*

Problem 1.3 Properties of matrix product: *Prove, when* $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ *are matrices and the products are defined*

i. $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$

ii. $(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$

Problem 1.4 Partitioning rules: *Prove*

i. *the partitioning rules.*

ii.

$$\begin{bmatrix} \mathbf{A}_{11} \cdots & \mathbf{A}_{1k} \\ \vdots & \vdots \\ \mathbf{A}_{r1} \cdots & \mathbf{A}_{rk} \end{bmatrix}^T = \begin{bmatrix} \mathbf{A}_{11}^T \cdots & \mathbf{A}_{r1}^T \\ \vdots & \vdots \\ \mathbf{A}_{1k}^T \cdots & \mathbf{A}_{rk}^T \end{bmatrix}$$

Problem 1.5 Solution space and column dependence structure: *Prove theorem 1.2.2 and Corollary 1.2.1.*

Problem 1.6 Algorithm for computing RRE: *Give an algorithm for converting any rectangular matrix into the RRE form. Give an upper bound for the number of arithmetical steps in your algorithm.*

Problem 1.7 Uniqueness of the RRE matrix: *Show that no RRE matrix is row equivalent to a distinct RRE matrix. Hence prove that every matrix is row equivalent to a unique matrix in the RRE form.*

Problem 1.8 RRE of special matrices:

i. *If* $\mathbf{A}$ *is a matrix with linearly independent columns what is its RRE form? If in addition* $\mathbf{A}$ *is square what is its RRE form?*

ii. *If* $\mathbf{A}, \mathbf{B}$ *are square such that* $\mathbf{AB} = \mathbf{I}$ *show that* $\mathbf{BA} = \mathbf{I}$.

iii. *Prove Theorem 1.2.6*

Problem 1.9 Existence and nature of solution for linear equations: *Consider the equation* $\mathbf{Ax} = \mathbf{b}$.

i. *Show that it has a solution*

(a) *iff* $r(\mathbf{A}) = r(\mathbf{A}|\mathbf{b})$.

(b) *iff whenever* $\lambda^T\mathbf{A} = \mathbf{0}, \lambda^T\mathbf{b}$ *is also zero.*

ii. *Show that a vector is a solution of the above equation iff it can be written in the form* $\mathbf{x}_o + \mathbf{x}_p$ *where* $\mathbf{x}_p$ *is a particular solution of the equation while* $\mathbf{x}_o$ *is a vector in the null space of* $\mathbf{A}$ *(i.e. a solution to the linear equation with* $\mathbf{b}$ *set equal to zero).*

iii. **Motivation for the matrix product:** *Why is the matrix product defined as in Problem 1.2? (In the above equation suppose we make the substitution* $\mathbf{x} = \mathbf{B}\mathbf{y}$*. What would the linear equation in terms of* $\mathbf{y}$ *be?)*

iv. **Linear dependence and logical consequence:** *The above equation may be regarded as a set of linear equations (one for each row of* $\mathbf{A}$*) each of which in turn could be thought of as a statement. Show that a linear equation is a logical consequence of others iff it is* **linearly** *dependent on the others.*

Problem 1.10 Positive definite matrices:

i. *Construct an example where* $\mathbf{A}$, $\mathbf{B}$ *are invertible but their sum is not.*

ii. *A matrix* $\mathbf{K}$ *is* **positive semidefinite (positive definite)** *iff* $\mathbf{x}^T\mathbf{K}\mathbf{x} \geq 0 \ \ \forall \mathbf{x} \neq 0$ *(*$\mathbf{x}^T\mathbf{K}\mathbf{x} > 0 \ \ \forall \mathbf{x} \neq 0$*). Show that*

(a) *a matrix is invertible if it is positive definite;*

(b) *sum of two positive semidefinite matrices (positive definite matrices) is positive semidefinite (positive definite);*

(c) *if* $\mathbf{K}$ *is a positive definite matrix,then* $\mathbf{AKA}^T$ *is positive semidefinite and if, further, rows of* $\mathbf{A}$ *are linearly independent, then* $\mathbf{AKA}^T$ *is positive definite;*

(d) *inverse of a symmetric positive definite matrix is also symmetric positive definite.*

Problem 1.11 Projection of a vector on a vector space: *Let* $\mathbf{x}$ *be a vector on* S *and let* $\mathcal{V}$ *be a vector space on* S*. Show that* $\mathbf{x}$ *can be uniquely decomposed as* $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$*, where* $\mathbf{x}_1 \in \mathcal{V}$ *and* $\mathbf{x}_2 \in \mathcal{V}^\perp$*.The vector* $\mathbf{x}_1$ *is called the* **projection** *of* $\mathbf{x}$ *on* $\mathcal{V}$ **along** $\mathcal{V}^\perp$*.*

Problem 1.12 Parity of a Permutation: *Show that if a permutation can be obtained by composing an odd number of transpositions it cannot also be obtained by composing an even number of transpositions.*

Problem 1.13 Graph of a permutation: *Define the graph* $\mathcal{G}_\sigma$ *of a permutation* σ *on* $\{1, 2, \cdots n\}$ *as follows:* $V(\mathcal{G}_\sigma) \equiv \{1, 2, \cdots, n\}$*; draw an edge with an arrow from* i *to* j *iff* $\sigma(i) = j$*.*

i. *Show that every vertex in this graph has precisely one arrow coming in and one going out. Hence, conclude that each connected component is a directed circuit.*

ii. Show that if $\mathcal{G}_\sigma$ has an odd (even) number of even length circuits then σ is odd (even).

Problem 1.14 Properties of the determinant: *Prove Theorem 1.2.8.*

Problem 1.15 Equivalence of definitions of a determinant: *Show that the usual definition of a determinant by expanding along a row or column is equivalent to the definition using permutations.*

Problem 1.16 Laplace expansion of the determinant: *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix. Show that*

$$det(\mathbf{A}) = \sum sgn(\sigma)\ det\left(\mathbf{A}\begin{pmatrix} r_1, & \cdots & , r_k \\ i_1, & \cdots & , i_k \end{pmatrix}\right) det\left(\mathbf{A}\begin{pmatrix} r_{k+1}, & \cdots & , r_m \\ i_{k+1}, & \cdots & , i_m \end{pmatrix}\right),$$

($\mathbf{A}\begin{pmatrix} d_1, & \cdots & , d_p \\ i_1, & \cdots & , i_p \end{pmatrix}$ *is the* $p \times p$ *matrix whose* (s,t) *entry is the* (d_s, i_t) *entry of* $\mathbf{A}$*), where the summation is over all subsets* $\{i_1, \cdots, i_k\}$ *of* $\{1, \cdots, n\}$
and $\sigma \equiv \begin{pmatrix} i_1, \cdots, i_k, i_{k+1} \cdots i_n \\ r_1, \cdots, r_k, r_{k+1} \cdots r_n \end{pmatrix}$ *i.e.,* $\sigma(i_j) \equiv r_j, j = 1, \cdots, n.$

Problem 1.17 Binet Cauchy Theorem: *Let* $\mathbf{A}$ *be an* $m \times n$ *and* $\mathbf{B}$ *an* $n \times m$ *matrix with* $m \leq n$*. If an* $m \times m$ *submatrix of* $\mathbf{A}$ *is composed of columns* $i_1, \cdots, i_m$*, the* **corresponding** $m \times m$ *submatrix of* $\mathbf{B}$ *is the one with rows* $i_1, \cdots, i_m$*. Prove the Binet Cauchy Theorem:* $det(\mathbf{AB}) = \sum$ *product of determinants of corresponding* $m \times m$ *submatrices of* $\mathbf{A}$ *and* $\mathbf{B}$.

1.3 Linear Inequality Systems

1.3.1 The Kuhn-Fourier Theorem

In this section we summarize basic results on inequality systems which we need later on in the book. Proofs are mostly omitted. They may be found in standard references such as [Stoer+Witzgall70] and [Schrijver86]. This section follows the former reference.

A **linear inequality system** is a set of constraints of the following kind on the vector $\mathbf{x} \in \Re^n$.

$$\left.\begin{array}{lcl} \mathbf{Ax} & = & \mathbf{a}_o \\ \mathbf{Bx} & > & \mathbf{b}_o \\ \mathbf{Cx} & \geq & \mathbf{c}_o \end{array}\right\} \qquad (I)$$

Here, $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are matrices, $\mathbf{a}_o, \mathbf{b}_o, \mathbf{c}_o$ are column vectors with appropriate number of rows. We say $\mathbf{x}_1 > \mathbf{x}_2 (\mathbf{x}_1 \geq \mathbf{x}_2)$ iff each component of $\mathbf{x}_1$ is greater than (greater than or equal to) the corresponding component of $\mathbf{x}_2$.

A **solution** of an inequality system is a vector which satisfies all the inequality constraints of the system. A constraint which is satisfied by every solution of an inequality system is said to be a **consequence** of the system. In particular, we are concerned with constraints of the kind $\mathbf{d}^T\mathbf{x} = \mathbf{d}_o$ or $> \mathbf{d}_o$ or $\geq \mathbf{d}_o$. A **legal linear combination** of the system (I) is obtained by linearly combining the equations and inequalities with real coefficients - α_i for the linear equations, and non-negative real coefficients β_j, γ_k for the '$>$' linear inequalities and '$\geq$' linear inequalities respectively. The resulting constraint would be a linear equation iff β_j, γ_k are all zero. It would be a '$>$' inequality iff at least one of the β_j's is nonzero. It would be a '$\geq$' inequality iff all of β_j are zero but at least one of the γ_k is nonzero. A legal linear combination is thus a consequence of the system. A legal linear combination, with at least one of the $\alpha_i, \beta_j, \gamma_k$ nonzero, that results in the LHS becoming zero is called a **legal linear dependence** of the system. Another important way of deriving consequence relations is by **weakening**. This means to weaken '$=$' to '$\geq$' and '$>$' to '$\geq$' and also in the case of '$>$' and '$\geq$' to lower the right side value.

Example 1.3.1 *Consider the system of linear inequalities:*

$$\begin{aligned} x_1 + 2x_2 &= 3 \\ 2x_1 + x_2 &= 4 \\ x_1 + x_2 &> 1 \\ 2x_1 + 3x_2 &> 2 \\ x_1 + 5x_2 &\geq 2 \\ -x_1 - 2x_2 &\geq 4. \end{aligned}$$

The legal linear combination corresponding to $\alpha_1 = 1, \alpha_2 = 1, \beta_1 = 0, \beta_2 = 0, \gamma_1 = 0, \gamma_2 = 0$ is
$3x_1 + 3x_2 = 7$;
that corresponding to $\alpha_1 = 1, \alpha_2 = 0, \beta_1 = 1, \beta_2 = 0, \gamma_1 = 1, \gamma_2 = 0$ is
$3x_1 + 8x_2 > 6$;
that corresponding to $\alpha_1 = 1, \alpha_2 = 0, \beta_1 = 0, \beta_2 = 0, \gamma_1 = 1, \gamma_2 = 0$ is
$2x_1 + x_2 \geq 5$.
The legal linear combination corresponding to $\alpha_1 = 1, \alpha_2 = 0, \beta_1 = 0, \beta_2 = 0, \gamma_1 = 0, \gamma_2 = 1$ is the zero relation
$0x_1 + 0x_2 \geq 7$.
Thus in this case, the system has a **legal linear dependence** that is a **contradiction**.

We can now state the fundamental theorem of Kuhn and Fourier [Fourier1826], [Kuhn56].

Theorem 1.3.1 (Kuhn-Fourier Theorem) *A linear inequality system has a solution iff no legal linear dependence is a contradiction.*

Sketch of the Proof of Theorem 1.3.1: First reduce the linear equations to the RRE form and express some of the variables in terms of the others. This substitution is now carried out also in the inequalities. So henceforth, without loss of generality, we may assume that we have only inequalities. If we prove the theorem for such a reduced system, it can be extended to one which has equalities also.

Suppose all variables have either zero coefficient or the same sign in all the inequalities. In this case it is easy to see that the system has a solution whether the coefficients are all zero or otherwise. If all the coefficients are zero we are done - the theorem is clearly true. If not, it is not possible to get a legal linear dependence without using zero coefficents. So the theorem is again true in this case.

We now present an elimination procedure which terminates at one of the above mentioned cases.
Let the inequalities be numbered $(1), \cdots, (r), (r+1), \cdots, (k)$. Let x_n be present with coefficient +1 in the inequalities $(1), \cdots, (r)$ and with coefficient -1 in the inequalities $(r+1), \cdots, (k)$. We create $r(k-r)$ inequalities without the variables x_n by adding each of the first r inequalities to each of the last $(k-r)$ inequalities. Note that if both the inequalities are of the $(\geq)$ kind, the addition would result in another of the $(\geq)$ kind and if one of them is of the $(>)$ kind, the addition would result in another of the $(>)$ kind.

If the original system has a solution, it is clear that the reduced system also has one. On the other hand, if the reduced system has a solution $(x'_1, \cdots, x'_{n-1})$ it is possible to find a value x'_n of x_n such that $(x'_1, \cdots, x'_{n-1}, x'_n)$ is a solution of the original system. We indicate how, below.
Let the inequalities added be

$$a_{i1}x_1 + \cdots + x_n \geq b_i$$

$$a_{j1}x_1 + \cdots - x_n > b_j$$

(The cases where both are $(\geq)$, both are $(>)$ or first inequality $(>)$ and second $(\geq)$ are similar.) The pair of inequalities can be written equivalently as

$$a_{j1}x_1 + \cdots + a_{j(n-1)}x_{n-1} - b_j > x_n \geq b_i - a_{i1}x_1 - \cdots - a_{i(n-1)}x_{n-1} \qquad (*)$$

The extreme left of the above inequality $(*)$ is always derived from the inequalities $(r+1)$ to (k) while the extreme right is always derived from the (1) to (r) inequalities. When $x'_1, \cdots, x'_{n-1}$ is substituted in the above inequality, it would be satisfied for every pair of inequalities, from $(j+1)$ to (k) on the extreme left and (1) to (j) on the extreme right. After substitution, let the least of the extreme left term be reached for inequality (p) and let the highest of the extreme right term be reached for inequality (q). Since

$$a_{p1}x'_1 + \cdots + a_{p(n-1)}x'_{n-1} - b_p > b_q - a_{q1}x'_1 - \cdots - a_{q(n-1)}x'_{n-1}$$

(this inequality results when (p) and (q) are added), we can find a value x'_n of x_n which lies between left and right sides of the above inequality. Clearly $(x'_1, \cdots, x'_n)$ is a solution of the original system.

If this procedure were repeated, we would reach a system where either all the coefficients have zero value or where the signs of the coefficients of a variable are all the same in all the inequalities. As mentioned before, for these cases the theorem is true.

□

As an immediate consequence we can prove the celebrated 'Farkas Lemma'.

Theorem 1.3.2 (Farkas Lemma) *The homogeneous system*

$$\mathbf{A}\ \mathbf{x} \leq 0$$

has the consequence

$$\mathbf{d}^T\mathbf{x} \leq 0$$

iff the row vector $\mathbf{d}^T$ *is a nonnegative linear combination of the rows of* $\mathbf{A}$.

Proof : By Kuhn-Fourier Theorem (Theorem 1.3.1), the system

$$\mathbf{A}^T\mathbf{y} = \mathbf{d}$$

$$\mathbf{y} \geq \mathbf{0}$$

has a solution iff

'$\mathbf{x}^T\mathbf{A}^T + \beta^T\mathbf{I} = \mathbf{0}, \beta^T \geq \mathbf{0}$' implies '$\mathbf{x}^T\mathbf{d} \leq \mathbf{0}$';
i.e., iff '$\mathbf{Ax} \leq \mathbf{0}$' implies '$\mathbf{d}^T\mathbf{x} \leq \mathbf{0}$.'

□

The analogue of 'vector spaces' for inequality systems is '**cones**'. A **cone** is a collection of vectors closed under addition and non-negative linear combination. It is easily verified that the solution set of $\mathbf{Ax} \geq \mathbf{0}$ is a cone. Such cones are called **polyhedral**. We say vectors $\mathbf{x}$, $\mathbf{y}$ (on the same set S) are **polar** iff $< \mathbf{x}, \mathbf{y} >$ (i.e., the dot product) is nonpositive. If $\mathcal{K}$ is a collection of vectors, the **polar of** $\mathcal{K}$, denoted by $\mathcal{K}^p$ is the collection of vectors polar to every vector in $\mathcal{K}$. Thus, Farkas Lemma states:

'Let $\mathcal{C}$ be the polyhedral cone defined by $\mathbf{Ax} \leq \mathbf{0}$. A vector $\mathbf{d}$ belongs to $\mathcal{C}^p$ iff $\mathbf{d}^T$ is a nonnegative linear combination of the rows of $\mathbf{A}$.'

1.3.2 Linear Programming

Let $\mathcal{S}$ be a linear inequality system with '$\leq$' and '$=$' constraints ('$\geq$' and '$=$' constraints). The **linear programming problem** or **linear program** is that of finding a solution $\mathbf{x}$ of $\mathcal{S}$ which maximizes a given linear function $\mathbf{c}^T\mathbf{x}$ (minimizes a given linear function $\mathbf{c}^T\mathbf{x}$). The linear function to be optimized is called the **objective function**. A solution of $\mathcal{S}$ is called a **feasible solution**, while a solution which optimizes $\mathbf{c}^T\mathbf{x}$ is called an **optimal solution**, of the linear programming problem. The **value** of a feasible solution is the value of the objective function on it.

The following linear programming problems are said to be **duals** of each other

Primal program

$$\text{Maximize} \qquad \mathbf{c}_1^T\mathbf{x}_1 + \mathbf{c}_2^T\mathbf{x}_2$$

$$\begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \end{pmatrix} \begin{matrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{matrix} = \mathbf{b}_1$$

$$\begin{pmatrix} \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \begin{matrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{matrix} \leq \mathbf{b}_2$$

$$\mathbf{x}_2 \geq 0$$

Dual program

$$\text{Minimize} \qquad \mathbf{b}_1^T \mathbf{y}_1 + \mathbf{b}_2^T \mathbf{y}_2$$

$$\begin{pmatrix} \mathbf{A}_{11}^T & \mathbf{A}_{21}^T \end{pmatrix} \begin{matrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{matrix} = \mathbf{c}_1$$

$$\begin{pmatrix} \mathbf{A}_{12}^T & \mathbf{A}_{22}^T \end{pmatrix} \begin{matrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{matrix} \geq \mathbf{c}_2$$

$$\mathbf{y}_2 \geq 0.$$

We now present the duality theorem of linear programming [von Neumann47], [Gale+Kuhn+Tucker51].

Theorem 1.3.3 *For dual pairs of linear programs the following statements hold:*

i. The value of each feasible solution of the minimization program is greater than or equal to the value of each feasible solution of the maximization program;

ii. if both programs have feasible solutions then both have optimal solutions and the optimal values are equal;

iii. if one program has an optimal solution then so does the other.

The usual proof uses Farkas Lemma, or more conveniently, Kuhn-Fourier Theorem. We only sketch it.
Sketch of Proof: Part (i) follows by the solution of Exercise 1.1.
Now we write down the inequalities of the primal and dual programs and another '$\leq$' inequality which is the **opposite** of the inequality in part (i). Part (ii) would be proved if this system of inequalities has a solution. We assume it has no solution and derive a contradiction by using Kuhn-Fourier Theorem.

□

Exercise 1.1 *Prove part (i) of Theorem 1.3.3.*

A very useful corollary of Theorem 1.3.3 is the following:

Corollary 1.3.1 *(***Complementary Slackness***)*
Let

$$\left\{\begin{array}{c}\max \mathbf{c}^T\mathbf{x}\\ \mathbf{A}\mathbf{x}=\mathbf{b}\\ \mathbf{x}\geq \mathbf{0}\end{array}\right\} \text{ and } \left\{\begin{array}{c}\min \mathbf{b}^T\mathbf{y}\\ \mathbf{A}^T\mathbf{y}\geq \mathbf{c}\end{array}\right\}$$

be dual linear programs. Let $\hat{\mathbf{x}}, \hat{\mathbf{y}}$ *be optimal solutions to the respective programs. Then,*

i. $\hat{x}_i > 0$ *implies* $(\mathbf{A}^T)_i\hat{\mathbf{y}} = c_i$,

ii. $(\mathbf{A}^T)_i\hat{\mathbf{y}} > c_i$ *implies* $\hat{\mathbf{x}}_i = 0$.

Proof : We have by part (ii) of Theorem 1.3.3 $\mathbf{c}^T\hat{\mathbf{x}} = \hat{\mathbf{y}}^T\mathbf{b}$, equivalently

$$\mathbf{c}^T\hat{\mathbf{x}} - \hat{\mathbf{y}}^T\mathbf{A}\hat{\mathbf{x}} = (\mathbf{c}^T - \hat{\mathbf{y}}^T\mathbf{A})\hat{\mathbf{x}} = 0.$$

The result now follows since $(\mathbf{c}^T - \hat{\mathbf{y}}^T\mathbf{A}) \geq 0$ and $\hat{\mathbf{x}} \geq \mathbf{0}$.

□

1.4 Solutions of Exercises

E 1.1: We use the linear programs given in the definition of dual linear programs. We have

$$\begin{aligned} \left(\begin{array}{cc}\mathbf{b}_1^T & \mathbf{b}_2^T\end{array}\right) \begin{array}{c}\mathbf{y}_1\\ \mathbf{y}_2\end{array} &\geq \left(\left(\begin{array}{cc}\mathbf{x}_1^T & \mathbf{x}_2^T\end{array}\right)\left[\begin{array}{cc}\mathbf{A}_{11} & \mathbf{A}_{12}\\ \mathbf{A}_{21} & \mathbf{A}_{22}\end{array}\right]^T\right) \begin{array}{c}\mathbf{y}_1\\ \mathbf{y}_2\end{array} \\ &\geq \left(\begin{array}{cc}\mathbf{x}_1^T & \mathbf{x}_2^T\end{array}\right)\left[\begin{array}{c}\mathbf{c}_1\\ \mathbf{c}_2\end{array}\right]. \end{aligned}$$

1.5 Solutions of Problems

Most of these problems can be found as standard results in undergraduate texts on linear algebra (see for instance [Hoffman+Kunze72]). We only give the solution to the last two problems. Here we follow [MacDuffee33], [Gantmacher59] respectively.

P 1.16: We state the following simple lemma without proof.

Lemma 1.5.1 *If* $\alpha_1, \cdots, \alpha_t$ *are permutations of* $\{1, \cdots, n\}$ *then* $sgn(\alpha_1\alpha_2\cdots\alpha_t) = (sgn(\alpha_1))(sgn(\alpha_2))\cdots(sgn(\alpha_t))$ *(where* $\alpha_i\alpha_j$ *denotes composition of permutations* α_i, α_j*).*

We have

$$\sum sgn(\sigma)\ det\left(\mathbf{A}\left(\begin{array}{ccc}r_1, & \cdots & ,r_k\\ i_1, & \cdots & ,i_k\end{array}\right)\right) det\left(\mathbf{A}\left(\begin{array}{ccc}r_{k+1}, & \cdots & ,r_m\\ i_{k+1}, & \cdots & ,i_m\end{array}\right)\right) =$$

$$\sum sgn(\sigma)(\sum sgn(\alpha)(a_{r_1\alpha(r_1)}\cdots a_{r_k\alpha(r_k)}))(\sum sgn(\beta)(a_{r_{k+1}\beta(r_{k+1})}\cdots a_{n\beta(n)})),$$

where α, β are permutations on the sets $\{r_1, \cdots, r_k\}$, $\{r_{k+1}, \cdots, r_n\}$ respectively. Let α' agree with α over $\{r_1, \cdots, r_k\}$ and over $\{r_{k+1}, \cdots, r_n\}$, with the identity permutation. Let β' agree with β over $\{r_{k+1}, \cdots, r_n\}$ and with the identity permutation over $\{r_1, \cdots, r_k\}$. So

$$\begin{aligned} LHS &= \sum sgn(\sigma) sgn(\alpha') sgn(\beta') (a_{r_1 \alpha(r_1)} \cdots a_{r_k \alpha(r_k)} a_{r_{k+1} \beta(r_{k+1})} \cdots a_{n \beta(n)}) \\ &= \sum sgn(\beta' \alpha' \sigma) (a_{i_1 \alpha\sigma(i_1)} \cdots a_{i_k \alpha\sigma(i_k)} a_{i_{k+1} \beta\sigma(i_{k+1})} \cdots a_{i_n \beta\sigma(i_n)}) \\ &= \sum sgn(\mu) (a_{i_1 \mu(i_1)} \cdots a_{i_k \mu(i_k)} a_{i_{k+1} \mu(i_{k+1})} \cdots a_{i_n \mu(i_n)}), \end{aligned}$$

where $\mu \equiv \beta' \alpha' \sigma$. Since the RHS is the usual definition of the determinant of $\mathbf{A}$, the proof is complete.

P 1.17: Let a_{ij}, b_{ij} denote respectively the $(i,j)^{th}$ entry of $\mathbf{A}, \mathbf{B}$. Then the matrix

$$\mathbf{AB} = \begin{bmatrix} \sum_{i_1=1}^n a_{1i_1} b_{i_1 1} & \cdots & \sum_{i_m=1}^n a_{1i_m} b_{i_m m} \\ \vdots & & \vdots \\ \sum_{i_1=1}^n a_{mi_1} b_{i_1 1} & \cdots & \sum_{i_m=1}^n a_{mi_m} b_{i_m m} \end{bmatrix}.$$

Now each column of $\mathbf{AB}$ can be thought of as the sum of n appropriate columns - for instance the transpose of the first column is made up of rows - a typical one being $(a_{1i_1} b_{i_1 1}, \cdots, a_{mi_1} b_{i_1 1})$. Using Theorem 1.2.8

$$\begin{aligned} det(\mathbf{AB}) &= \sum_{i_1, \cdots, i_m} det\left(\begin{bmatrix} a_{1i_1} b_{i_1 1} & \cdots & a_{1i_m} b_{i_m m} \\ \vdots & & \vdots \\ a_{mi_1} b_{i_1 1} & \cdots & a_{mi_m} b_{i_m m} \end{bmatrix} \right) \\ &= \sum (b_{i_1 1} \cdots b_{i_m m}) \ det\left(\mathbf{A} \begin{pmatrix} 1, & \cdots & ,m \\ i_1, & \cdots & ,i_m \end{pmatrix} \right), \end{aligned}$$

where $\mathbf{A} \begin{pmatrix} 1, & \cdots & ,m \\ i_1, & \cdots & ,i_m \end{pmatrix}$ is the $m \times m$ matrix which has the first m rows of $\mathbf{A}$ in the same order as in $\mathbf{A}$ but whose j^{th} column is the i_j^{th} column of $\mathbf{A}$. So, again by Theorem 1.2.8,

$$\det(\mathbf{AB}) = \sum_{k_1, \cdots, k_m} det\left(\mathbf{A} \begin{pmatrix} 1, & \cdots & ,m \\ k_1, & \cdots & ,k_m \end{pmatrix} \right) (sgn(\sigma)) \ b_{\sigma(k_1)1} \cdots b_{\sigma(k_m)m},$$

where $k_1 < \cdots < k_m, \{k_1, \cdots, k_m\} = \{i_1, \cdots, i_m\}$ and σ is the permutation $\begin{pmatrix} k_1, & \cdots & ,k_n \\ i_1, & \cdots & ,i_n \end{pmatrix}$, i.e.,

$$\sigma(k_j) = i_j.$$

So,

$$det(\mathbf{AB}) =$$

$$\sum_{\substack{k_1, \cdots, k_m \\ k_1 < \cdots < k_m}} det\left(\mathbf{A}\begin{pmatrix} 1, & \cdots & , m \\ k_1, & \cdots & , k_m \end{pmatrix}\right) det\left(\mathbf{B}\begin{pmatrix} k_1, & \cdots & , k_m \\ 1, & \cdots & , m \end{pmatrix}\right).$$

Bibliography

[Fourier1826] J. Fourier: Solution d'une question particulière du calcul des inégalités. *Oeuvres II* (1826) 317-328.

[Gale+Kuhn+Tucker51] D. Gale, H.W. Kuhn and A.W. Tucker: Linear Programming and the theory of games. In: *Activity Analysis of production and allocation* (John Wiley, New York, 1951).

[Gantmacher59] F.R.Gantmacher: *The Theory of Matrices.* Volumes I,II (translated from the Russian) (Chelsea, NewYork, 1959).

[Hoffman+Kunze72] K.Hoffman and R.Kunze: *Linear Algebra* (2nd ed.) (Prentice-Hall International, inc., Englewood Cliffs, 1972). (Prentice-Hall of India, 1972).

[Jacobson74] N.Jacobson: *Basic Algebra I&II*(W.H.Freeman and Co.,U.S.A,1974).

[Kuhn56] H.W. Kuhn: Solvability and consistency for linear equations and inequalities. *American Mathematical Monthly* **63** (1956) 217-232.

[MacDuffee33] C.C.MacDuffee: *The Theory of Matrices* (Springer, Berlin, 1933) (reprinted Chelsea, NewYork, 1946).

[Schrijver86] A. Schrijver: *Theory of Linear and Integer Programming* (John Wiley & Sons, New York, 1986).

[Stoer+Witzgall70] J. Stoer and C. Witzgall: *Convexity and Optimization in Finite Dimensions* **I** (Springer, Berlin, 1970).

[von Neumann47] J. von Neumann: Discussion of a maximum problem. Unpublished working paper. Institute for Advanced Studies, Princeton,New Jersey (1947). Reprinted in
John von Neumann, Collected Works, vol VI. (Pergamon Press, Oxford, 1963).