Optimization Class Lecture Notes

Contents

CHAPTER 1

Classical Optimization

1. Review: Linear Algebra and Quadratic Forms

Throughout this course, we will try to find minimal and maximal values of functions $f : D \to \Re$, where D is a subset of $\Re^n$ for a certain $n \geq 1$. We start by reviewing some of the algebraic and analytic properties of $\Re^n$ and functions of several variables.

1.1. Vectors. Elements of $\Re^n$ are written as column vectors, using either bold faced letters or by placing arrows above letters. Components of vectors are denoted by lower indices.

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \in \Re^n$$

or

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} \in \Re^k$$

Vectors can be added and multiplied by scalars component-wise:

$$\vec{x} + \vec{y} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix}$$

$$r\vec{x} = r \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} rx_1 \\ \vdots \\ rx_n \end{bmatrix}$$

The scalar product or dot-product of two vectors is defined by

$$\vec{x} \cdot \vec{y} = x_1 y_1 + \cdots + x_n y_n$$

The vector all whose components is called the zero-vector:

$$\mathbf{o} = \vec{0} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

The i^{th} unit vector is the vector with an entry of 1 in the i^{th} component, and all other components 0 :

$$\vec{e}_i = \mathbf{e}_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Rules for those algebraic operations were discussed in a course on linear algebra. Those rules include various forms of commutative, associative and distributive laws.

Vectors $\vec{a}_1, \ldots, \vec{a}_m \in \Re^n$ are linearly independent if

$$r_1\vec{a}_1 + \cdots + r_m\vec{a}_m = \vec{0}$$

implies that $r_1 = \cdots = r_m = 0$. Vectors that are not linearly independent are called linearly dependent.

A maximal linearly independent collection of vectors $\vec{a}_1, \ldots, , \vec{a}_m \in \Re^n$ is called a basis. If $B = (\vec{a}_1, \ldots, \vec{a}_m)$ is a basis, then $m = n$, and every vector $\vec{b} \in \Re^n$ can be written uniquely in the form

$$\vec{b} = r_1\vec{a}_1 + \cdots + r_n\vec{a}_n$$

The numbers $r_1, \ldots, r_n$ are called the coordinates of $\vec{b}$ with respect to the basis B, and written as

$$\left[\vec{b}\right]_B = \begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix}$$

Two vectors $\vec{a}$ and $\vec{b}$ are called orthogonal to each other, if

$$\vec{a} \cdot \vec{b} = 0.$$

The length of a vector $\vec{a}$ is defined as

$$\|\vec{a}\| = \sqrt{\vec{a} \cdot \vec{a}}$$

The Schwartz inequality says that

$$\left|\vec{a} \cdot \vec{b}\right| \leq \|\vec{a}\| \left\|\vec{b}\right\|$$

The angle α between $\vec{\alpha}$ and $\vec{b}$ is defined by

$$\alpha = \arccos \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \left\|\vec{b}\right\|}$$

A basis $B = \left[\vec{b}_1, \ldots, \vec{b}_n\right]$ of $\Re^n$ is called an orthonormal basis, if the $\vec{b}_i$ are pair-wise orthogonal and have length 1 :

$$\vec{b}_i \cdot \vec{b}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

If $B = \left[\vec{b}_1, \ldots \vec{b}_n\right]$ are an orthonormal basis, then the coordinates of a vector $\vec{b}$ are given by $r_i = \vec{b} \cdot \vec{b}_i$:

$$\begin{aligned} \vec{b} &= \left(\vec{b} \cdot \vec{b}_1\right) \vec{b}_1 + \cdots + \left(\vec{b} \cdot \vec{b}_n\right) \vec{b}_n \\ \left[\vec{b}\right]_B &= \begin{bmatrix} \vec{b} \cdot \vec{b}_1 \\ \vdots \\ \vec{b} \cdot \vec{b}_n \end{bmatrix} \end{aligned}$$

The canonical unit vectors form an orthonormal basis.

EXAMPLE 1. *Complete the vectors*

$$\begin{aligned} \vec{x}_1 &= \frac{1}{5} \begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix} \\ \vec{x}_2 &= \frac{1}{5} \begin{bmatrix} -4 \\ 3 \\ 0 \end{bmatrix} \end{aligned}$$

to an orthonormal basis. Also, find the coordinate of

$$\vec{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

with respect to this basis.

The third vector will be the cross product of $\vec{x}_1$ and $\vec{x}_2$: $\frac{1}{5}[3,4,0] \times \frac{1}{5}[-4,3,0] = \frac{1}{25}[0,0,25] = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$. The components of $\vec{b}$ are

$$\begin{aligned} r_1 &= \frac{11}{5} \\ r_2 &= \frac{2}{5} \\ r_3 &= 3 \end{aligned}$$

1.2. Matrices. A matrix with m rows and n columns is a rectangular scheme of real numbers:

$$\begin{aligned} T &= \left[t_{ij}\right]_{1 \leq i \leq m, 1 \leq j \leq n} \\ &= \begin{bmatrix} t_{11} & t_{12} & t_{13} & \ldots & t_{1n} \\ t_{21} & t_{22} & t_{23} & \ldots & t_{2n} \\ t_{31} & t_{32} & t_{33} & \ldots & t_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{m1} & t_{m2} & t_{m3} & \ldots & t_{mn} \end{bmatrix} \end{aligned}$$

The index i is called the row index and the index j is called the column index.

Matrices can be multiplied by scalars. Further, if two matrices T and S have the same number of rows and columns, then S and T can be added component-wise:

$$\begin{aligned} rS &= r\left[s_{ij}\right]_{1\leq i\leq m,1\leq j\leq n} = \left[rs_{ij}\right]_{1\leq i\leq m,1\leq j\leq n} \\ S+T &= \left[s_{ij}\right]_{1\leq i\leq m,1\leq j\leq n} + \left[t_{ij}\right]_{1\leq i\leq m,1\leq j\leq n} \\ &= \left[s_{ij}+t_{ij}\right]_{1\leq i\leq m,1\leq j\leq n} \end{aligned}$$

If the number of columns of a matrix S is equal to the number of rows of a matrix T, then we can define the product ST. If $S = \left[s_{ij}\right]_{1\leq i\leq m,1\leq j\leq n}$ and $T = \left[t_{jk}\right]_{1\leq j\leq n,1\leq k\leq l}$, then

$$ST = \left[r_{ik}\right]_{1\leq i\leq m,1\leq k\leq l}$$

where

$$r_{ik} = \sum_{j=1}^{n} s_{ij}t_{jk}$$

i.e. r_{ik} is obtained from multiply the i^{th} row of S by the k^{th} row of T.

The transpose of a matrix A is defined by interchanging row and column indices:

$$\begin{aligned} A &= \left[a_{ij}\right]_{1\leq i\leq m,1\leq j\leq n} \\ A^T &= \left[a_{ji}\right]_{1\leq j\leq n,1\leq i\leq m} \end{aligned}$$

Again, in a course of linear algebra, various commutative, associative and distributive laws for matrix operations are verified. It is important to notice that matrix multiplication is not commutative: If A and B are two matrices so that AB and BA are both defined, then it is not necessarily true that $AB = BA$. For example, if

$$\begin{aligned} A &= \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \\ B &= \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} AB &= \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} =: \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \\ BA &= \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \end{aligned}$$

Vectors can be viewed as special matrices: An element $\vec{v} \in \Re^n$ can be thought to be a matrix with n rows and one column. Similarly, a matrix with one row and n columns is sometimes called a row-vector.

1.3. Linear Transformations. If a matrix M with m rows and n columns is multiplied by a (column) vector $\vec{v} \in \Re^n$, the result is a (column) vector $M\vec{v} \in \Re^m$. In this way, we obtain a map from $\Re^n$ to $\Re^m$:

$$\begin{aligned} T_M &: \Re^n \to \Re^m \\ T_M\left(\vec{v}\right) &= M\vec{v} \end{aligned}$$

The associative and distributive properties imply:

$$\begin{aligned} T_M\left(\vec{v}+\vec{w}\right) &= T_M\left(\vec{v}\right) + T_M\left(\vec{w}\right) \\ T_M\left(r\vec{v}\right) &= rT_M\left(\vec{v}\right) \end{aligned}$$

and

$$T_M(\mathbf{o}) = \mathbf{o}$$

The first two properties of T_M are used to define linear maps: A map $T : \Re^n \to \Re^m$ is called linear, provided that for all vectors $\vec{v}, \vec{w} \in \Re^n$ and all scalars $r \in \Re$ we have

$$\begin{aligned} T(\vec{v}+\vec{w}) &= T(\vec{v}) + T(\vec{w}) \\ T(r\vec{v}) &= rT(\vec{v}) \end{aligned}$$

Every linear map $T : \Re^n \to \Re^m$ can be represented as matrix multiplication by a matrix M. The matrix M is uniquely determined and can be found by writing the vectors $T(\mathbf{e}_i)$ into the i^{th} column of M.

EXAMPLE 2. *Verify that*

$$T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x+y-3z \\ -2x+5y \end{bmatrix}$$

is linear and find the matrix representing T.

Since

$$T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} 1 & 1 & -3 \\ -2 & 5 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

we find that

$$M = \begin{bmatrix} 1 & 1 & -3 \\ -2 & 5 & 0 \end{bmatrix}$$

Therefore, T is given by matrix multiplication and hence linear.

EXAMPLE 3. *Is the map T given by*

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x+1 \\ y-1 \end{bmatrix}$$

linear? - This map is not linear: We find that

$$\begin{aligned} T\left(2\begin{bmatrix} x \\ y \end{bmatrix}\right) &= T\begin{bmatrix} 2x \\ 2y \end{bmatrix} = \begin{bmatrix} 2x+1 \\ 2y-1 \end{bmatrix} \\ 2T\begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} 2x+2 \\ 2y-2 \end{bmatrix} \end{aligned}$$

Those values are different, hence the map is not linear.

EXAMPLE 4. *Is the map T given by*

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x^2+y \\ x-y^2 \end{bmatrix}$$

linear?

We find that

$$T\begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ -2 \end{bmatrix}$$

and

$$T\begin{bmatrix}2\\0\end{bmatrix}+T\begin{bmatrix}0\\2\end{bmatrix}=\begin{bmatrix}4\\2\end{bmatrix}+\begin{bmatrix}2\\-4\end{bmatrix}=\begin{bmatrix}6\\-2\end{bmatrix}=T\begin{bmatrix}2\\2\end{bmatrix}$$

Even though these two values are identical, this does not imply that T is linear. Indeed,

$$T\begin{bmatrix}1\\1\end{bmatrix}+T\begin{bmatrix}1\\1\end{bmatrix}=\begin{bmatrix}3\\0\end{bmatrix}+\begin{bmatrix}3\\0\end{bmatrix}\neq\begin{bmatrix}6\\-2\end{bmatrix}=T\begin{bmatrix}2\\2\end{bmatrix}$$

Since the values for $T\begin{bmatrix}1\\1\end{bmatrix}+T\begin{bmatrix}1\\1\end{bmatrix}$ and $T\left(\begin{bmatrix}1\\1\end{bmatrix}+\begin{bmatrix}1\\1\end{bmatrix}\right)=T\begin{bmatrix}2\\2\end{bmatrix}$ are different, the map is not linear.

DEFINITION 1. *Let A be a square matrix. If $\vec{v}$ is a non-zero vector, and if λ is a number so that*

$$A\vec{v}=\lambda\vec{v}$$

then λ is called an eigenvalue of A with eigenvector $\vec{v}$.

The equation

$$A\vec{v}=\lambda\vec{v}$$

can be rewritten as

$$(A-\lambda I)\,\vec{v}=\vec{0}$$

Since $\vec{v}\neq\vec{0}$, the matrix $A-\lambda I$ is singular (non-invertible), hence eigenvalues are characterized by the equation

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

The function $\chi(\lambda)=\det(A-\lambda I)$ is called the characteristic polynomial of A, and the zeros of the characteristic polynomial are exactly the eigenvalues of A. Even though we will be dealing only with real matrices and real numbers, it is sometimes convenient to consider complex eigenvalues and eigenvectors.

Recall that a matrix A that satisfies $A=A^T$ is called a symmetric matrix.

THEOREM 1. *Assume that A is a symmetric $n\times n$ matrix. Then A has only real eigenvalues, and $\Re^n$ has an orthonormal basis consisting of eigenvectors of A.*

If A is a symmetric $n\times n$ matrix, and if we write the orthonormal basis of eigenvectors of $\Re^n$ into the columns of a matrix S, then $S^{-1}=S^T$ and S^TAS is a diagonal matrix: Only entries on the diagonal can be different from 0, and the eigenvectors of A can be found on the diagonal of S^TAS (counted in their multiplicity).

EXAMPLE 5. *If*

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

find an orthonormal basis of eigenvectors of A and a matrix S so that the product $S^{-1}AS$ is a diagonal matrix.

SOLUTION 1. *The characteristic polynomial is given by* $\det = 4\lambda + 3\lambda^2 - \lambda^3$, *and the zeros of this polynomial are* $4, 0$ *and* -1. *For* $\lambda = 0$ *we find the eigenvectors by solving*

$$\begin{bmatrix} 1 & \sqrt{3} & 1 \\ \sqrt{3} & 1 & \sqrt{3} \\ 1 & \sqrt{3} & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

The general solution of this equation is given by $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = r \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$. *Similarly the eigenvectors for* $\lambda = 1$ *are given by* $r \begin{bmatrix} 1 \\ -\sqrt{3} \\ 1 \end{bmatrix}$ *and the eigenvectors for* $\lambda = 4$ *are given by* $r \begin{bmatrix} 1 \\ \frac{2}{3}\sqrt{3} \\ 1 \end{bmatrix}$. *Normalizing those eigenvectors to length 1 gives the matrix* S :

$$S = \begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{\frac{10}{3}}} \\ 0 & -\frac{\sqrt{3}}{\sqrt{5}} & \frac{2}{3}\frac{\sqrt{3}}{\sqrt{\frac{10}{3}}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{\frac{10}{3}}} \end{bmatrix}$$

Indeed, $S^{-1} = \begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{\frac{10}{3}}} \\ 0 & -\frac{\sqrt{3}}{\sqrt{5}} & \frac{2}{3}\frac{\sqrt{3}}{\sqrt{\frac{10}{3}}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{\frac{10}{3}}} \end{bmatrix}^{-1} = \begin{bmatrix} -\frac{1}{2}\sqrt{2} & 0 & \frac{1}{2}\sqrt{2} \\ \frac{1}{5}\sqrt{5} & -\frac{1}{5}\sqrt{3}\sqrt{5} & \frac{1}{5}\sqrt{5} \\ \frac{1}{10}\sqrt{3}\sqrt{10} & \frac{1}{5}\sqrt{10} & \frac{1}{10}\sqrt{3}\sqrt{10} \end{bmatrix}$

$= S^T$ *and*

$$S^T A S = \begin{bmatrix} -\frac{1}{2}\sqrt{2} & 0 & \frac{1}{2}\sqrt{2} \\ \frac{1}{5}\sqrt{5} & -\frac{1}{5}\sqrt{3}\sqrt{5} & \frac{1}{5}\sqrt{5} \\ \frac{1}{10}\sqrt{3}\sqrt{10} & \frac{1}{5}\sqrt{10} & \frac{1}{10}\sqrt{3}\sqrt{10} \end{bmatrix} \begin{bmatrix} 1 & \sqrt{3} & 1 \\ \sqrt{3} & 1 & \sqrt{3} \\ 1 & \sqrt{3} & 1 \end{bmatrix} \begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{\frac{10}{3}}} \\ 0 & -\frac{\sqrt{3}}{\sqrt{5}} & \frac{2}{3}\frac{\sqrt{3}}{\sqrt{\frac{10}{3}}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{\frac{10}{3}}} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

A function $p(x_1, x_2, \ldots, x_n)$ is a quadratic form, if it can be written in the form

$$p(x_1, x_2, \ldots, x_n) = [x_1, \ldots, x_n] A \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

where A is a symmetric form.

Examples of quadratic forms are

$$\begin{aligned} p(x, y) &= x^2 + y^2 \\ p(x, y) &= x^2 - y^2 \\ p(x, y) &= x^2 + xy - y^2 \end{aligned}$$

and

$$p(u, v, w, x) = u^2 - v^2 - w^2 + x^2 - 2xy - 2wx$$

In matrix notation, those quadratic forms can be written as

$$\begin{aligned} x^2+y^2 &= (x,y)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} \\ x^2-y^2 &= (x,y)\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} \\ x^2+xy-y^2 &= (x,y)\begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & -1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} \\ u^2-v^2-w^2+x^2-2xy-2wx &= (u,v,w,x,y)\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & -1 & 1 & -1 \\ 0 & 0 & 0 & -1 & 0 \end{bmatrix}\begin{bmatrix} u \\ v \\ w \\ x \\ y \end{bmatrix} \end{aligned}$$

As the following example illustrates, the eigenvalues and eigenvectors of A are particular importance when we are dealing with quadratic forms:

EXAMPLE 6. *Find the minimum of* $f(x,y) = x^2+y^2$ *subject to the condition* $x^2+xy-y^2=1$.

SOLUTION 2. *In matrix form, the constraint* $x^2+xy-y^2=1$ *can be written as*

$$[x,y]\begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & -1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = 1$$

The eigenvalues of $A = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & -1 \end{bmatrix}$ are $\pm\frac{1}{2}\sqrt{5}$; the eigenvectors for $\frac{1}{2}\sqrt{5}$ are multiples of $\begin{bmatrix} \sqrt{5}+2 \\ 1 \end{bmatrix}$, and the eigenvectors for $-\frac{1}{2}\sqrt{5}$ are multiples of $\begin{bmatrix} -\sqrt{5}+2 \\ 1 \end{bmatrix}$. Writing the normalized eigenvectors into the columns of S gives

$$S = \begin{bmatrix} \frac{\sqrt{5}+2}{\sqrt{10+4\sqrt{5}}} & \frac{-\sqrt{5}+2}{\sqrt{10-4\sqrt{5}}} \\ \frac{1}{\sqrt{10+4\sqrt{5}}} & \frac{1}{\sqrt{10-4\sqrt{5}}} \end{bmatrix}$$

Again, $S^{-1} = S^T = \begin{bmatrix} \frac{1}{2}\sqrt{2}\frac{\sqrt{5}+2}{\sqrt{2\sqrt{5}+5}} & \frac{1}{2}\frac{\sqrt{2}}{\sqrt{2\sqrt{5}+5}} \\ \frac{1}{2}\sqrt{2}\frac{-\sqrt{5}+2}{\sqrt{-2\sqrt{5}+5}} & \frac{1}{2}\frac{\sqrt{2}}{\sqrt{-2\sqrt{5}+5}} \end{bmatrix}$, and

$$S^T A S = \begin{bmatrix} \frac{1}{2}\sqrt{5} & 0 \\ 0 & -\frac{1}{2}\sqrt{5} \end{bmatrix}$$

Since S is the matrix of a rotation (try $\cos\alpha = \frac{1}{2}\sqrt{2}\frac{\sqrt{5}+2}{\sqrt{2\sqrt{5}+5}}$ and $\sin\alpha = \frac{1}{2}\sqrt{2}\frac{-\sqrt{5}+2}{\sqrt{-2\sqrt{5}+5}}$), This implies

$$A = S\begin{bmatrix} \frac{1}{2}\sqrt{5} & 0 \\ 0 & -\frac{1}{2}\sqrt{5} \end{bmatrix} S^T$$

If we let

$$\begin{bmatrix} \xi \\ \eta \end{bmatrix} = S^T\begin{bmatrix} x \\ y \end{bmatrix}$$

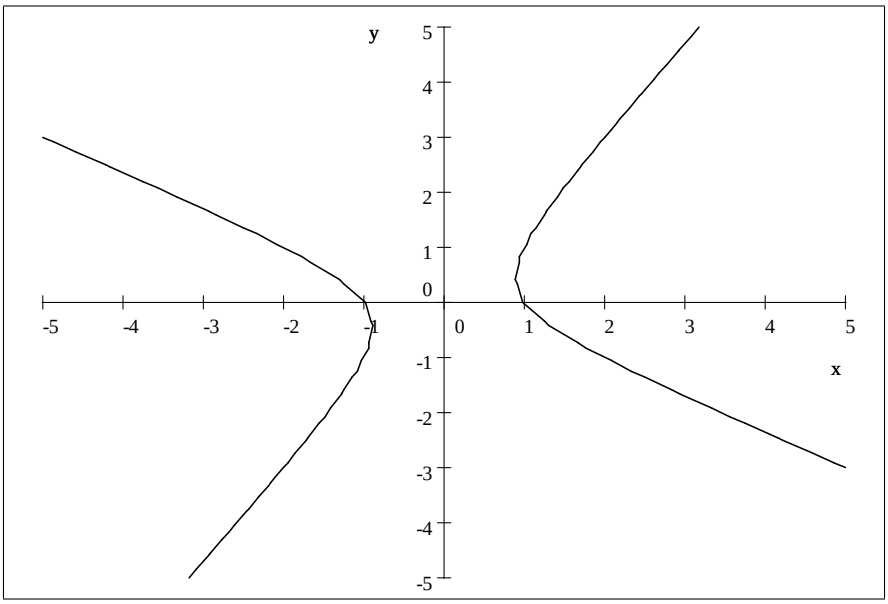
Since S is the matrix of a rotation (try $\cos\alpha = \frac{1}{2}\sqrt{2}\frac{\sqrt{5}+2}{\sqrt{2\sqrt{5}+5}}$ and $\sin\alpha = \frac{1}{2}\sqrt{2}\frac{-\sqrt{5}+2}{\sqrt{-2\sqrt{5}+5}}$), it follows that $\xi^2+\eta^2 = x^2+y^2$ - we could also verify this as follows:

$$\begin{aligned}
\xi^2+\eta^2 &= [\xi,\eta]\begin{bmatrix}\xi\\ \eta\end{bmatrix}\\
&= \begin{bmatrix}\xi\\ \eta\end{bmatrix}^T\begin{bmatrix}\xi\\ \eta\end{bmatrix}\\
&= \left(S^T\begin{bmatrix}x\\ y\end{bmatrix}\right)^T S^T\begin{bmatrix}x\\ y\end{bmatrix}\\
&= [x,y]\,SS^T\begin{bmatrix}x\\ y\end{bmatrix}\\
&= [x,y]\,SS^{-1}\begin{bmatrix}x\\ y\end{bmatrix}\\
&= [x,y]\begin{bmatrix}x\\ y\end{bmatrix}\\
&= x^2+y^2
\end{aligned}$$

It therefore does not matter whether we minimize x^2+y^2 are $\xi^2+\eta^2$. In the new coordinate system, the equation for the constraint reads

$$\begin{aligned}
(\xi,\eta)\begin{pmatrix}\frac{1}{2}\sqrt{5} & 0\\ 0 & -\frac{1}{2}\sqrt{5}\end{pmatrix}\begin{pmatrix}\xi\\ \eta\end{pmatrix} &= 1\\
\frac{1}{2}\sqrt{5}\xi^2-\frac{1}{2}\sqrt{5}\eta^2 &= 1\\
\xi^2-\eta^2 &= \frac{2}{\sqrt{5}}
\end{aligned}$$

This is an hyperbola, and the minimum distance to the origin occurs at $\xi = \pm\sqrt{\frac{2}{\sqrt{5}}}$, $\eta = 0$. Hence the minimum of x^2+y^2 with respect to $x^2+xy-y^2=1$ is given by $\xi^2+\eta^2 = \sqrt{\frac{2}{\sqrt{5}}}^2 + 0^2 = \frac{2}{5}\sqrt{5}$.



2. Review: The Derivative Matrix

Let $F : D \subseteq \Re^n \to \Re^m$ be function. Then there are functions $f_i : D \subseteq \Re^n \to \Re$ so that

$$F(x_1, \ldots, x_n) = [f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n)]$$

If all the coordinate functions are differentiable, then we can form the derivative matrix or Jacobian of as follows:

$$F'(t_1, \ldots, t_m) = \begin{bmatrix} \frac{\partial}{\partial x_1} f_1(t_1, \ldots, t_n) & \frac{\partial}{\partial x_2} f_1(t_1, \ldots, t_n) & \cdots & \frac{\partial}{\partial x_n} f_1(t_1, \ldots, t_n) \\ \frac{\partial}{\partial x_1} f_2(t_1, \ldots, t_n) & \frac{\partial}{\partial x_2} f_2(t_1, \ldots, t_n) & \cdots & \frac{\partial}{\partial x_n} f_2(t_1, \ldots, t_n) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1} f_m(t_1, \ldots, t_n) & \frac{\partial}{\partial x_2} f_m(t_1, \ldots, t_n) & \cdots & \frac{\partial}{\partial x_n} f_m(t_1, \ldots, t_n) \end{bmatrix}$$

The derivative matrix can be used to find linear approximations of differentiable functions:

$$F(\vec{x}) \approx F'(\vec{x}_0)(\vec{x} - \vec{x}_0) + F(\vec{x}_0)$$

EXAMPLE 7. *Find a linear approximation of* $F(x) = \begin{bmatrix} \sin x \\ \cos x \end{bmatrix}$ *at* $x_0 = \frac{\pi}{4}$. -

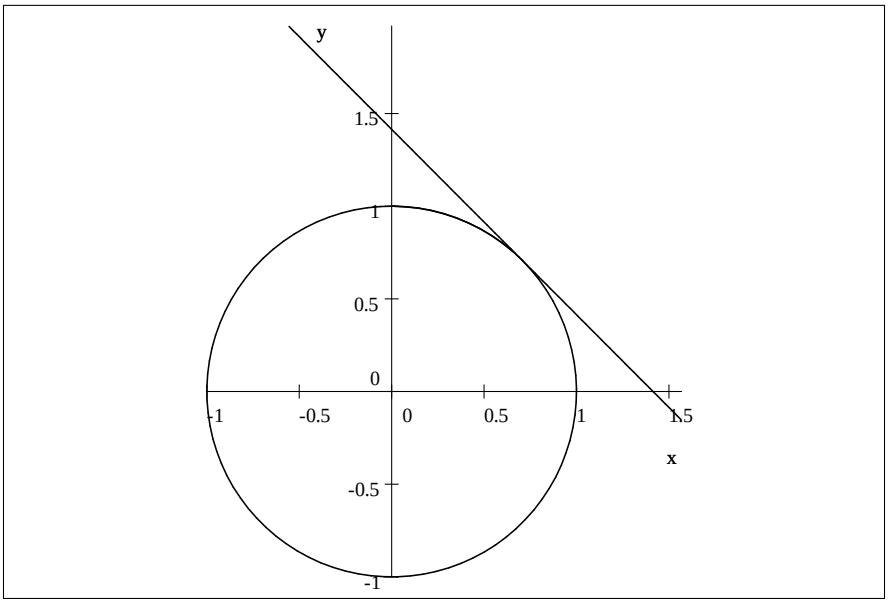
We have to find the derivative matrix of F :

$$\begin{aligned} F'(x_0) &= \begin{bmatrix} \cos x_0 \\ -\sin x_0 \end{bmatrix} (x - x_0) \\ &= \begin{bmatrix} \frac{1}{2}\sqrt{2} \\ -\frac{1}{2}\sqrt{2} \end{bmatrix} \left(x - \frac{\pi}{4}\right) \end{aligned}$$

Hence

$$\begin{bmatrix} \sin x \\ \cos x \end{bmatrix} \approx \begin{bmatrix} \frac{1}{2}\sqrt{2} \\ -\frac{1}{2}\sqrt{2} \end{bmatrix} \left(x - \frac{\pi}{4}\right) + \begin{bmatrix} \frac{1}{2}\sqrt{2} \\ \frac{1}{2}\sqrt{2} \end{bmatrix}.$$

Plotting both the function and its approximation gives the following picture:

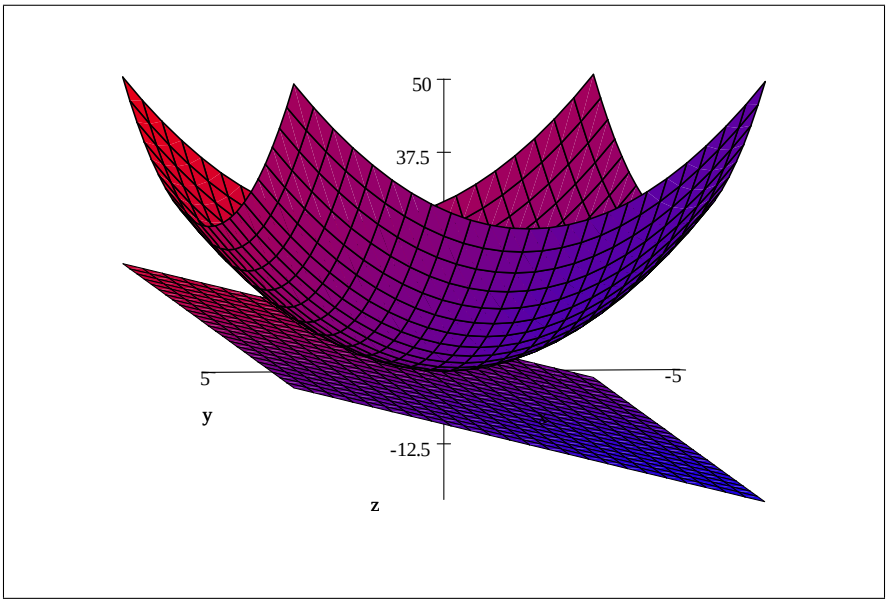


EXAMPLE 8. *Find a linear approximation of* $F(x,y) = x^2 + y^2$ *at* $x = [1,1]$.

The derivative matrix of $x^2 + y^2$ is $[2x, 2y]$, and therefore the approximation is given by

$$\begin{aligned} F(x,y) &\approx [2,2] \begin{bmatrix} x-1 \\ y-1 \end{bmatrix} + 2 \\ &= 2x + 2y - 2 \end{aligned}$$

Plotting both graphs results in the following picture:



EXAMPLE 9. *Find a linear approximation of* $F(x,y) = \left[x^2 + y^2, x^2 - y^2\right]$ *at* $x = [2,1]$.

In this case, the derivative matrix is equal to

$$F'(x,y) = \begin{bmatrix} 2x & 2y \\ 2x & -2y \end{bmatrix}$$

At $x = [2,1]$ we obtain

$$F'(2,1) = \begin{bmatrix} 4 & 2 \\ 4 & -2 \end{bmatrix}$$

Hence

$$\begin{bmatrix} x^2 + y^2 \\ x^2 - y^2 \end{bmatrix} \approx \begin{bmatrix} 4 & 2 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} x-2 \\ y-1 \end{bmatrix} + \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$

Now let us consider two differentiable functions $f : D \subseteq \Re^p \to \Re^n$ and $g : E \subseteq \Re^n \to \Re^m$. Assume that for every vector $\vec{x} \in D$ the value $f(\vec{x})$ belongs to E. Then the composition $g \circ f$ is defined and yields a differentiable function $g \circ f : D \subseteq \Re^p \to \Re^m$. The chain rule allows us to compute the derivative matrix of $g \circ f$, if the derivative matrices of f and g are known:

$$(g \circ f)'(\vec{x}) = g'(f(\vec{x})) \cdot f'(\vec{x})$$

EXAMPLE 10. *Assume that* $f(x,y) = \left[x+y, x^2-y^2, x^2+y^2\right]$ *and* $g(r,s,t) = [r+s, r-s, r+t, r-t]$. *Find*

(1) *The function* $g \circ f$.
(2) *The derivative matrix of* $g \circ f$ *without using the chain rule.*
(3) *The derivative matrix of* $g \circ f$ *by applying the chain rule.*

First, we find

$$\begin{aligned}(g \circ f)(x,y) &= g(f(x,y)) \\ &= g\left(x+y, x^2-y^2, x^2+y^2\right) \\ &= \left[x+y+x^2-y^2, x+y-x^2+y^2, x+y+x^2+y^2, x+y-x^2-y^2\right]\end{aligned}$$

Hence the derivative matrix of $g \circ f$ is equal to

$$(g \circ f)'(x,y) = \begin{bmatrix} 1+2x & 1-2y \\ 1-2x & 1+2y \\ 1+2x & 1+2y \\ 1-2x & 1-2y \end{bmatrix}$$

The chain rule give the same result:

$$\begin{aligned} f'(x,y) &= \begin{bmatrix} 1 & 1 \\ 2x & -2y \\ 2x & 2y \end{bmatrix} \\ g'(r,s,t) &= \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & -1 \end{bmatrix}\end{aligned}$$

We now would have to express r, s and t in terms of x and y using the indentities $r = x+y$, $s = x^2-y^2$ and $t = x^2+y^2$. In this case, this is not necessary because the matrix $g'(r,s,t)$ contains only constants. So we obtain:

$$\begin{aligned}(g \circ f)'(x,y) &= g'(f(x.y)) \cdot f'(x,y) \\ &= \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2x & -2y \\ 2x & 2y \end{bmatrix} \\ &= \begin{bmatrix} 2x+1 & -2y+1 \\ -2x+1 & 2y+1 \\ 2x+1 & 2y+1 \\ -2x+1 & -2y+1 \end{bmatrix}\end{aligned}$$

3. Review: Gradients and Level Sets

If $m = 1$, i.e. if $f : D \subseteq \Re^n \to \Re$ is a real-valued function, then the Jacobian of f is also called the gradient of f :

$$grad\,(f)\,(\vec{x}) = \nabla f\,(\vec{x}) = [f_{x_1}\,(\vec{x}), \ldots, f_{x_n}\,(\vec{x})]$$

The gradient can be used to find directional derivatives. If $\vec{u} \in \Re^n$ is a vector, then we define

$$\frac{\partial f}{\partial \vec{u}}\,(\vec{x}) = \lim_{h \to 0} \frac{f\,(\vec{x} + h\vec{u}) - f\,(\vec{x})}{h}$$

Using the gradient of f, we find that

$$\frac{\partial f}{\partial \vec{u}}\,(\vec{x}) = grad\,(f)\,(\vec{x}) \cdot \vec{u}$$

The Schwartz inequality yields

$$\left|\frac{\partial f}{\partial \vec{u}}\,(\vec{x})\right| \leq \|grad\,(f)\,(\vec{x})\| \cdot \|\vec{u}\|$$

or

$$\frac{\partial f}{\partial \vec{u}}\,(\vec{x}) \leq \|grad\,(f)\,(\vec{x})\| \cdot \|\vec{u}\|$$

If, in addition, $\|\vec{u}\| = 1$, then

$$\frac{\partial f}{\partial \vec{u}}\,(\vec{x}) \leq \|grad\,(f)\,(\vec{x})\|$$

In the last equation, equality hold precisely if $grad\,(f)\,(\vec{x})$ and $\vec{u}$ point in the same direction. Hence the rate of change is maximal if $\vec{u}$ points in the direction of $grad\,(f)\,(\vec{x})$ (and, similarly, the rate of change is minimal, if $\vec{u}$ and $grad\,(f)\,(\vec{x})$ point in opposite directions).

EXAMPLE 11. *Find the direction in which the rate of change is maximal at* (x_0, y_0)*, if* $f\,(x, y) = x^2 - y^2$*.* -

The gradient at $[x_0, y_0]$ is $[2x_0, -2y_0]$, and this point in the direction of $\frac{1}{\sqrt{x_0^2+y_0^2}}\,[x_0, -y_0]$

Next, we discuss the concept of level sets. If $f : D \subseteq \Re^n \to \Re$ is a function, and if $c \in \Re$ is a constant, then the solution set of

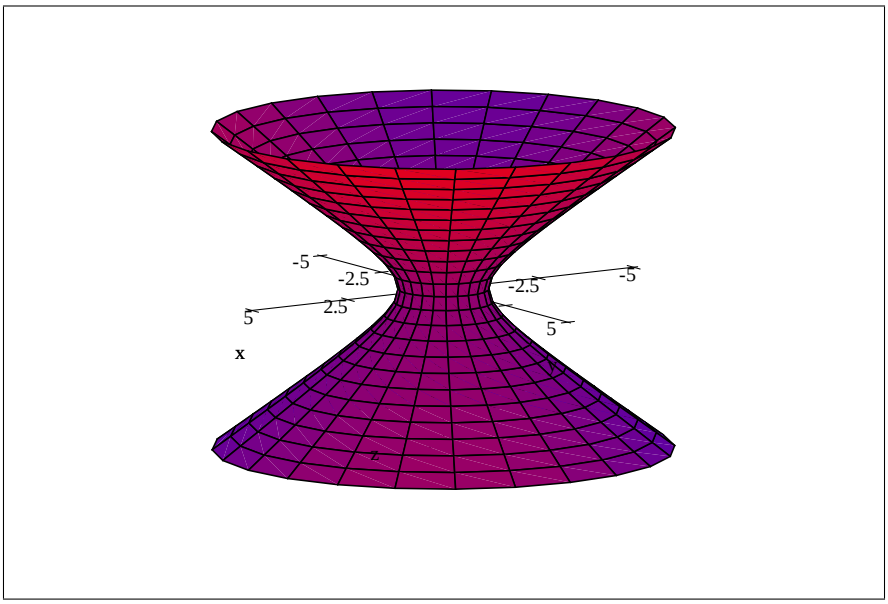
$$f\,(\vec{x}) = c$$

is called a level set.

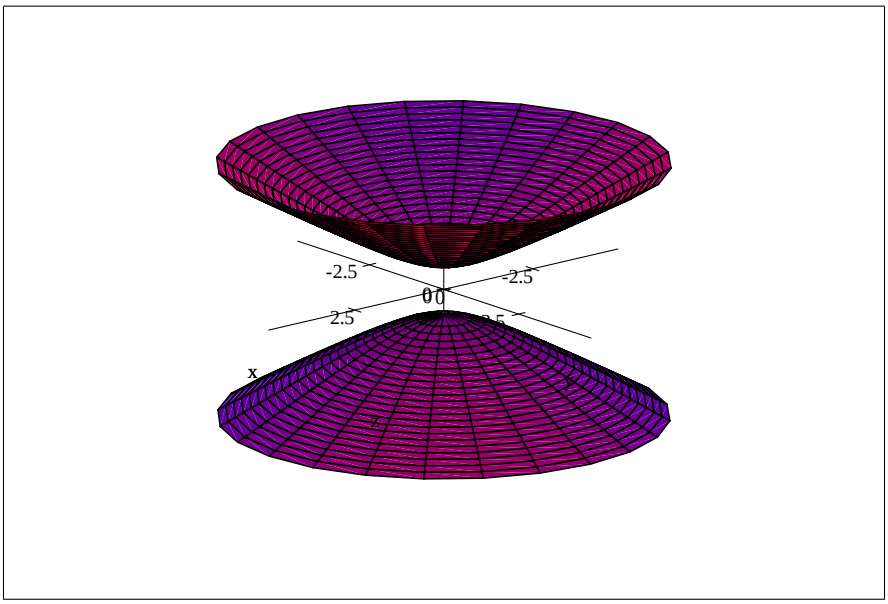
EXAMPLE 12. *Describe the level sets of* $f\,(x, y, z) = x^2+y^2+z^2$ *and* $f\,(x, y, z) = x^2 + y^2 - z^2$*.* - *The level sets of the first function are spheres. The second function has level set*

$$x^2 + y^2 - z^2 = c$$

If $c > 0$*, then we obtain hyperboloids with one shell:*



If $c < 0$, we find hyperboloids with two shells:



Usually, a level set in $\Re^2$ describes a curve in $\Re^2$. More generally, a level set in $\Re^n$ will be surface in $\Re^n$. The gradient of f will be orthogonal to those level sets.

EXAMPLE 13. *Find a vector orthogonal to the surface given by $x^3 - 2xy + y^3 + z^3 = 1$ at $[x, y, z] = [1, 1, 1]$. Also, find the tangent plane to the surface at this point.*

The gradient of $f(x, y, z) = x^3 - 2xy + y^3 + z^3$ is given by

$$\nabla f(x, y, z) = \left[3x^2 - 2y, 3y^2 - 2x, 3z^2\right]$$

If $x = y = z = 1$, then

$$\nabla f\,(1,1,1) = [1,1,3]$$

Hence the vector $[1,1,3]$ is perpendicular to the surface. The equation of the tangent plane is given by

$$[1,1,3] \cdot [x-1, y-1, z-1] = 0$$

or

$$x + y + 3z = 5$$

4. Review: Taylor's Formula

Let us start with a theorem from Math 9C:

THEOREM 2 (Taylor). *Let $f(x)$ be a function, defined on an open interval (a,b), and let $a < x_0 < b$. If $f(x)$ is m times continuously differentiable, then*

$$f(x) = f(x_0) + \frac{(x-x_0)}{1!} f^{(1)}(x_0) + \frac{(x-x_0)^2}{2!} f^{(2)}(x_0) + \cdots + \frac{(x-x_0)^{m-1}}{(m-1)!} f^{(m-1)}(x_0) + R_m$$

where

$$R_m = \frac{(x-x_0)^m}{m!} f^{(m)}(\xi)$$

for a suitable real number ξ between x_0 and x.

EXAMPLE 14. *Let $f(x) = x^3 - x^2 + x - 1$. Find the Taylor approximation of degree 3 at $x_0 = 1$. What is R_4?*

We compute:

$$\begin{aligned} f(x_0) &= f(1) = 0 \\ f'(x_0) &= f'(1) = 2 \\ f''(x_0) &= f''(1) = 4 \\ f'''(x_0) &= f'''(1) = 6 \\ f''''(x) &= 0 \end{aligned}$$

Hence

$$R_4 = 0$$

and

$$\begin{aligned} f(x) &= f(1) + \frac{f'(1)}{1!}(x-1) + \frac{f''(1)}{2!}(x-1)^2 + \frac{f'''(1)}{3!}(x-1)^3 + R_4 \\ &= 2(x-1) + 2(x-1)^2 + (x-1)^3 \end{aligned}$$

In order to generalize this formula to n dimensions, it is convenient in introduce Landau's symbol O:

DEFINITION 2. *Let $f, g : D \subseteq \Re^n \to \Re^m$ be functions, and assume that for a certain fixed value of $\rho > 0$, $\|\vec{x}\| < \rho$ implies that $\vec{x} \in D$. We write $f(\vec{x}) = O(g(\vec{x}))$ to mean that the quotient $\frac{\|f(\vec{x})\|}{\|g(\vec{x})\|}$ is bounded near $\vec{0}$, i.e. there is a number $K > 0$ and a value of δ with $0 < \delta \leq \rho$ so that $\|\vec{x}\| < \delta$ implies that $\frac{\|f(\vec{x})\|}{\|g(\vec{x})\|} \leq K$.*

EXAMPLE 15. *Let $f(\vec{x}) = \cos(\|\vec{x}\|)(\vec{x} - \vec{x}_0)$ and $g(x) = \vec{x} - \vec{x}_0$. Show that $f(\vec{x}) = O(g(\vec{x}))$.*

We now can rephrase Taylor's theorem as follows:

THEOREM 3 (Taylor). *Let $f(x)$ be a function, defined on an open interval (a,b), and let $a < x_0 < b$. If $f(x)$ is m times continuously differentiable, then*

$$f(x) = f(x_0) + \frac{(x-x_0)}{1!} f^{(1)}(x_0) + \frac{(x-x_0)^2}{2!} f^{(2)}(x_0) + \cdots + \frac{(x-x_0)^{m-1}}{(m-1)!} f^{(m-1)}(x_0) + O(x-x_0)^m$$

We need a version of Taylor's theorem that is valid for functions of several variables.

Let $f(\vec{x}) = f(x_1, \ldots, x_n) : D \to \Re$ be a real valued function, where $D \subseteq \Re^n$ is an open set. Let $\vec{x}_0 \in D$ be a point so that all partial derivatives $\frac{\partial f}{\partial x_i}(\vec{x}_0)$ exist.We define the gradient of f at $\vec{x}_0 \in D$ by

$$\nabla f(\vec{x}_0) = \left[\frac{\partial f}{\partial x_1}(\vec{x}_0), \ldots, \frac{\partial f}{\partial x_n}(\vec{x}_0) \right]$$

If all mixed partials $\frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{x}_0)$ also exist, then we define the Hessian matrix of f at $\vec{x}_0$ by

$$\begin{aligned} H(f)(\vec{x}_0) &= \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{x}_0) \right]_{1 \le i,j \le n} \\ &= \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_{ni} \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix} \end{aligned}$$

If all mixed partials exists and are continuous on an open domain containing $\vec{x}_0$, then

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{x}_0) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\vec{x}_0)$$

Hence, in this case the Hessian matrix is a symmetric matrix.

EXAMPLE 16. *Let* $f(x,y,z) = \frac{1}{\sqrt{x^2+y^2+z^2}}$. *Find the gradient and the Hessian matrix of* f *at* $\vec{x}_0 = (1,-1,3)$.

We compute:

$$\begin{aligned} \nabla f(x,y,z) &= \left[-\frac{x}{(x^2+y^2+z^2)^{\frac{3}{2}}}, \quad -\frac{y}{(x^2+y^2+z^2)^{\frac{3}{2}}}, \quad -\frac{z}{(x^2+y^2+z^2)^{\frac{3}{2}}} \right] \\ \nabla f(1,-1,3) &= \left[-\frac{1}{121}\sqrt{11}, \quad \frac{1}{121}\sqrt{11}, \quad -\frac{3}{121}\sqrt{11} \right] \end{aligned}$$

and

$$H(f) = \begin{bmatrix} -\frac{1}{(x^2+y^2+z^2)^{\frac{3}{2}}} + 3\frac{x^2}{(x^2+y^2+z^2)^{\frac{5}{2}}} & 3x\frac{y}{(x^2+y^2+z^2)^{\frac{5}{2}}} & 3x\frac{z}{(x^2+y^2+z^2)^{\frac{5}{2}}} \\ 3x\frac{y}{(x^2+y^2+z^2)^{\frac{5}{2}}} & -\frac{1}{(x^2+y^2+z^2)^{\frac{3}{2}}} + 3\frac{y^2}{(x^2+y^2+z^2)^{\frac{5}{2}}} & 3y\frac{z}{(x^2+y^2+z^2)^{\frac{5}{2}}} \\ 3x\frac{z}{(x^2+y^2+z^2)^{\frac{5}{2}}} & 3y\frac{z}{(x^2+y^2+z^2)^{\frac{5}{2}}} & -\frac{1}{(x^2+y^2+z^2)^{\frac{3}{2}}} + 3\frac{z^2}{(x^2+y^2+z^2)^{\frac{5}{2}}} \end{bmatrix}$$

$$H(f)(1,-1,3) = \begin{bmatrix} -\frac{8}{1331}\sqrt{11} & -\frac{3}{1331}\sqrt{11} & \frac{9}{1331}\sqrt{11} \\ -\frac{3}{1331}\sqrt{11} & -\frac{8}{1331}\sqrt{11} & -\frac{9}{1331}\sqrt{11} \\ \frac{9}{1331}\sqrt{11} & -\frac{9}{1331}\sqrt{11} & \frac{16}{1331}\sqrt{11} \end{bmatrix}$$

We will need the following special case of Taylor's Theorem:

THEOREM 4 (Taylor). *Let* $f(\vec{x})$ *be a real-valued function defined on* $D \subseteq \Re^n$. *Assume that for a certain value of* $\rho > 0$, *the open ball* $\{\vec{x} : \|\vec{x} - \bar{x}_0\| < \rho\}$ *belongs to the domain* D *of* f. *Assume all (mixed) partials of degree* 3 *or less exist and are continuous, and let* $H(f)(\vec{x}_0)$ *be the Hessian matrix of* f, *evaluated at* $\vec{x}_0$. *Then*

$$f(\vec{x}) = f(\vec{x}_0) + \nabla f(\vec{x}_0) \cdot (\vec{x} - \vec{x}_0) + \frac{1}{2!}(\vec{x} - \vec{x}_0)^T H(f)(\vec{x}_0)(\vec{x} - \vec{x}_0) + O\left(\|\vec{x} - \vec{x}_0\|^3\right)$$

EXAMPLE 17. *Use Taylor's theorem to find an approximation of degree* 2 *of* $f(\vec{x}) = \exp\left(x^2+y^2\right)$ *at* $(x_0, y_0) = (1,2)$.

If $f(x,y) = \exp\left(x^2+y^2\right)$ then

$$\begin{aligned} \nabla f(x,y) &= \left[\begin{array}{cc} 2xe^{x^2+y^2}, & 2ye^{x^2+y^2} \end{array}\right] \\ \nabla f(1,2) &= \left[\begin{array}{cc} 2e^5, & 4e^5 \end{array}\right] \end{aligned}$$

and

$$\begin{aligned} H(f)(x,y) &= \left[\begin{array}{cc} 2e^{x^2+y^2}+4x^2e^{x^2+y^2} & 4xye^{x^2+y^2} \\ 4xye^{x^2+y^2} & 2e^{x^2+y^2}+4y^2e^{x^2+y^2} \end{array}\right] \\ &= e^{x^2+y^2}\left[\begin{array}{cc} 2+4x^2 & 4xy \\ 4xy & 2+4y^2 \end{array}\right] \\ H(f)(1,2) &= 2e^5\left[\begin{array}{cc} 9 & 4 \\ 4 & 9 \end{array}\right] \end{aligned}$$

It follows that

$$\begin{aligned} \exp\left(x^2+y^2\right) &= e^5 + e^5[2,4]\left[\begin{array}{c} x-1 \\ y-2 \end{array}\right] + \frac{1}{2!}2e^5[x-1,y-2]\left[\begin{array}{cc} 9 & 4 \\ 4 & 9 \end{array}\right]\left[\begin{array}{c} x-1 \\ y-2 \end{array}\right] + \\ &\quad + O\left(\sqrt{(x-1)^2+(y-2)^2}^3\right) \\ &= e^5 + e^5(2x+4y-10) + e^5\left(8xy-44y-34x+9x^2+9y^2+61\right) + \\ &\quad + O\left(\sqrt{(x-1)^2+(y-2)^2}^3\right) \\ &= e^5\left(8xy-40y-32x+9x^2+9y^2+52\right) + O\left(\sqrt{(x-1)^2+(y-2)^2}^3\right) \end{aligned}$$

5. Local and Gobal Extreme Values

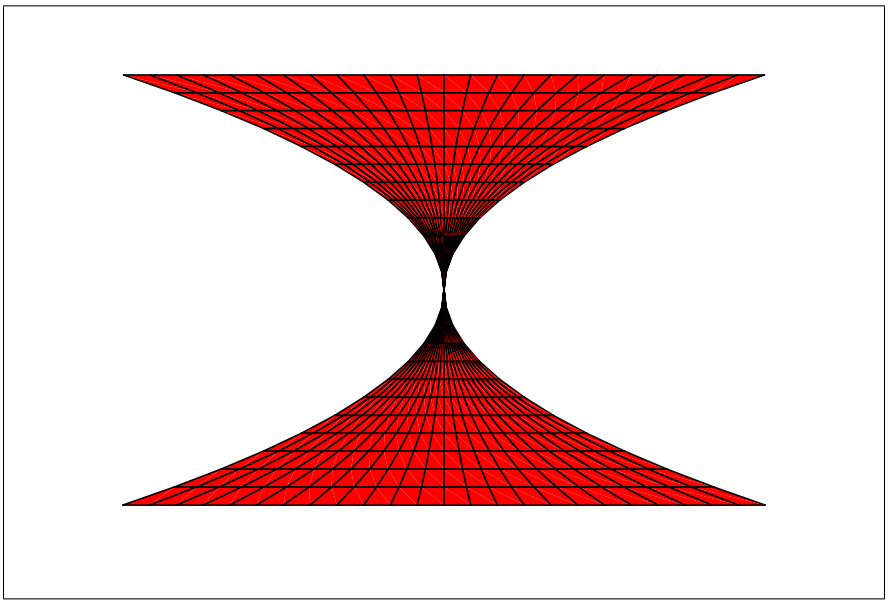
Let $f : D \subseteq \Re^n \to \Re$ be a real valued function. We would like to find the extreme values of f in D. For this purpose, we first of all assume that D is open:

DEFINITION 3. *Let $D \subseteq \Re^n$ be a subset. If for each $\vec{x} \in D$ there is a number $\varepsilon > 0$ so that all vectors $\vec{y} \in \Re^n$ satisfying $\|\vec{x} - \vec{y}\| < \varepsilon$ also belong to D, then D is called an open set.*

Open sets are very frequently defined by strict inequalities:

EXAMPLE 18. *The set $\{(x, y) : |y| < x^2\}$ is open. Sketch this set.*

The set $\{(x, y) : |y| < x^2\}$ consists of all points (x, y) not belonging to the shaded region:



This set is indeed open: Assume that $\vec{x} = (x_1, x_2) \in \{(x, y) : |y| < x^2\}$. Then $|x_2| < x_1^2$. Pick $0 < \varepsilon < |x_1|$ so that $|x_2| + \varepsilon < (|x_1| - \varepsilon)^2$. If $\vec{y} = (y_1, y_2)$ is a point satisfying $\|\vec{x} - \vec{y}\| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} < \varepsilon$, then

$$(x_1 - y_1)^2 + (x_2 - y_2)^2 < \varepsilon^2$$

Hence

$$\begin{aligned} |x_1 - y_1| &< \varepsilon \\ |x_2 - y_2| &< \varepsilon \end{aligned}$$

It follows that $|x_1| = |x_1 - y_1 + y_1| \leq |x_1 - y_1| + |y_1| < |y_1| + \varepsilon$, and therefore

$$|y_1| > |x_1| - \varepsilon$$

Similarly, $|y_2| = |y_2 - x_2 + x_2| \leq |y_2 - x_2| + |x_2| < |x_2| + \varepsilon$, i.e.

$$|y_2| < |x_2| + \varepsilon$$

We now conclude that

$$\begin{aligned} |y_2| &< |x_2| + \varepsilon \\ &< (|x_1| - \varepsilon)^2 \\ &< |y_1|^2 \end{aligned}$$

and hence the point (y_1, y_2) belongs to the set $\left\{(x, y) : |y| < x^2\right\}$.

DEFINITION 4. *Let $f : D \subseteq \Re^n \to \Re$ be a real valued function, defined on a set D.*

1. *We say that f has a local minimum at $\vec{x}_0 \in D$, if there is a number $\varepsilon > 0$ so that $\|\vec{x}_0 - \vec{y}\| < \varepsilon$ and $\vec{y} \in D$ imply that $f(y) \geq f(\vec{x}_0)$.*
2. *Similarly, f has a local maximum at $\vec{x}_0 \in D$, if there is a number $\varepsilon > 0$ so that $\|\vec{x}_0 - \vec{y}\| < \varepsilon$ and $\vec{y} \in D$ imply that $f(y) \leq f(\vec{x}_0)$.*
3. *If f has either a local maximum or a local minimum at $\vec{x}_0 \in D$, then f has a local extreme value at $\vec{x}_0$.*
4. *If $\vec{x}_0 \in D$, and if $f(\vec{x}) \geq f(\vec{x}_0)$ for all $x \in D$, then we say that f has a global mimimum at $\vec{x}_0$.*
5. *If $\vec{x}_0 \in D$, and if $f(\vec{x}) \leq f(\vec{x}_0)$ for all $x \in D$, then we say that f has a global maximum at $\vec{x}_0$.*
6. *Global maxima and global minima are also called global extreme values.*

DEFINITION 5. *Let $f : D \subseteq \Re^n \to \Re$ be a real valued function, defined on an open set D. Then $\vec{x}_0 \in D$ is a critical point for f, if either one of the partial derivative of f does not exist at $\vec{x}_0$, or else $grad(f)(\vec{x}_0) = \vec{0}$, i.e. all partial derivatives at $\vec{x}_0$ vanish.*

THEOREM 5. *If $f : D \subseteq \Re^n \to \Re$ be a real valued function, defined on an open set D, and if f has a local extreme value at $\vec{x}_0$, then $\vec{x}_0$ is a critical point of f.*

PROOF. For functions of one varible, this is a well-known theorem from Calculus, which can be easily extended to functions of several variables. If $f(x_1, \ldots, x_n)$ has a local extreme value at $\vec{x}_0 = (a_1, \ldots, a_n)$, then for each index i between 1 and n, the function $g(t)$ defined by

$$g(t) = f(a_1, \ldots, a_{i-1}, t, a_{i+1}, \ldots a_n)$$

has also a local extreme value at $t = a_i$. Hence either $g'(a_i)$ does not exist, or else $g'(a_i) = 0$. Since

$$g'(t) = \frac{\partial f}{\partial x_i}(a_1, \ldots, a_{i-1}, t, a_{i+1}, \ldots a_n)$$

the assertion of the theorem follows. □

EXAMPLE 19. *Find the local and global extreme values of*

$$f(x, y) = x^2 + xy + y^2 - x$$

subject to the constraint

$$\begin{aligned} |y| &< x^2 + 1 \\ x &< 2 \end{aligned}$$

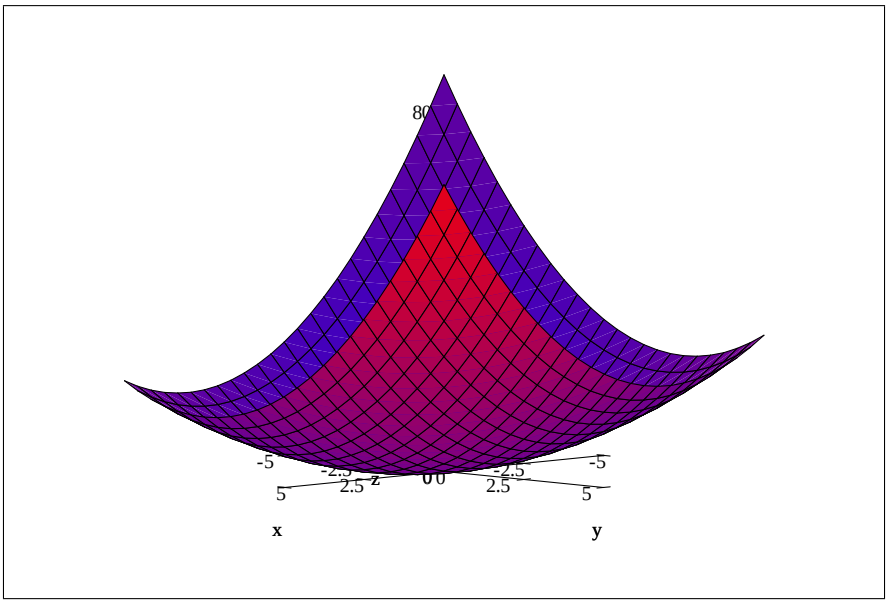
All partials have to vanish:

$$\begin{aligned} f_x &= 2x + y - 1 = 0 \\ f_y &= x + 2y = 0 \end{aligned}$$

This leads

$$\begin{aligned} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ \begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} \frac{2}{3} \\ -\frac{1}{3} \end{bmatrix} \end{aligned}$$

This solutions satisfies the constraint, so it is a feasable critical point. We plot $x^2 + xy + y^2 - x$



Our guess: The critical point presents a global minimum. To distinguish between maxima and minima, a second derivative test is needed.

EXAMPLE 20. *Find the local and global extreme values of*

$$x^4 + 2x^2y^2 + y^4 - 2x^2 + 1$$

subject to

$$-1 < x + y < 2$$

Again, all partial have to vanish:

$$\begin{aligned} 4x^3 + 4xy^2 - 4x &= 0 \\ 4x^2y + 4y^3 &= 0 \end{aligned}$$

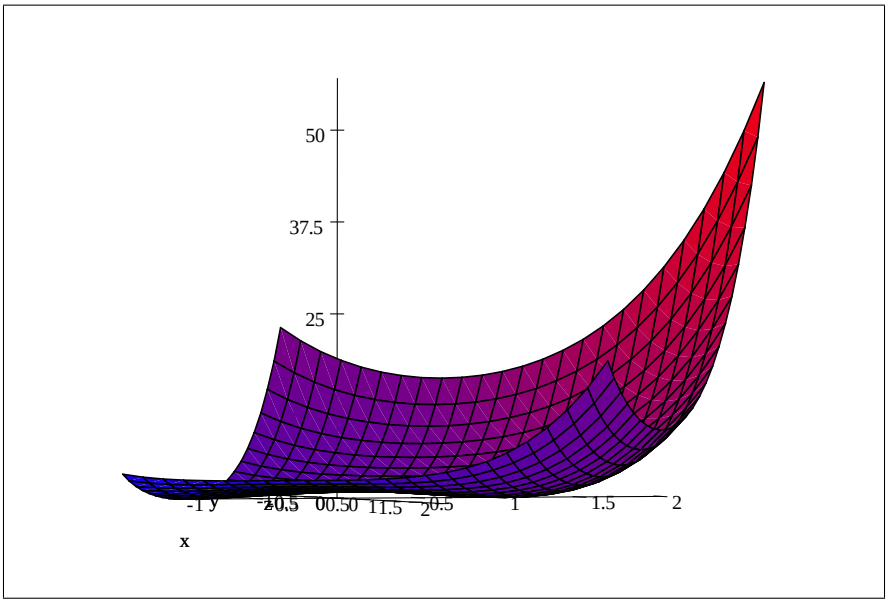
Factoring those equations gives

$$\begin{aligned} x\left(x^2 + y^2 - 1\right) &= 0 \\ y\left(x^2 + y^2\right) &= 0 \end{aligned}$$

If $x^2 + y^2 = 0$, then $x = y = 0$, and this is indeed a solution of both equations.

If $x^2 + y^2 \neq 0$, then $y = 0$, and $x \neq 0$. Hence $x\left(x^2 - 1\right) = 0$ implies that $x^2 = 1$, i.e. $x \pm 1$.

Candidates for critical points are $(0, 0)$, $(1, 0)$, and $(-1, 0)$. Only the first two points satisfy the constraint, i.e the critical points are $(0, 0)$ and $(1, 0)$. We plot the function $x^4 + 2x^2y^2 + y^4 - 2x^2 + 1$:



Again, we need a test deciding whether a critical point is a local minimum or a local maximum.

6. Local Extreme Values for Functions of One Variable

As we have seen in the previous section, local extreme values can only occur at critical points. In this section, we will review some additional tests from Calculus for functions of one variable. New might be the fact that these additional tests can be derived from Taylor's Formula.

THEOREM 6. *Let $f(x)$ be a function that is m times continuously differentiable on an open interval (a,b). Let c be a critical point of f in (a,b) so that the first $m-1$ derivatives of f vanish at c:*

$$f'(c) = f''(c) = \cdots = f^{(m-1)}(c) = 0.$$

Assume that $f^{(m)}(c) \neq 0$.

1. *If m is odd, then f does not have a local extreme value at $x = c$.*
2. *If m is even and if $f^{(m)}(c) > 0$, then f has a local minimum at $x = c$.*
3. *If m is even and if $f^{(m)}(c) < 0$, then f has a local maximum at $x = c$.*

PROOF. Using Taylor's formula, we obtain

$$f(x) = f(c) + \frac{(x-c)}{1!} f^{(1)}(c) + \frac{(x-c)^2}{2!} f^{(2)}(c) + \cdots + \frac{(x-c)^{m-1}}{(m-1)!} f^{(m-1)}(c) + R_m$$

where

$$R_m = \frac{(x-c)^m}{m!} f^{(m)}(\xi)$$

for a suitable real number ξ between c and x. Since $f'(c) = f''(c) = \cdots = f^{(m-1)}(c) = 0$, the expression reduces to

$$f(x) = f(c) + \frac{(x-c)^m}{m!} f^{(m)}(\xi)$$

Since $f^{(m)}(x)$ is a continuous function of x and since $f^{(m)}(c) \neq 0$, we can find a number $\varepsilon > 0$ so that $|x-c| < \varepsilon$ implies that $f^{(m)}(c) \neq 0$. In particular, we have either $f^{(m)}(x) > 0$ or $f^{(m)}(x) < 0$ for all values of x with $|x-c| < \varepsilon$. Since ξ is between x and c, this also implies that either $f^{(m)}(\xi) > 0$ or $f^{(m)}(\xi) < 0$ for all such values of ξ.

Assume first that m is odd. Then $f^{(m)}(\xi)$ does not change its sign while $(x-c)^m$ has a change of sign at $x = c$. As a consequence, the remainder $\frac{(x-c)^m}{m!} f^{(m)}(\xi)$ changes its sign $x = c$, and therefore $f(x)$ is greater than $f(c)$ for some values of x near c and smaller for some other values of x near c. Hence $f(x)$ cannot have a local extreme value at $x = c$.

Next, we assume that m is even. In this case, the function $(x-c)^m$ is always positive, at can be 0 only for $x = c$. Hence the signs of $\frac{(x-c)^m}{m!} f^{(m)}(\xi)$ and $f^{(m)}(c)$ agree as long as $|x-c| < \varepsilon$.

If $f^{(m)}(c) > 0$, then $\frac{(x-c)^m}{m!} f^{(m)}(\xi) > 0$ for values of x with $0 < |x-c| < \varepsilon$. For those values of x we obtain

$$\begin{aligned} f(x) &= f(c) + \frac{(x-c)^m}{m!} f^{(m)}(\xi) \\ &< f(c) \end{aligned}$$

and therefore f has a local minimum at $x = c$.

Similarly, if $f^{(m)}(c) < 0$, we find that $\frac{(x-c)^m}{m!} f^{(m)}(\xi) < 0$ for all values of x with $0 < |x-c| < \varepsilon$, and therefore

$$\begin{aligned} f(x) &= f(c) + \frac{(x-c)^m}{m!} f^{(m)}(\xi) \\ &> f(c) \end{aligned}$$

In this case, the function f has a local maximum at $x = c$. □

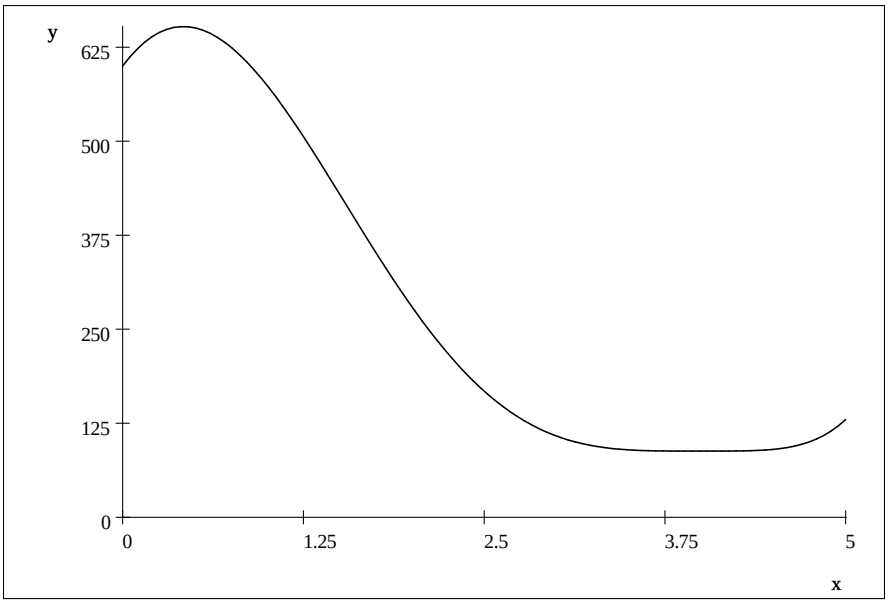
DEFINITION 6. *If f is differentiable on (a, b) and if $f'(c) = 0$, but f has not a local extreme value at $x = c$, then we say that f has a saddle point at $x = c$.*

EXAMPLE 21. *The function $f(x) = 256x - 320x^2 + 50x^4 - 13x^5 + x^6 + 600$ has a critical point at $x = 4$. Does f have a local maximum, a local minimum or a saddle point at $x = 4$?*

We compute:

$$\begin{aligned} f'(x) &= 200x^3 - 640x - 65x^4 + 6x^5 + 256 \\ f'(4) &= 0 \\ f''(x) &= 600x^2 - 260x^3 + 30x^4 - 640 \\ f''(4) &= 0 \\ f'''(x) &= 1200x - 780x^2 + 120x^3 \\ f'''(4) &= 0 \\ f''''(x) &= 360x^2 - 1560x + 1200 \\ f''''(4) &= 720 \end{aligned}$$

The forth derivative is the first higher derivative that is different from 0. Since 4 is an even number, and since $f''''(4) > 0$, the function has a local minimum at $x = 4$.



7. Local Extreme Values for Functions of Several Variables

In order to discuss local extreme values for functions of several variables, we need some additional tools from linear algebra:

DEFINITION 7. *Let A be a symmetric real $n \times n$ - matrix.*

(1) *If $\vec{x}^T A\vec{x} > 0$ for all non-zero vectors $\vec{0} \neq \vec{x} \in \Re^n$, then A is called positive definite.*
(2) *If $\vec{x}^T A\vec{x} < 0$ for all non-zero vectors $\vec{0} \neq \vec{x} \in \Re^n$, then A is called negative definite.*

THEOREM 7. *Let A be a symmetric real matrix*

(1) *A is positive definite if and only if all eigenvalues of A are strictly positive. If λ and Λ are the smallest and largest eigenvalue of A, respectively, then for each vector $\vec{x} \in \Re^n$ we have*

$$\lambda \|\vec{x}\|^2 \leq \vec{x}^T A\vec{x} \leq \Lambda \|\vec{x}\|^2$$

(2) *Similarly, A is negative definite if and only if all eigenvalues of A are strictly negative. If λ and Λ are the smallest and largest eigenvalue of A, respectively, then for each vector $\vec{x} \in \Re^n$ we have*

$$\lambda \|\vec{x}\|^2 \geq \vec{x}^T A\vec{x} \geq \Lambda \|\vec{x}\|^2$$

PROOF. Again, We pick an orthonormal basis B of eigenvectors of A. We write all the vectors of B into the columns of a matrix S. Then S is an orthogonal matrix (i.e. $S^T = S^{-1}$) and $S^T AS$ is a diagonal matrix having the eigenvalues of A on the diagonal:

$$S^T AS = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n \end{bmatrix}$$

Every orthogonal matrix preserves the Euclidean length of vectors:

$$\|S\vec{a}\| = \|\vec{a}\|$$

Now let $\vec{x} \in \Re^n$. We define $\vec{y} = S^{-1}\vec{x}$. Then

$$\begin{aligned} \vec{x}^T A\vec{x} &= (S\vec{y})^T A (S\vec{y}) \\ &= \vec{y}^T \left(S^T AS\right) \vec{y} \\ &= \vec{y}^T \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n \end{bmatrix} \vec{y} \\ &= \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 \end{aligned}$$

If all eigenvalues are positive, then the last expression is positive, hence $\vec{x}^T A\vec{x} > 0$ for all vectors $\vec{x}$. Moreover, if λ and Λ are the smallest and largest eigenvalues,

respectively, then

$$\begin{aligned}
\lambda \|\vec{x}\|^2 &= \lambda \|S\vec{y}\|^2 \\
&= \lambda \|\vec{y}\|^2 \\
&= \lambda \left(y_1^2 + \cdots + y_n^2\right) \\
&\leq \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 \\
&= \vec{x}^T A \vec{x} \\
&\leq \Lambda \left(y_1^2 + \cdots + y_n^2\right) \\
&= \Lambda \|\vec{y}\|^2 \\
&= \Lambda \|S\vec{y}\|^2 \\
&= \Lambda \|\vec{x}\|^2
\end{aligned}$$

This proves the first assertion. The proof of the second statement is similar. □

THEOREM 8. *The matrix*

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

is positive definite if and only if

$$a > 0 \text{ and } ac > b^2$$

Similarly, A is negative definite if and only if

$$a < 0 \text{ and } ac > b^2$$

PROOF. The eigenvalues of A are the solutions of

$$\det \begin{bmatrix} a - \lambda & b \\ b & c - \lambda \end{bmatrix} = 0$$

i.e. the solutions of the quadratic equation

$$(a - \lambda)(c - \lambda) - b^2 = 0$$

Since

$$\lambda^2 - (a + c)\lambda + ac - b^2 = 0$$

implies that

$$\lambda_{1,2} = \frac{1}{2}\left(a + c \pm \sqrt{(a - c)^2 + 4b^2}\right)$$

both values of λ are strictly positive if and only if

$$a + c > \sqrt{(a - c)^2 + 4b^2}$$

Hence the conditions

$$\begin{aligned}
a + c &\geq 0 \text{ and} \\
(a + c)^2 &> (a - c)^2 + 4b^2
\end{aligned}$$

have to hold simultaneously. We can rephrase this as

$$\begin{aligned}
a + c &\geq 0 \text{ and} \\
ac &> b^2
\end{aligned}$$

Especially, a and c have to have the same sign. Since $a+c \geq 0$, both have to be positive. Therefore the statement

$$\begin{aligned} a &> 0 \text{ and} \\ ac &> b^2 \end{aligned}$$

is equivalent to the fact that the matrix A is positive definite.

The proof of the second statement is similar. □

The concept of definite matrices is useful if we would like to classify local extreme values for function of several variables:

THEOREM 9. *Let $f : D \subseteq \Re^n \to \Re$ be a real valued function, defined on an open set D. Assume that f is three times differentiable, and that $\vec{x}_0$ is critical point of f. Let $H(f)(\vec{x}_0)$ be the Hessian matrix of f.*

1. *If $H(f)(\vec{x}_0)$ is positive definite, i.e. if all eigenvalues of $H(f)(\vec{x}_0)$ are strictly positive, then f has a local minimum at $\vec{x}_0$.*
2. *If $-H(f)(\vec{x}_0)$ is positive definite, i.e. if all eigenvalues of $H(f)(\vec{x}_0)$ are strictly negative, then f has a local maximum at $\vec{x}_0$.*
3. *If $H(f)(\vec{x}_0)$ has strictly positive as well as strictly negative eigenvalues, then f has neither a local maximum nor a local minimum at $\vec{x}_0$*

PROOF. We know from Taylor's theorem that

$$f(\vec{x}) = f(\vec{x}_0) + \nabla f(\vec{x}_0) \cdot (\vec{x}-\vec{x}_0) + \frac{1}{2!}(\vec{x}-\vec{x}_0)^T H(f)(\vec{x}_0)(\vec{x}-\vec{x}_0) + O\left(\|\vec{x}-\vec{x}_0\|^3\right)$$

Since $\vec{x}_0$ is a critical point, the gradient of f at $\vec{x}_0$ vanishes, hence

$$f(\vec{x}) = f(\vec{x}_0) + \frac{1}{2!}(\vec{x}-\vec{x}_0)^T H(f)(\vec{x}_0)(\vec{x}-\vec{x}_0) + O\left(\|\vec{x}-\vec{x}_0\|^3\right)$$

Assume that $H(f)(\vec{x}_0)$ is positive definite, and let λ be the smallest eigenvalue of $H(f)(\vec{x}_0)$. We know from the previous result that

$$\frac{1}{2!}(\vec{x}-\vec{x}_0)^T H(f)(\vec{x}_0)(\vec{x}-\vec{x}_0) \geq \frac{1}{2}\lambda \|\vec{x}-\vec{x}_0\|^2$$

Similarly, there is a constant M so that so that

$$\left| f(\vec{x}) - f(\vec{x}_0) - \frac{1}{2!}(\vec{x}-\vec{x}_0)^T H(f)(\vec{x}_0)(\vec{x}-\vec{x}_0) \right| \leq M \|\vec{x}-\vec{x}_0\|^3$$

as long as $\|\vec{x}-\vec{x}_0\|$ is sufficiently small. Hence for values of $\vec{x}$ close to $\vec{x}_0$ we find that

$$-M\|\vec{x}-\vec{x}_0\|^3 \leq f(\vec{x}) - f(\vec{x}_0) - \frac{1}{2!}(\vec{x}-\vec{x}_0)^T H(f)(\vec{x}_0)(\vec{x}-\vec{x}_0)$$

If we allow only values of $\vec{x}$ so that

$$\|\vec{x}-\vec{x}_0\| < \frac{\lambda}{2M}$$

then

$$\begin{aligned} 0 &\leq \left(\frac{1}{2}\lambda - M\left\|\vec{x}-\vec{x}_0\right\|\right)\left\|\vec{x}-\vec{x}_0\right\|^2 \\ &= \frac{1}{2}\lambda\left\|\vec{x}-\vec{x}_0\right\|^2 - M\left\|\vec{x}-\vec{x}_0\right\|^3 \\ &\leq \frac{1}{2!}\left(\vec{x}-\vec{x}_0\right)^T H\left(f\right)\left(\vec{x}_0\right)\left(\vec{x}-\vec{x}_0\right) - M\left\|\vec{x}-\vec{x}_0\right\|^3 \\ &\leq f\left(\vec{x}\right) - f\left(\vec{x}_0\right) \end{aligned}$$

For those values of $\vec{x}$ it follows that $f\left(\vec{x}_0\right) \leq f\left(\vec{x}\right)$. Hence f has a local minimum at $\vec{x}_0$.

Similarly, we can show that f has a local maximum at critical points for which $H\left(f\right)\left(\vec{x}_0\right)$ is negative definite. □

Note that the theorem makes no prediction for cases where the Hessian has 0 as an eigenvalue.

For functions of two variables, we have the following result:

COROLLARY 1. *If $f : D \subseteq \Re^2 \to \Re$ is a three times differentiable and that (x_0, y_0) is a critical point of f.*

1. *If $f_{xx}\left(x_0,y_0\right) f_{yy}\left(x_0,y_0\right) > f_{xy}\left(x_0,y_0\right)^2$, then f has a local extreme value at (x_0, y_0).*
   (a) *If $f_{xx}\left(x_0,y_0\right) < 0$, then f then this local extreme value is a local maximum..*
   (b) *If $f_{xx}\left(x_0,y_0\right) > 0$, then f has a local minimum at $\left(x_0,y_0\right)$.*
2. *If $f_{xx}\left(x_0,y_0\right) f_{yy}\left(x_0,y_0\right) < f_{xy}\left(x_0,y_0\right)^2$, then f has a saddle point at (x_0, y_0)*

EXAMPLE 22. *Find all local maxima and local minima for the following functions*

1. $f\left(x,y\right) = x^4 + y^4$,
2. $f\left(x,y\right) = x^2 - y^2$,.
3. $f\left(x,y,z\right) = x^2 + y^2 + z^2 - xyz$.

EXAMPLE 23. *For the following functions and constraints, find all local extreme values.*

1. $f\left(x,y\right) = x^2 + xy + y^2 - 1$ *subject to the constraints* $|y| < x^2 + 1$, $x < 2$.
2. $f\left(x,y\right) = x^4 + 2x^2y^2 + y^4 - 2x^2 + 1$, *subject to* $-1 < x + y < 2$.

8. Equality Constraints

Many problems in optimization theory are of the form

$$\begin{aligned} \text{Maximize (or minimize) } y &= f(\vec{x}) \\ \text{subject to the condition } g(\vec{x}) &= \vec{0} \end{aligned}$$

where $f : D \subseteq \Re^n \to \Re$ and $g : E \subseteq \Re^n \to \Re^m$ are functions of n variables $\vec{x} = [x_1, \ldots, x_n]^T \in \Re^n$.Of course, we have to assume that the domain of f contains the solution set $g(\vec{x}) = \vec{0}$.

Since the range of g is contained in $\Re^m$, there are m function $g_1, \ldots, g_m : E \subseteq \Re^n \to \Re$ so that $g(\vec{x}) = [g_1(\vec{x}), \ldots, g_m(\vec{x})]^T$. Hence the optimization problem can be rewritten as

$$\begin{aligned} \text{Maximize (or minimize) } y &= f(\vec{x}) \\ \text{subject to the condition } g_1(\vec{x}) &= 0 \\ g_2(\vec{x}) &= 0 \\ &\vdots \\ g_m(\vec{x}) &= 0 \end{aligned}$$

EXAMPLE 24. *Let $f(x, y) = x^2 - xy + y^2$. Find the maximum and the minimum of $f(x, y)$ under the condition that $x - y = 5$.*

In this case, the function g is given by $g(x, y) = x - y - 5$, The equation $g(x, y) = x - y - 5 = 0$ can be solved for one of the variables:

$$y = x - 5$$

Hence $f(x, y) = x^3 - xy + y^2 = x^2 - x(x - 5) + (x - 5)^2 = x^2 - 5x + 25$. Since

$$\frac{d}{dx}\left(x^2 - 5x + 25\right) = 2x - 5$$

this function has only one critical point, namely at $x = \frac{2}{5}$. At this critical point, the function has a minimum, namely $f\left(\frac{2}{5}, \frac{2}{5} - 5\right) = \frac{579}{25}$. There is no maximum.

EXAMPLE 25. *Let $f(x, y, z) = x^2 + 2y^2 + 3z^2$. Find the extreme values of this function under the conditions that*

$$\begin{aligned} x^2 + y^2 + z^2 &= 1 \\ x + y + z &= 1 \end{aligned}$$

In this case, the constraints are given by two functions g_1 and g_2

$$\begin{aligned} g_1(x, y, z) &= x^2 + y^2 + z^2 - 1 \\ g_2(x, y, z) &= x + y + z - 1 \end{aligned}$$

We can use the two constraints to express two of the variables in terms of the third variable. Indeed, the simultaneous solutions of $g_1(x, y, z) = 0$ and $g_2(x, y, z) = 1$ represent the intersect of the unit sphere $x^2 + y^2 + z^2 = 1$ with the plane $x + y + z = 1$, and this intersection is a circle in $\Re^3$. We compute:

$$\begin{aligned} z &= 1 - x - y \\ x^2 + y^2 + (1 - x - y)^2 &= 1 \end{aligned}$$

The last equation leads to

$$xy - y - x + x^2 + y^2 = 0$$

Solving this equation for x gives

$$x_{1,2} = \frac{1}{2} - \frac{1}{2}y \pm \frac{1}{2}\sqrt{2y - 3y^2 + 1}$$

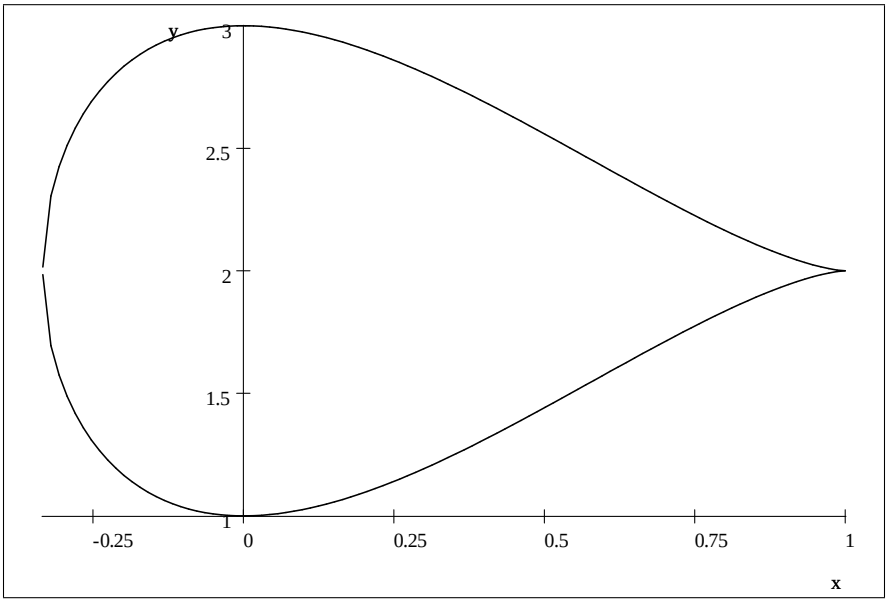
Hence

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \frac{1}{2} - \frac{1}{2}y \pm \frac{1}{2}\sqrt{2y - 3y^2 + 1} \\ y \\ \frac{1}{2} - \frac{1}{2}y \mp \frac{1}{2}\sqrt{2y - 3y^2 + 1} \end{bmatrix}$$

This leads to

$$\begin{aligned} f(x,y,z) &= x^2 + 2y^2 + 3z^2 \\ &= \left(\frac{1}{2} - \frac{1}{2}y \pm \frac{1}{2}\sqrt{2y - 3y^2 + 1}\right)^2 + 2y^2 + 3\left(\frac{1}{2} - \frac{1}{2}y \mp \frac{1}{2}\sqrt{2y - 3y^2 + 1}\right)^2 \\ &= \pm(y-1)\sqrt{2y - 3y^2 + 1} + 2 \\ &= \pm(y-1)\sqrt{(3y+1)(1-y)} + 2 \\ &= \mp(1-y)^{3/2}(3y+1)^{1/2} + 2 \end{aligned}$$

Depending on the choice of the sign, we end up with two possible functions. Both functions are defined for $-\frac{1}{3} \leq y \leq 1$



The candidates for extreme values of the functions are the endpoints of the interval for y, namely $y = -\frac{1}{3}$, $y = 1$ and the solutions of

$$\frac{d}{dy}\left(\pm(1-y)^{3/2}(3y+1)^{1/2} + 2\right) = 0$$

Since

$$\frac{d}{dy}\left(\pm(1-y)^{3/2}(3y+1)^{1/2} + 2\right) = \pm\left(\frac{3}{2\sqrt{3y+1}}\left(\sqrt{1-y} - y\sqrt{1-y}\right) - \frac{3}{2}\sqrt{1-y}\sqrt{3y+1}\right)$$

we find that

$$\begin{aligned}\frac{3}{2\sqrt{3y+1}}\left(\sqrt{1-y}-y\sqrt{1-y}\right) &= \frac{3}{2}\sqrt{1-y}\sqrt{3y+1}\\ (1-y) &= (3y+1)\end{aligned}$$

Hence $y=0$. At $y=0$, we find that

$$\pm(1-y)^{3/2}(3y+1)^{1/2}+2=\pm 1+2$$

We obtain the following candidates for extreme values:

(1) $y=-\frac{1}{3}$. Then

$$\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}\frac{1}{2}-\frac{1}{2}y\pm\frac{1}{2}\sqrt{2y-3y^2+1}\\ y\\ \frac{1}{2}-\frac{1}{2}y\mp\frac{1}{2}\sqrt{2y-3y^2+1}\end{bmatrix}=\begin{bmatrix}\frac{2}{3}\\ -\frac{1}{3}\\ \frac{2}{3}\end{bmatrix}$$

In this case, $f(x,y,z)=f\left(\frac{2}{3},-\frac{1}{3},\frac{2}{3}\right)=2$

(2) $y=1$ implies

$$\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}\frac{1}{2}-\frac{1}{2}y\pm\frac{1}{2}\sqrt{2y-3y^2+1}\\ y\\ \frac{1}{2}-\frac{1}{2}y\mp\frac{1}{2}\sqrt{2y-3y^2+1}\end{bmatrix}=\begin{bmatrix}0\\1\\0\end{bmatrix}$$

The function value at this point is that $f(0,1,0)=2$

(3) $y=0$. Then

$$\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}\frac{1}{2}-\frac{1}{2}y\pm\frac{1}{2}\sqrt{2y-3y^2+1}\\ y\\ \frac{1}{2}-\frac{1}{2}y\mp\frac{1}{2}\sqrt{2y-3y^2+1}\end{bmatrix}=\begin{bmatrix}\frac{1}{2}\pm\frac{1}{2}\\ 0\\ \frac{1}{2}\mp\frac{1}{2}\end{bmatrix}$$

So we obtain two critical points namely $[x,y,z]=[1,0,0]$ with function value $f(1,0,0)=1$ and $[x,y,z]=[0,0,1]$ with function value $f(0,0,1)=3$.

We conclude that the maximum is 3, obtained at $[0,0,1]$, and the minimum is 1, obtained at $[1,0,0]$.

In the next section, we will discuss the Jacobian method, with automates many of the steps we used in the previous example.

9. Standard Forms

We will be discussing problems of the form:
Maximize

$$f(\vec{x}) = \vec{a} \cdot \vec{x}$$

subject to the conditions

$$\begin{aligned} \vec{b}_1 \cdot \vec{x} &\leq r_1 \\ &\vdots \\ \vec{b}_k \cdot \vec{x} &\leq r_n \\ x_1 &\geq 0 \\ &\vdots \\ x_n &\geq 0 \end{aligned}$$

where $\vec{x} = (x_1, \ldots, x_n), \vec{a}, \vec{b}_1, \ldots, \vec{b}_k \in \Re^n$.

The function $f(\vec{x})$ is linear and therefore convex. The constraint describes a closed convex set. If this set is also bounded, then we know that convex functions defined on convex sets obtain their maximum values at extreme points So the strategy would be to identify all extreme points of this polytope, evaluate the linear function at those extreme points, and then take the point that yields the maximum value.

9.0.1. *Dealing the problem of finding minimum values:* If we are asked to minimize $f(\vec{x}) = \vec{a} \cdot \vec{x}$, we might as well maximize $-f(\vec{x}) = (-\vec{a}) \cdot \vec{x}$.

9.0.2. *Dealing with* $\geq$:. If one of the constraints is of the form

$$\vec{b} \cdot \vec{x} \geq r$$

we multiply the inequality by -1, and obtain

$$\left(-\vec{b}\right) \cdot \vec{x} \leq -r$$

9.0.3. *Replacing inequalities by equalities:* Every inequality

$$\vec{b} \cdot \vec{x} \leq r$$

is replaced by

$$\begin{aligned} \vec{b} \cdot \vec{x} + y &= r \\ y &\geq 0 \end{aligned}$$

9.0.4. *Dealing with negative right-hand-sides:* Each equation of the form

$$\vec{b} \cdot \vec{x} = r$$

where $r < 0$ is replaced by

$$\left(-\vec{b}\right) \cdot \vec{x} = -r$$

In this way, we produce a *standard form* of a linear programming problem:

Maximize the objective function

$$f(\vec{x}) = \vec{a} \cdot \vec{x}$$

subject to the constraints

$$\begin{aligned} B\vec{x} &= \vec{b} \\ \vec{x} &\geq 0 \end{aligned}$$

where

$$\vec{b} \geq \vec{0}$$

The features of a problem in standard form are:

(1) The objective function is to be maximized.
(2) All constraints except the nonnegativity conditions are strict equations.
(3) The independent variables are all nonnegative.
(4) The constant to the right of each equality sign in each constraint is nonnegative.

Later, when we discuss duality, we will also use different standarized forms of linear programming problems.

EXAMPLE 26. *Consider the problem:*

Minimize

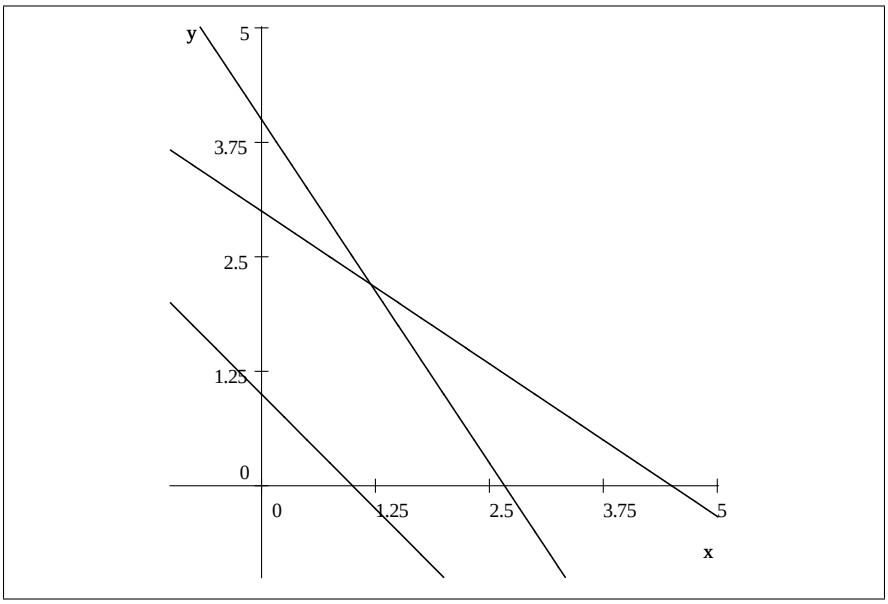
$$f(x, y) = 4x - 2y$$

subject to

$$\begin{aligned} 3x + 2y &\leq 8 \\ 2x + 3y &\leq 9 \\ x + y &\geq 1 \\ x &\geq 0 \\ y &\geq 0 \end{aligned}$$

Find the corresponding standard form for this problem.

Even though this is not part of the problem, we first plot the feasible region:



First, we would like to find the corresponding problem for finding maximum values:
Maximize

$$f(x,y) = -4x + 2y$$

subject to

$$\begin{aligned} 3x + 2y &\leq 8 \\ 2x + 3y &\leq 9 \\ x + y &\geq 1 \\ x &\geq 0 \\ y &\geq 0 \end{aligned}$$

Then we convert $x + y \geq 1$ into $-x - y \leq -1$:
Maximize

$$f(x,y) = -4x + 2y$$

subject to

$$\begin{aligned} 3x + 2y &\leq 8 \\ 2x + 3y &\leq 9 \\ -x - y &\leq -1 \\ x &\geq 0 \\ y &\geq 0 \end{aligned}$$

Next, we we add slack variables:
Maximize

$$f(x,y) = -4x + 2y$$

subject to

$$\begin{array}{rcrcrcrcrcl} 3x & + & 2y & + & u & & & & & = & 8 \\ 2x & + & 3y & & & + & v & & & = & 9 \\ -x & - & y & & & & & + & w & = & -1 \\ & & & & & & & & x & \geq & 0 \\ & & & & & & & & y & \geq & 0 \\ & & & & & & & & u & \geq & 0 \\ & & & & & & & & v & \geq & 0 \\ & & & & & & & & w & \geq & 0 \end{array}$$

The third equation need to be multiplied by -1:
Maximize

$$f(x,y) = -4x + 2y$$

subject to

$$\begin{array}{rcrcrcrcrcl} 3x & + & 2y & + & u & & & & & = & 8 \\ 2x & + & 3y & & & + & v & & & = & 9 \\ x & + & y & & & & & - & w & = & 1 \\ & & & & & & & & x & \geq & 0 \\ & & & & & & & & y & \geq & 0 \\ & & & & & & & & u & \geq & 0 \\ & & & & & & & & v & \geq & 0 \\ & & & & & & & & w & \geq & 0 \end{array}$$

Finally, the problem is written in matrix form:

Maximize

$$f(\vec{x}) = (-4, 2, 0, 0, 0) \begin{pmatrix} x \\ y \\ u \\ v \\ w \end{pmatrix}$$

subject to

$$\begin{pmatrix} 3 & 2 & 1 & 0 & 0 \\ 2 & 3 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ u \\ v \\ w \end{pmatrix} = \begin{pmatrix} 8 \\ 9 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} x \\ y \\ u \\ v \\ w \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

or:

Maximize

$$f(\vec{x}) = (-4, 2, 0, 0, 0) \cdot \vec{x}$$

subject to

$$\begin{pmatrix} 3 & 2 & 1 & 0 & 0 \\ 2 & 3 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & -1 \end{pmatrix} \vec{x} = \begin{pmatrix} 8 \\ 9 \\ 1 \end{pmatrix}$$
$$\vec{x} \geq \vec{0}$$

Note that the equations

$$\begin{pmatrix} 3 & 2 & 1 & 0 & 0 \\ 2 & 3 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & -1 \end{pmatrix} \vec{x} = \begin{pmatrix} 8 \\ 9 \\ 1 \end{pmatrix}$$
$$\vec{x} \geq \vec{0}$$

and

$$\begin{aligned} 3x + 2y &\leq 8 \\ 2x + 3y &\leq 9 \\ x + y &\geq 1 \\ x &\geq 0 \\ y &\geq 0 \end{aligned}$$

have exactly the same solutions. A bijection between both sets is given by the map

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \\ 8 - 3x - 2y \\ 9 - 2x - 3y \\ x + y - 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 8 \\ 9 \\ -1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -3 & -2 \\ -2 & -3 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

This map is the sum of a linear map and a constant, i.e. an affine map. Bijective affine maps preserve convex sets and their extreme points. Hence it does not matter whether we look for the extreme points of the original convex set or the extreme points of the new convex set - in our concrete example, the extreme points of the solutions of

$$\begin{pmatrix} 3 & 2 & 1 & 0 & 0 \\ 2 & 3 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & -1 \end{pmatrix} \vec{x} = \begin{pmatrix} 8 \\ 9 \\ 1 \end{pmatrix}$$
$$\vec{x} \geq \vec{0}$$

10. The Jacobian Method

In order to explain the Jacobian Method, we need the Implicit Function Theorem. Let $g : D \subseteq \Re^n \to \Re^m$ be a differentiable function, and let $\vec{x}_0 \in D$. Assume that the rank of the Jacobian matrix

$$g'(x_0) = \begin{bmatrix} \frac{\partial g_1}{\partial x_1}(\vec{x}_0) & \frac{\partial g_1}{\partial x_2}(\vec{x}_0) & \dots & \frac{\partial g_1}{\partial x_n}(\vec{x}_0) \\ \frac{\partial g_2}{\partial x_1}(\vec{x}_0) & \frac{\partial g_2}{\partial x_2}(\vec{x}_0) & \dots & \frac{\partial g_2}{\partial x_n}(\vec{x}_0) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(\vec{x}_0) & \frac{\partial g_m}{\partial x_2}(\vec{x}_0) & \dots & \frac{\partial g_m}{\partial x_n}(\vec{x}_0) \end{bmatrix}$$

is equal to m. Then there are indices $1 \leq i_1 < i_2 < \cdots < i_m \leq n$ so that the matrix

$$\begin{bmatrix} \frac{\partial g_1}{\partial x_{i_1}}(\vec{x}_0) & \frac{\partial g_1}{\partial x_{i_2}}(\vec{x}_0) & \dots & \frac{\partial g_1}{\partial x_{i_m}}(\vec{x}_0) \\ \frac{\partial g_2}{\partial x_{x_1}}(\vec{x}_0) & \frac{\partial g_2}{\partial x_{i_2}}(\vec{x}_0) & \dots & \frac{\partial g_2}{\partial x_{i_m}}(\vec{x}_0) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_{i_1}}(\vec{x}_0) & \frac{\partial g_m}{\partial x_{i_2}}(\vec{x}_0) & \dots & \frac{\partial g_m}{\partial x_{i_m}}(\vec{x}_0) \end{bmatrix}$$

is invertible. We will now reorder the variables so that $i_1 = 1$, $i_2 = 2, \dots, i_m = m$. Then we have

$$\vec{x} = [w_1, \dots, w_m, z_1, \dots, z_{m-n}]$$

and

$$g'(\vec{x}_0) = [J, C]$$

with

$$J = \begin{bmatrix} \frac{\partial g_1}{\partial w_1}(\vec{x}_0) & \frac{\partial g_1}{\partial w_2}(\vec{x}_0) & \dots & \frac{\partial g_1}{\partial w_m}(\vec{x}_0) \\ \frac{\partial g_2}{\partial w_1}(\vec{x}_0) & \frac{\partial g_2}{\partial w_2}(\vec{x}_0) & \dots & \frac{\partial g_2}{\partial w_m}(\vec{x}_0) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial w_1}(\vec{x}_0) & \frac{\partial g_m}{\partial w_2}(\vec{x}_0) & \dots & \frac{\partial g_m}{\partial w_m}(\vec{x}_0) \end{bmatrix}$$

$$C = \begin{bmatrix} \frac{\partial g_1}{\partial z_1}(\vec{x}_0) & \frac{\partial g_1}{\partial z_2}(\vec{x}_0) & \dots & \frac{\partial g_1}{\partial z_{n-m}}(\vec{x}_0) \\ \frac{\partial g_2}{\partial z_1}(\vec{x}_0) & \frac{\partial g_2}{\partial z_2}(\vec{x}_0) & \dots & \frac{\partial g_2}{\partial z_{n-m}}(\vec{x}_0) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial z_1}(\vec{x}_0) & \frac{\partial g_m}{\partial z_2}(\vec{x}_0) & \dots & \frac{\partial g_m}{\partial z_{n-m}}(\vec{x}_0) \end{bmatrix}$$

The variable $w_1, \dots, w_m$ are called the state variable, and the variables $y_1, \dots, y_{n-m}$ are called decision variables. The matrix J is also called the Jacobian matrix (even though this may conflict with the name for the matrix $g'(\vec{x})$, which in this context will be called the derivative matrix). The matrix C is called the control matrix.

It is always important to identify state variables and decision variables before attempting to solve a problem. The choice has to be made in such a way that the matrix J is invertible at $\vec{x}_0$.

THEOREM 10. *Let $D \subseteq \Re^m$ be an open set, and let $g : D \subseteq \Re^m \times \Re^{n-m} \to \Re^m$ be a differentiable function. Further, let $\vec{x}_0 = [\vec{w}_0, \vec{z}_0] \in D$ be a point so that*

$g(\vec{x}_0) = 0$. *Assume that the derivative matrix*

$$g'(x_0) = \begin{bmatrix} \frac{\partial g_1}{\partial x_1}(\vec{x}_0) & \frac{\partial g_1}{\partial x_2}(\vec{x}_0) & \dots & \frac{\partial g_1}{\partial x_n}(\vec{x}_0) \\ \frac{\partial g_2}{\partial x_1}(\vec{x}_0) & \frac{\partial g_2}{\partial x_2}(\vec{x}_0) & \dots & \frac{\partial g_2}{\partial x_n}(\vec{x}_0) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(\vec{x}_0) & \frac{\partial g_m}{\partial x_2}(\vec{x}_0) & \dots & \frac{\partial g_m}{\partial x_n}(\vec{x}_0) \end{bmatrix} = [J, C]$$

has the property that J is invertible. Then there a number $\varepsilon > 0$ and a uniquely determined function

$$\begin{aligned} w &= h(\vec{z}) \\ \begin{bmatrix} w_1 \\ \vdots \\ w_m \end{bmatrix} &= \begin{bmatrix} h_1(\vec{z}) \\ \vdots \\ h_m(\vec{z}) \end{bmatrix} \end{aligned}$$

so that

$$h(\vec{z}_0) = \vec{w}_0$$

and

$$g(h(\vec{z}), \vec{z}) = 0$$

for all values of $\vec{z}$ with $\|\vec{z} - \vec{z}_0\| < \varepsilon$. In addition, the number $\varepsilon > 0$ can be choosen so that for all values of $(\vec{w}, \vec{z})$ with $\|\vec{w} - \vec{w}_0\| < \varepsilon$ and $\|\vec{z} - \vec{z}_0\| < \varepsilon$ we have

$$g(\vec{w}, \vec{z}) = 0 \iff \vec{w} = h(\vec{z})$$

This theorem says that "we can solve for the state variables in terms of the decision variables."

We use this theorem in order to find critical points. Assume that we would like to

$$\begin{aligned} \text{Find the local extreme values of } y &= f(\vec{x}) \\ \text{subject to the condition } g(\vec{x}) &= \vec{0} \end{aligned}$$

where g is given as before. Assume that $\vec{x}_0$ is a point where a local extreme value occurs. We write

$$\begin{aligned} \vec{x}_0 &= [\vec{w}_0, \vec{z}_0] \\ g(\vec{x}_0) &= 0 \end{aligned}$$

Then, close to $\vec{x}_0$, we only have to consider points that meet the constraints, i.e. points that are of the form

$$\vec{x} = [h(\vec{z}), \vec{z}]^T$$

where h is a function that satisfies the properties of the implicit function theorem. Hence, there is a number $\varepsilon > 0$ so that only values of the form

$$y = f(h(\vec{z}), \vec{z})$$

with $\|\vec{z} - \vec{z}_0\| < \varepsilon$ need to be considered. It follows that at for a local extreme value the gradient of the function

$$f_1(\vec{z}) = f(h(\vec{z}), \vec{z})$$

vanishes:

$$\nabla f_1(\vec{w}) = \vec{0}$$

The chain rule implies that

$$\begin{aligned}\nabla f_1(\vec{z}) &= \nabla f(h(\vec{z}),\vec{z}) \\ &= \left[\frac{\partial f}{\partial w_1},\ldots,\frac{\partial f}{\partial w_m},\frac{\partial f}{\partial z_1},\ldots,\frac{\partial f}{\partial z_{n-m}}\right]\cdot\begin{bmatrix}\frac{\partial h_1}{\partial z_1} & \frac{\partial h_1}{\partial z_2} & \cdots & \frac{\partial h_1}{\partial z_{n-m}}\\ \frac{\partial h_2}{\partial z_1} & \frac{\partial h_2}{\partial z_2} & \cdots & \frac{\partial h_2}{\partial z_{n-m}}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial h_m}{\partial z_1} & \frac{\partial h_m}{\partial z_2} & \cdots & \frac{\partial h_m}{\partial z_{n-m}}\\ 1 & 0 & \ldots & 0\\ 0 & 1 & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & 1\end{bmatrix} \\ &= \left[\frac{\partial f}{\partial w_1},\ldots,\frac{\partial f}{\partial w_m},\frac{\partial f}{\partial z_1},\ldots,\frac{\partial f}{\partial z_{n-m}}\right]\cdot\begin{bmatrix}h'(\vec{z})\\ I\end{bmatrix}\end{aligned}$$

The derivative matrix of $h'(\vec{w})$ be computed from $g'(h(\vec{z}),\vec{z})$, using the chain rule again: Since $g(h(\vec{z}),\vec{z})=\vec{0}$, we find that

$$\begin{aligned}\vec{0} &= g'(h(\vec{z}),\vec{z}) \\ &= [J,C]\begin{bmatrix}h'(z)\\ I\end{bmatrix}\end{aligned}$$

i.e.

$$J\cdot h'(\vec{z})+C=0$$

or

$$h'(\vec{z})=-J^{-1}C$$

This leads to

$$\nabla f_1(\vec{z})=\left[\frac{\partial f}{\partial w_1},\ldots,\frac{\partial f}{\partial w_m},\frac{\partial f}{\partial z_1},\ldots,\frac{\partial f}{\partial z_{n-m}}\right]\cdot\begin{bmatrix}-J^{-1}C\\ I\end{bmatrix}$$

At a critical point, this gradient has to vanish:

$$\begin{aligned}\nabla f_1(\vec{z}) &= \vec{0} \\ \left[\frac{\partial f}{\partial w_1},\ldots,\frac{\partial f}{\partial w_m},\frac{\partial f}{\partial z_1},\ldots,\frac{\partial f}{\partial z_{n-m}}\right]\cdot\begin{bmatrix}-J^{-1}C\\ I\end{bmatrix} &= \vec{0}\end{aligned}$$

We use this method to repeat the last example of the previous section:

EXAMPLE 27. *Let $f(x,y,z)=x^2+2y^2+3z^2$. Find the extreme values of this function under the conditions that*

$$\begin{aligned}x^2+y^2+z^2 &= 1\\ x+y+z &= 1\end{aligned}$$

The function $g(x,y,z)$ is given by

$$g(x,y,z)=\begin{bmatrix}x^2+y^2+z^2-1\\ x+y+z-1\end{bmatrix}$$

First, we have to find the derivative matrix of g :

$$g'(x,y,z)=\begin{bmatrix}2x & 2y & 2z\\ 1 & 1 & 1\end{bmatrix}$$

Next, we have to find the state variables and the decision variable: x and y will be state variables, if the matrix

$$\begin{bmatrix} 2x & 2y \\ 1 & 1 \end{bmatrix}$$

is invertible, i.e. if $x \neq y$. We consider three cases:

Case 1. $x \neq y$. Then x and y are state variables and z a decision variable. Hence

$$\begin{aligned} J &= \begin{bmatrix} 2x & 2y \\ 1 & 1 \end{bmatrix} \\ C &= \begin{bmatrix} 2z \\ 1 \end{bmatrix} \\ J^{-1}C &= \begin{bmatrix} 2x & 2y \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 2z \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} \frac{y-z}{y-x} \\ \frac{z-x}{y-x} \end{bmatrix} \end{aligned}$$

and $\left[\frac{\partial f}{\partial w_1}, \ldots, \frac{\partial f}{\partial w_m}\right] J^{-1}C = \left[\frac{\partial f}{\partial z_1}, \ldots, \frac{\partial f}{\partial z_{n-m}}\right]$ leads to

$$\begin{aligned} \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right] \begin{bmatrix} \frac{y-z}{y-x} \\ \frac{z-x}{y-x} \end{bmatrix} &= \frac{\partial f}{\partial z} \\ [2x, 4y] \begin{bmatrix} \frac{y-z}{y-x} \\ \frac{z-x}{y-x} \end{bmatrix} &= 6z \end{aligned}$$

We obtain the equation

$$\begin{aligned} 2x(y-z) + 4y(z-x) - 6z(y-x) &= 0 \\ 2xz - xy - yz &= 0 \end{aligned}$$

Hence every critical point satisfies the three equations

$$\begin{aligned} x^2 + y^2 + z^2 &= 1 \\ x + y + z &= 1 \\ 2xz &= y(x+z) \end{aligned}$$

The solutions are: the three unit vectors:

$$\begin{aligned} [x, y, z] &= [1, 0, 0] \text{ with function value } f(1, 0, 0) = 1 \\ [x, y, z] &= [0, 1, 0] \text{ with function value } f(0, 1, 0) = 2 \\ [x, y, z] &= [0, 0, 1] \text{ with function value } f(0, 0, 1) = 3 \end{aligned}$$

Case 2 $y \neq z$. In this case, y and z are state variables and x is the decision variable. In this case, we should reorder to variables so that y is the first variable, z is the second variable and x is the third variable. So we obtain a function

$$f^*(y, z, x) = f(x, y, z) = 3y^2 + 2z^2 + x^2$$

and a constraint

$$g^*(y, z, x) = \begin{bmatrix} y^2 + z^2 + x^2 - 1 \\ y + z + x - 1 \end{bmatrix}$$

Hence

$$\begin{aligned} J &= \begin{bmatrix} 2y & 2z \\ 1 & 1 \end{bmatrix} \\ C &= \begin{bmatrix} 2x \\ 1 \end{bmatrix} \\ J^{-1}C &= \begin{bmatrix} 2y & 2z \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 2x \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} \frac{z-x}{z-y} \\ \frac{x-y}{z-y} \end{bmatrix} \end{aligned}$$

and $\left[\frac{\partial f^*}{\partial w_1}, \ldots, \frac{\partial f^*}{\partial w_m}\right] J^{-1}C = \left[\frac{\partial f^*}{\partial z_1}, \ldots, \frac{\partial f^*}{\partial z_{n-m}}\right]$ leads to

$$\begin{aligned} \left[\frac{\partial f^*}{\partial y}, \frac{\partial f^*}{\partial z}\right] \begin{bmatrix} \frac{z-x}{z-y} \\ \frac{x-y}{z-y} \end{bmatrix} &= \frac{\partial f^*}{\partial x} \\ [4y, 6z] \begin{bmatrix} \frac{z-x}{z-y} \\ \frac{x-y}{z-y} \end{bmatrix} &= 2x \end{aligned}$$

We obtain the equation

$$2xz - xy - yz = 0$$

Hence every critical point satisfies the three equations

$$\begin{aligned} x^2 + y^2 + z^2 &= 1 \\ x + y + z &= 1 \\ 2xz &= y\,(x + z) \end{aligned}$$

These equations are identical with the three equations we obtained in Case 1.

Case 3. $x = y$ and $y = z$. Then $x = y = z$, and $x + y + z = 1$ leads to $x = y = z = \frac{1}{3}$. But this point does not satisfy the constraint $x^2 + y^2 + z^2 = 1$.

Hence we found three candidates for extreme points, namely the three unit vectors. The minimum is obtained at $[x, y, z] = [1, 0, 0]$ and has value 1; the maximum is equal to 3 and obtained at $[x, y, z] = [0, 0, 1]$. The solution is of course identical with the answer obtained in the last section. The additional candidate for extreme values in the last section originate from endpoints of the range for y, i.e. from the requirement that $-\frac{1}{3} \leq y \leq 1$.

11. Lagrange Multipliers

The Method of Lagrange Multipliers is a consequence of the Jacobian Method. Again, we consider the following optimization problem:

$$\begin{aligned} \text{Find the local extreme values of } y &= f(\vec{x}) \\ \text{subject to the conditions } g_1(\vec{x}) &= 0 \\ &\vdots \\ g_m(\vec{x}) &= 0 \end{aligned}$$

where f is a function of n variables, and where $g = [g_1, \dots g_m]^T : D \subseteq \Re^n \to \Re^m$. Let $\vec{x}_0$ be a point where a local extreme value occurs. As before, we assume that the derivative matrix $g'(\vec{x}_0)$ has rank m. After reordering the variables, if necessary, we can write

$$g'(\vec{x}_0) = [J, C]$$

where the Jacobian J is invertible, and where C is the control matrix. Jacobi's Method implies that at $\vec{x}_0$ we have

$$\left[\frac{\partial f}{\partial w_1}, \dots, \frac{\partial f}{\partial w_m}, \frac{\partial f}{\partial z_1}, \dots, \frac{\partial f}{\partial z_{n-m}}\right] \cdot \begin{bmatrix} -J^{-1}C \\ I \end{bmatrix} = \vec{0}$$

where the $w_i's$ are the state variables and the $z_j's$ are the decision variables. We can rewrite this equation as

$$\left[\frac{\partial f}{\partial w_1}, \dots, \frac{\partial f}{\partial w_m}\right] J^{-1}C = \left[\frac{\partial f}{\partial z_1}, \dots, \frac{\partial f}{\partial z_{n-m}}\right]$$

We now introduce a vector $\vec{\lambda} = [\lambda_1, \dots, \lambda_m]$ by

$$[\lambda_1, \dots, \lambda_m] = \left[\frac{\partial f}{\partial w_1}, \dots, \frac{\partial f}{\partial w_m}\right] J^{-1}$$

The $\lambda_k's$ are called Lagrange multipliers. We find that

$$\begin{aligned} [\lambda_1, \dots, \lambda_m] J &= \left[\frac{\partial f}{\partial w_1}, \dots, \frac{\partial f}{\partial w_m}\right] \\ [\lambda_1, \dots, \lambda_m] C &= \left[\frac{\partial f}{\partial z_1}, \dots, \frac{\partial f}{\partial z_{n-m}}\right] \end{aligned}$$

If we reassemble the derivative matrix $g'(\vec{x}_0) = [J, C]$, we obtain

$$[\lambda_1, \dots, \lambda_m][J, C] = \left[\frac{\partial f(\vec{x}_0)}{\partial w_1}, \dots, \frac{\partial f(\vec{x}_0)}{\partial w_m}, \frac{\partial f}{\partial z_1}, \dots, \frac{\partial f}{\partial z_{n-m}}\right]$$

or

$$\begin{aligned} [\lambda_1, \dots, \lambda_m] g'(\vec{x}_0) &= \left[\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_m}, \frac{\partial f}{\partial x_{m+1}}, \dots, \frac{\partial f}{\partial x_n}\right] \\ [\lambda_1, \dots, \lambda_m] \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1} & \frac{\partial g_m}{\partial x_2} & \cdots & \frac{\partial g_m}{\partial x_n} \end{bmatrix} &= \left[\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}\right] \end{aligned}$$

Writing this matrix equation coordinate-wise yields the following theorem:

THEOREM 11. *Let f, $g_1, \ldots, g_m$ be differentiable functions of n variables, and let $\vec{x}_0$ be a solution of the problem*

$$\begin{aligned} \text{Find the local extreme values of } y &= f(\vec{x}) \\ \text{subject to the conditions } g_1(\vec{x}) &= 0 \\ &\vdots \\ g_m(\vec{x}) &= 0 \end{aligned}$$

Then either the rank of the matrix $g'(\vec{x}_0) = \left[\frac{\partial g_i(\vec{x}_0)}{\partial x_j}\right]_{1\leq i\leq m, 1\leq j\leq n}$ is less than m or else there are number $\lambda_1, \ldots, \lambda_m$ so that

$$\begin{aligned} \frac{\partial f(\vec{x}_0)}{\partial x_1} &= \lambda_1 \frac{\partial g_1(\vec{x}_0)}{\partial x_1} + \cdots + \lambda_m \frac{\partial g_m(\vec{x}_0)}{\partial x_1} \\ &\vdots \\ \frac{\partial f(\vec{x}_0)}{\partial x_n} &= \lambda_1 \frac{\partial g_1(\vec{x}_0)}{\partial x_n} + \cdots + \lambda_m \frac{\partial g_m(\vec{x}_0)}{\partial x_n} \end{aligned}$$

EXAMPLE 28. *Let $f(x,y,z) = x^2 + 2y^2 + 3z^2$. Find the extreme values of this function under the conditions that*

$$\begin{aligned} x^2 + y^2 + z^2 &= 1 \\ x + y + z &= 1 \end{aligned}$$

The method of Lagrange multipliers leads to the three additional equations

$$\begin{aligned} 2x &= 2x\lambda_1 + \lambda_2 \\ 4y &= 2y\lambda_1 + \lambda_2 \\ 6z &= 2z\lambda_1 + \lambda_2 \end{aligned}$$

We can write this equation as

$$\begin{bmatrix} 2x & 1 \\ 2y & 1 \\ 2z & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 2x \\ 4y \\ 6z \end{bmatrix}$$

This equation has only a solution for λ_1 and λ_2 if

$$\begin{aligned} \det \begin{bmatrix} 2x & 1 & -2x \\ 2y & 1 & -4y \\ 2z & 1 & -6z \end{bmatrix} &= 0 \\ 4xy - 8xz + 4yz &= 0 \\ xy - 2xz + yz &= 0 \\ (x+z)\,y &= 2xz \end{aligned}$$

Since $x + y + z = 1$, we find that

$$\begin{aligned} (1-y)\,y &= 2x(1-x-y) \\ y - y^2 &= 2x(1-x-y) \\ y - y^2 &= 2x - 2x^2 - 2xy \end{aligned}$$

Since $x^2+y^2+z^2=1$, we also find that

$$\begin{aligned} x^2+y^2+(1-x-y)^2 &= 1 \\ 2xy-2y-2x+2x^2+2y^2 &= 0 \\ -2y+2y^2 &= 2x-2x^2-2xy \end{aligned}$$

Hence

$$\begin{aligned} y-y^2 &= -2y+2y^2 \\ y^2-y &= 0 \end{aligned}$$

We conclude that either $y=0$ or $y=1$.

If $y=0$, then $y-y^2=2x-2x^2-2xy$ reduces to $x=x^2$, hence either $x=0$ and $z=1$ or else $x=1$ and $z=0$.

If $y=1$ then $x^2+y^2+z^2=1$ implies that $x=y=0$.

We now have three possible candidates for an extreme value, namely, $[x,y,z]=[1,0,0]$ (minimum), $[x,y,z]=[0,1,0]$ and $[x,y,z]=[0,0,1]$ (maximum).

As an application, we prove Hadamard's Theorem:

THEOREM 12. *Let $A=[a_{i,j}]_{1\le i.,j\le n}$ be a square matrix, and assume that*

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2 = m$$

Then

$$\det A \le \left(\frac{m}{n}\right)^{n/2}$$

PROOF. We know that

$$\det A = \sum_{\sigma\in S_n} sign\,(\sigma)\prod_{i=1}^{n}\alpha_{i\sigma(i)}$$

is a function of the n^2 variables a_{ij}. So we have to solve the following problem:

$$\begin{aligned} \text{Maximize } \det A &= \sum_{\sigma\in S_n} sign\,(\sigma)\prod_{i=1}^{n}\alpha_{i\sigma(i)} \\ \text{subject to } \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2 &= m \end{aligned}$$

In the following, A is a matrix where the maximum is obtained. The method of Lagrange Multipliers leads to the equations

$$\frac{\partial \det A}{\partial a_{ij}} = 2\lambda a_{ij}$$

We develop the the i^{th} row of $\det A$:

$$\det A = \sum_{k=1}^{n}(-1)^{i+k}\,a_{ik}\det A_{ik}$$

None of the determinants $\det A_{ik}$ depends on a_{ij}, hence

$$\frac{\partial \det A}{\partial a_{ij}} = (-1)^{i+j}\det A_{ij}$$

It follows that

$$(-1)^{i+j} \det A_{ij} = 2\lambda a_{ij}$$

Multiplying these equation by a_{ij} leads to

$$(-1)^{i+j} a_{ij} \det A_{ij} = 2\lambda a_{ij}^2$$

Summing over j gives

$$\begin{aligned} \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij} &= 2\lambda \sum_{j=1}^{n} a_{ij}^2 \\ \det A &= 2\lambda \sum_{j=1}^{n} a_{ij}^2 \end{aligned}$$

We conclude that for all values of i we have

$$\sum_{j=1}^{n} a_{ij}^2 = \frac{\det A}{2\lambda}$$

Summing over all values of i gives

$$m = n \frac{\det A}{2\lambda}$$

and therefore

$$\det A = 2\lambda \frac{m}{n}$$

If $\det A = 0$, then the statement is obvious. So we may assume that A is invertible. By Cramer's rule, we find that

$$A^{-1} = \frac{1}{\det A} \left[(-1)^{i+j} \det A_{ji} \right]$$

Hence

$$\begin{aligned} A^{-1} &= \frac{2\lambda}{\det A} [a_{ji}] \\ &= \frac{2\lambda}{\det A} A^T \end{aligned}$$

It follows that

$$\begin{aligned} \det A^{-1} &= \left(\frac{2\lambda}{\det A} \right)^n \det \left(A^T \right) \\ \frac{1}{\det A} &= \left(\frac{2\lambda}{\det A} \right)^n \det (A) \\ (\det A)^{n-2} &= (2\lambda)^n \end{aligned}$$

Substituting $\det A = 2\lambda \frac{m}{n}$, i.e. $2\lambda = \frac{n}{m} \det A$ gives

$$\begin{aligned} (\det A)^{n-2} &= \left(\frac{n}{m} \det A \right)^n \\ \left(\frac{m}{n} \right)^n &= (\det A)^2 \\ \det A &= \left(\frac{m}{n} \right)^{n/2} \end{aligned}$$

□

12. Regional Constraints and the Kuhn-Tucker Conditions

We start with a general theorem. Actually, this theorem was already used in the proof of Hadamard's theorem. In this proof, we assume that A is a matrix maximizing the determinant $\det A$ with respect to the constraints $\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2 = m$. The existence of such a matrix is guaranteed by the following theorem:

THEOREM 13. *Suppose that we are given a continuous function $f : D \to \Re$, where $D \subseteq \Re^n$ is a closed, bounded region. Then f has a maximum and a minimum on D.*

In order to find the maximum and minimum, we separate D into the interior and the boundary:

$$D = D^o \cup \partial D$$

The interior is given by

$$\{\vec{x} : \text{there is } \varepsilon > 0 \text{ so that } \|\vec{x} - \vec{y}\| < \varepsilon \text{ implies } \vec{y} \in D\}$$

and the boundary is the rest:

$$\partial D = D \setminus D^o$$

EXAMPLE 29. *Find the interior and the boundary for the following regions:*

(1) $D = \left\{(x,y) : x^2 + y^2 \leq 1\right\}$ - *has interior* $D^o = \left\{(x,y) : x^2 + y^2 < 1\right\}$ *and the boundary is given by* $\left\{(x,y) : x^2 + y^2 = 1\right\}$
(2) $D = \{(x,y) : |x| + |y| \leq 1\}$ - *similar.*

The examples suggest that the interior is given by strict inequalities, whereas the boundary is defined by using equations. This is normally the case, but needs to be verified in each problem.

If we would like to find extreme values on regions, we use the following approach:

(1) Find all critical points of f, i.e. find all solutions of $\nabla f(\vec{x}) = 0$. Check which of those solutions belong to D^o. These solutions are called feasible solutions.
(2) Discuss the behavior of f on ∂D. Use Lagrange multipliers!

EXAMPLE 30. *Find the maximum and the minimum of $f(x,y) = \sin\left(2x^2 + y^2\right)$, if $x^2 + 2y^2 \leq 1$.*

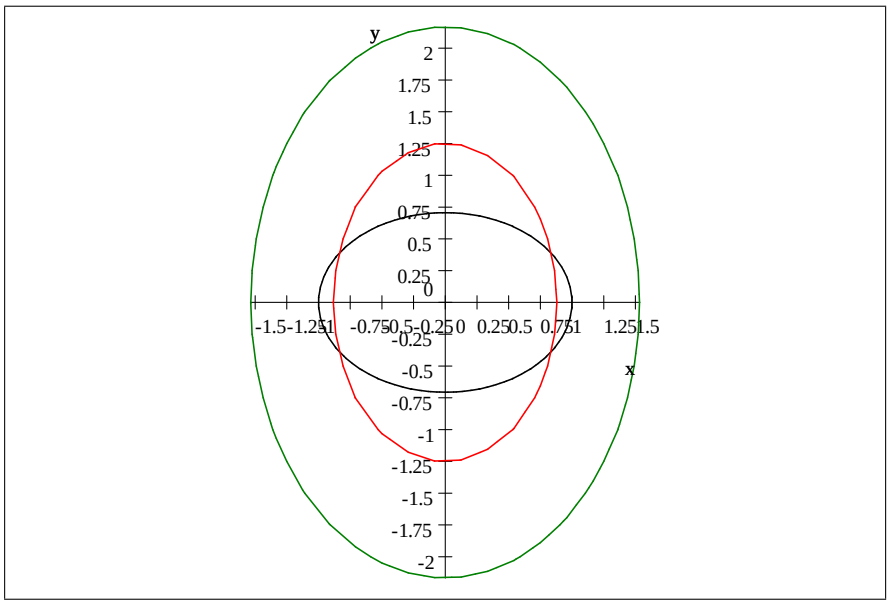
(1) The interior is given by $x^2 + 2y^2 < 1$. The gradient needs to be zero:

$$\begin{aligned} 4x\cos\left(2x^2 + y^2\right) &= 0 \\ 2y\cos\left(2x^2 + y^2\right) &= 0 \end{aligned}$$

So either $\cos\left(2x^2 + y^2\right) = 0$ or $x = y = 0$. Hence

$$\begin{aligned} x &= y = 0 \text{ or} \\ 2x^2 + y^2 &= \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \ldots \end{aligned}$$

We plot the constraint $x^2 + 2y^2 = 1$ (black) and the solutions of $2x^2 + y^2 = \frac{\pi}{2}$ (red) and $2x^2 + y^2 = \frac{\pi}{2}$ (green):



So only the solutions of $2x^2 + y^2 = \frac{\pi}{2}$ are feasible points, and for all of those, $\sin\left(2x^2 + y^2\right) = 1$. Therefore the maximum value in the interior is equal to 1, occurring at all points (x, y) with $2x^2 + y^2 = \frac{\pi}{2}$ that also satisfy $x^2 + 2y^2 < 1$. Since $\sin\left(2 \times 0^2 + 0^2\right) = 0$, the minimum value is 0, occurring at $(0, 0)$.

(2) Next, we consider the boundary, and use Lagrange multipliers: We find that

$$\begin{aligned} x\left(2\cos\left(2x^2 + y^2\right) + \lambda\right) &= 0 \\ y\left(\cos\left(2x^2 + y^2\right) + 2\lambda\right) &= 0 \\ x^2 + 2y^2 &= 1 \end{aligned}$$

One of x and y has to be different from 0.

(a) If $x = 0$, then $y = \pm\frac{1}{2}\sqrt{2}$
(b) If $y = 0$, then $x = \pm 1$
(c) If $x \neq 0$ and $y \neq 0$, then we can divide the first two equations by x and y, respectively. We then write the equations in matrix form:

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} \cos\left(2x^2 + y^2\right) \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Since the matrix is invertible, we find that $\cos\left(2x^2 + y^2\right) = 0$. hence $2x^2 + y^2 = \frac{\pi}{2}$.

(3) Conclusion:

(a) The minimum value of $f(x, y) = \sin\left(2x^2 + y^2\right)$ is equal to 0, and it occurs at $(x, y) = (0, 0)$.
(b) The maximum value of $f(x, y) = \sin\left(2x^2 + y^2\right)$ is equal to 1, and it occurs at all points satisfying $2x^2 + y^2 = \frac{\pi}{2}$ that also satisfy $x^2 + 2y^2 \leq 1$.

.

THEOREM 14 (Kuhn-Tucker). *Assume that $f(\vec{x})$ and $g(\vec{x})$ are differentiable functions of $\vec{x}$. If f has a local extremum at $\vec{x}_0$ in the feasible region $R = \{\vec{x} \in \Re^n : g(\vec{x}) \leq 0\}$, then, under the additional condition that $\nabla g(\vec{x}_0) \neq \vec{0}$, it is necessary that*

$$\begin{aligned} \frac{\partial f}{\partial x_i}(\vec{x}_0) + \lambda \frac{\partial g}{\partial x_i}(\vec{x}_0) &= 0 \\ g(\vec{x}_0) &\leq 0 \\ \lambda g(\vec{x}_0) &= 0 \text{ (i.e. } \lambda = 0 \text{ or } g(\vec{x}_0) = 0) \\ \lambda &\leq 0 \text{ (for a maximum) or} \\ \lambda &\geq 0 \text{ (for a minimum)} \end{aligned}$$

PROOF. Let $\vec{x}_0$ be a feasible local extreme value. We show that this implies there is a λ so that

$$\frac{\partial f}{\partial x_i}(\vec{x}_0) + \lambda \frac{\partial g}{\partial x_i}(\vec{x}_0) = 0$$

and

$$\begin{aligned} \lambda g(\vec{x}_0) &= 0 \\ \lambda &\geq 0 \text{ for a minimum} \\ \lambda &\leq 0 \text{ for a maximum} \end{aligned}$$

There are two cases two consider:

(1) $g(\vec{x}_0) < 0$. In this case, $\vec{x}_0$ belongs to interior of the feasible region and therefore is a critical point, i.e. all partial derivatives are vanishing at $\vec{x}_0$. We can pick $\lambda = 0$. Hence

$$\lambda g(\vec{x}_0) = 0$$

(2) $g(\vec{x}_0) = 0$. It follows that

$$\lambda g(\vec{x}_0) = 0$$

for each value of λ. Furthermore, the method of Lagrange implies the existence of a λ so that
$\frac{\partial f}{\partial x_i}(\vec{x}_0) + \lambda \frac{\partial g}{\partial x_i}(\vec{x}_0) = 0$ for all i.

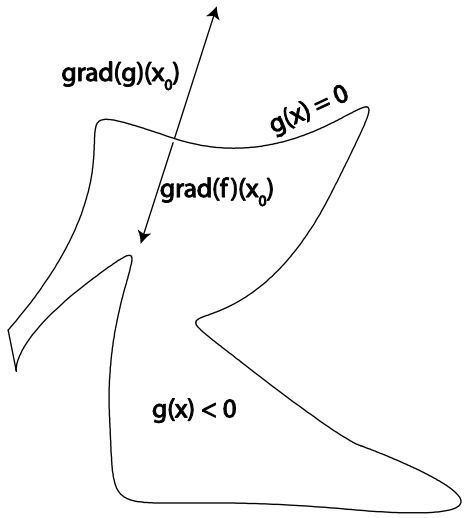
Let us assume that f has a local maximum at $\vec{x}_0$ We have to show that we can find a solution for λ with $\lambda \leq 0$. If $\nabla f(\vec{x}_0) = \vec{0}$, then we also can pick $\lambda = 0$. So we may assume that $\nabla f(\vec{x}_0) \neq \vec{0}$. Using the method of Lagrange multipliers, we find that the local extrema of $f(\vec{x})$ under the condition $g(\vec{x}) = 0$ are among the solutions of the equations

$$\begin{aligned} \frac{\partial f}{\partial x_i}(\vec{x}_0) + \lambda \frac{\partial g}{\partial x_i}(\vec{x}_0) &= 0 \text{ or} \\ \lambda \frac{\partial g}{\partial x_i}(\vec{x}_0) &= -\frac{\partial f}{\partial x_i}(\vec{x}_0) \end{aligned}$$

We find that

$$\nabla f(\vec{x}_0) = -\lambda \nabla g(\vec{x}_0)$$

i.e. $\nabla f(\vec{x}_0)$ and $\nabla g(\vec{x}_0)$ are parallel. The vector $\nabla g(\vec{x}_0)$ points in the direction of largest ascend of $g(\vec{x}_0)$. Since $\nabla g(\vec{x}_0) \neq \vec{0}$, we find that $g(\vec{x}_0 + \varepsilon \nabla g(\vec{x}_0)) > 0$. So $\vec{x}_0 + \varepsilon \nabla g(\vec{x}_0)$ does not belong to the feasible region for each $\varepsilon > 0$, and hence $\nabla g(\vec{x}_0)$ points to the outside of this region.



If $\lambda > 0$, then $\nabla f(\vec{x}_0) = -\lambda \nabla g(\vec{x}_0)$ points to the inside of the feasible region, i.e. $\vec{x}_0 + \varepsilon \nabla f(\vec{x}_0)$ belongs to the feasible region for small values of $\varepsilon > 0$. Since $\nabla f(\vec{x}_0)$ points in the direction of largest ascent of $f(\vec{x}_0)$, it follows that $f(\vec{x}_0 + \varepsilon \nabla f(\vec{x}_0)) > f(\vec{x}_0)$ for sufficiently small values of $\varepsilon > 0$, and hence $f(\vec{x}_0)$ cannot be a local maximum. Therefore, $\lambda \leq 0$ for a local maximum.

Similarly $\lambda < 0$ implies that $f(\vec{x}_0)$ cannot be a local minimum □

EXAMPLE 31. *Maximize*

$$f(x,y) = \left(1+x^2\right) e^{-y^2}$$

subject to

$$x^2 - y^2 \leq 1$$

If we really would like to use the previous theorem, then we have to consider the equations

$$\begin{aligned} F(x,y,\lambda) &= \left(1+x^2\right) e^{-y^2} + \lambda\left(x^2 - y^2 - 1\right) \\ \left(x^2 - y^2 - 1\right) &\leq 0 \\ 2xe^{-y^2} + 2x\lambda &= 0 \\ -2y\left(1+x^2\right) e^{-y^2} - 2y\lambda &= 0 \\ \lambda\left(x^2 - y^2 - 1\right) &= 0 \end{aligned}$$

Hence we have two cases:

(1) If $\lambda = 0$, then

$$\begin{aligned} 2xe^{-y^2} &= 0 \\ -2y\left(1+x^2\right)e^{-y^2} &= 0 \end{aligned}$$

and hence $x = y = 0$. Since $0^2 - 0^2 \leq 1$, this is a feasible point.

(2) If $\lambda \neq 0$, then $x^2 - y^2 - 1 = 0$, hence $y^2 = x^2 - 1$. We arrive at

$$\begin{aligned} x\left(e^{1-x^2} + \lambda\right) &= 0 \\ \sqrt{x^2-1}\left[\left(1+x^2\right)e^{1-x^2} - \lambda\right] &= 0 \end{aligned}$$

Since $x = 0$ violates the constraint $x^2 - y^2 - 1 = 0$, we find that $x \neq 0$. Hence $\lambda = -e^{1-x^2} < 0$. So in this case every local extreme value will be a local maximum. The second equation leads to

$$\begin{aligned} \sqrt{x^2-1}\left[\left(1+x^2\right)e^{1-x^2} + e^{1-x^2}\right] &= 0 \\ \sqrt{x^2-1}\left[\left(2+x^2\right)\right]e^{1-x^2} &= 0 \end{aligned}$$

Hence

$$\sqrt{x^2-1} = 0$$

or

$$x = \pm 1$$

If $x = \pm 1$, then $y = 0$.

So the points $(0,0)$ and $\pm(1,0)$ are the only candidates for local extreme values. The theorem of Kuhn and Tucker implies:

(1) Only $(0,0)$ could be a either a local maximum or a local minimum. At $(0,0)$, the Hessian of the function $f(x,y) = \left(1+x^2\right)e^{-y^2}$, is equal to

$$H(f)(0,0) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$$

so there is a saddle point at $[0,0]$

(2) The points $\pm[1,0]$ could only be local maxima. At this points, the Hessian matrix cannot be used, since we find ourselves at the boundary of the feasible region. If we let

$$\begin{aligned} x &= \pm 1 + \varepsilon \\ y &= \varepsilon \end{aligned}$$

then

$$\begin{aligned} f(x,y) &= f(\pm 1 + \varepsilon, \varepsilon) \\ &= \left(1 + (\pm 1 + \varepsilon)^2\right)e^{-\varepsilon^2} \\ &= \left(2 \pm 2\varepsilon + \varepsilon^2\right)e^{-\varepsilon^2} \end{aligned}$$

If one of the points $\pm[1,0]$ would be a local maximum, then we also would have to have a local maximum of $\left(2 \pm 2\varepsilon + \varepsilon^2\right)e^{-\varepsilon^2}$ at $\varepsilon = 0$. The derivatives of the functions

$$h_{1/2}(\varepsilon) = \left(2 \pm 2\varepsilon + \varepsilon^2\right)e^{-\varepsilon^2}$$

at $\varepsilon = 0$ are equal to

$$\frac{d}{d\varepsilon}\left(\left(2 \pm 2\varepsilon + \varepsilon^2\right) e^{-\varepsilon^2}\right)_{\varepsilon=0} = \pm 2$$

Since both values are different from 0, we cannot have a local extreme value at $\varepsilon = 0$.

We state the Kuhn-Tucker Theorem for two constraints:

THEOREM 15. *Assume that $f(\vec{x})$ and $g_1(\vec{x})$ and $g_2(\vec{x})$ are differentiable functions of $\vec{x}$. If f has a local maximum (minimum) at $\vec{x}_0$ in the feasible region R of the problem: Find the local maxima (minima) of*

$$y = f(\vec{x})$$

subject to the constraints

$$\begin{aligned} g_1(\vec{x}) &\leq 0 \\ g_2(\vec{x}) &\leq 0 \end{aligned}$$

then, under the additional condition that $\nabla g_1(\vec{x}_0)$ and $\nabla g_2(\vec{x}_0)$ are linearly independent, it is necessary that

$$\begin{aligned} \frac{\partial f}{\partial x_i}(\vec{x}_0) + \lambda \frac{\partial g_1}{\partial x_i}(\vec{x}_0) + \mu \frac{\partial g_2}{\partial x_i}(\vec{x}_0) &= 0 \\ g_1(\vec{x}_0) &\leq 0 \text{ and } g_2(\vec{x}_0) \leq 0 \\ \lambda g_1(\vec{x}_0) &= 0 \text{ and } \mu g_2(\vec{x}_0) = 0 \\ \lambda &\leq 0 \text{ and } \mu \leq 0 \text{ (for a maximum) or} \\ \lambda &\geq 0 \text{ and } \mu \geq 0 \text{ (for a minimum)} \end{aligned}$$

EXAMPLE 32. *Find the maximum and minimum of*

$$f(x, y) = 3x^2 - 2y^2$$

if

$$\begin{aligned} 4x^2 + y^2 &\leq 1 \\ x^2 + 4y^2 &\leq 1 \end{aligned}$$

EXAMPLE 33. *Find the local maxima and minima of*

$$f(x, y) = 3x^2 + 2y^3$$

if

$$\begin{aligned} x + y &\leq 1 \\ x - y &\geq 1 \end{aligned}$$

13. Convex Sets

Convex functions are occuring in many applications. We already know from the second derivative test that a function f that is convex if and only if the second derivative f'' is positive. We will have to generalize this result to functions of several variables. But first, we will have to give a definition of convex functions in more algebraic terms.

The natural domain of convex functions will be convex subsets of $\Re^n$. We start with a quick review of some topics from linear algebra:

DEFINITION 8. *A subspace $V \subseteq \Re^n$ is a linear subspace, if*

(1) $\vec{0} \in V$
(2) *If $\vec{u}, \vec{v} \in V$ and if r and s are numbers, then $r\vec{u} + s\vec{v} \in V$.*

Every subspace V has a dimension. If $\vec{u}_1, \ldots, \vec{u}_k \in V$ are linearly independent vectors so that every $\vec{v} \in V$ is linear combination of $\vec{u}_1, \ldots, \vec{u}_k$, then $\vec{u}_1, \ldots, \vec{u}_k$ is called a basis of V, and the (uniquely determined) number k is called the dimension of V. In this case, we write $k = \dim V$.

If V is a subspace of $\Re^n$ and if $\vec{x}_0$ is a fixed vector, then

$$\vec{x}_0 + V = \{\vec{x}_0 + x : x \in V\}$$

is called an affine subspace. The dimension of $\vec{x}_0 + V$ is defined to be the dimension of V.

Affine subspaces can be described as solution sets of equations:

THEOREM 16. *If $A \subseteq \Re^n$ is an affine subspace, then there is an $m \times n$ matrix M and a vector $b \in \Re^m$ so that*

$$A = \left\{\vec{x} \in \Re^n : M\vec{x} = \vec{b}\right\}$$

Conversely, if $A = \left\{\vec{x} \in \Re^n : M\vec{x} = \vec{b}\right\}$ for a certain $m \times n$ matrix M, then A is an affine subspace.

Subspace of $\Re^n$ of dimension $n-1$ are called hyperplanes:

DEFINITION 9. *If $\vec{x}_1, \ldots, \vec{x}_{n-1}$ are $n-1$ linearly independent vectors in $\Re^n$, then*

$$\begin{aligned} H &= span\,(\vec{x}_1, \ldots, \vec{x}_{n-1}) \\ &= \{r_1\vec{x}_1 + \cdots + r_{n-1}\vec{x}_{n-1} \in \Re^n : r_1, \ldots, r_{n-1} \in \Re\} \end{aligned}$$

is a hyperplane in $\Re^n$. If H is a hyperplane, and if $\vec{x}_0 \in \Re^n$ is a fixed vector, then

$$\vec{x}_0 + H = \{\vec{x}_0 + x : x \in H\}$$

is called an affine hyperplane.

Hyperplanes are vector spaces in their own right, i.e. if H is a hyperplane, then $\vec{0} \in H$ and if $\vec{x}, \vec{y} \in H$ and if r and s are scalars, then $r\vec{x} + r\vec{y} \in H$. Moreover, the dimension of H is $n-1$.

Affine hyperplanes are not necessarily vector spaces. Affine hyperplanes may also describes as solution sets of an equation:

THEOREM 17. *A subset $G \subseteq \Re^n$ is an affine hyperplane if and only if there are numbers $a_1, \ldots, a_n, b \in \Re$ so that not all of the $a_i's$ are equal to 0 and so that*

$$G = \{\vec{x} \in \Re^n : a_1x_1 + \cdots + a_nx_n = b\}$$

The affine hyperplane $G = \{\vec{x} \in \Re^n : a_1x_1 + \cdots + a_nx_n = b\}$ is a hyperplane if and only if $b = 0$.

If we define $\vec{a} = [a_1, \ldots, a_n]$, then we can write G in the form

$$G = \{\vec{x} \in \Re^n : \vec{a} \cdot \vec{x} = b\}$$

DEFINITION 10. *Two affine hyperplanes G_1 and G_2 satisfying either $G_1 = G_2$ or $G_1 \cap G_2 = \emptyset$ are called parallel.*

THEOREM 18. *Let G_1 and G_2 be two hyperplanes of $\Re^n$. Then the following statements are equivalent:*

(1) *G_1 and G_2 are parallel.*
(2) *There is a hyperplane H and vectors $\vec{x}_1, \vec{x}_2 \in \Re^n$ so that*

$$\begin{aligned} G_1 &= \vec{x}_1 + H \\ G_2 &= \vec{x}_2 + H \end{aligned}$$

(3) *There is a vector $\vec{b} \in \Re^n$ so that*

$$\begin{aligned} G_1 &= \vec{b} + G_2 \\ &= \left\{\vec{b} + \vec{x} : \vec{x} \in G_2\right\}. \end{aligned}$$

THEOREM 19. *If G_1 and G_2 are two affine hyperplanes, then either G_1 and G_2 are parallel, or else there is an affine subspace A of dimension $n-2$ so that $G_1 \cap G_2 = A$.*

Let G be a hyperplane of $\Re^n$. Then it separates $\Re^n$ into two half spaces. If

$$G = \{\vec{x} \in \Re^n : a_1x_1 + \cdots + a_nx_n = d\},$$

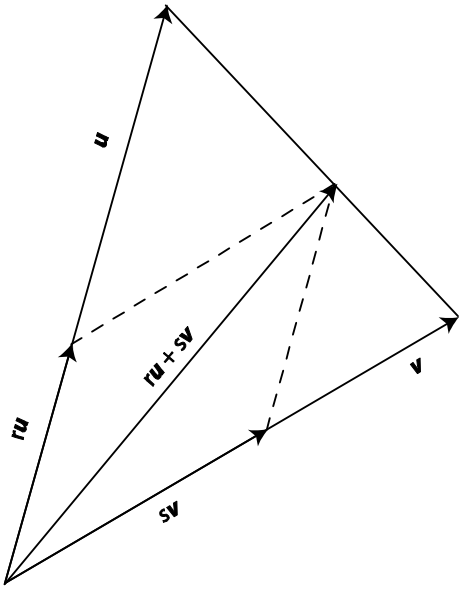
then the two half spaces are given by

$$\begin{aligned} H_1 &= \{\vec{x} \in \Re^n : a_1x_1 + \cdots + a_nx_n \geq d\} \\ H_2 &= \{\vec{x} \in \Re^n : a_1x_1 + \cdots + a_nx_n \leq d\} \end{aligned}$$

DEFINITION 11. *Let $\vec{u}, \vec{v} \in \Re^n$ be two vectors. A convex combination of $\vec{u}$ and $\vec{v}$ is a vector of the form $r\vec{u} + (1-r)\vec{v}$, where $0 \leq r \leq 1$. The line segment between $\vec{u}$ and $\vec{v}$ is denoted by $conv\{\vec{u}, \vec{v}\}$ and consists of all convex combinations of $\vec{u}$ and $\vec{v}$:*

$$conv\{\vec{u}, \vec{v}\} = \{r\vec{u} + (1-r)\vec{v} : 0 \leq r \leq 1\}$$

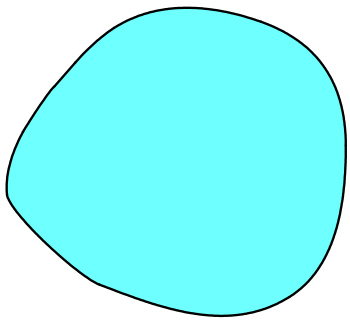
As the following illustration shows, the convex combinations of $\vec{u}$ and $\vec{v}$ fill up the line segment from $\vec{u}$ to $\vec{v}$.



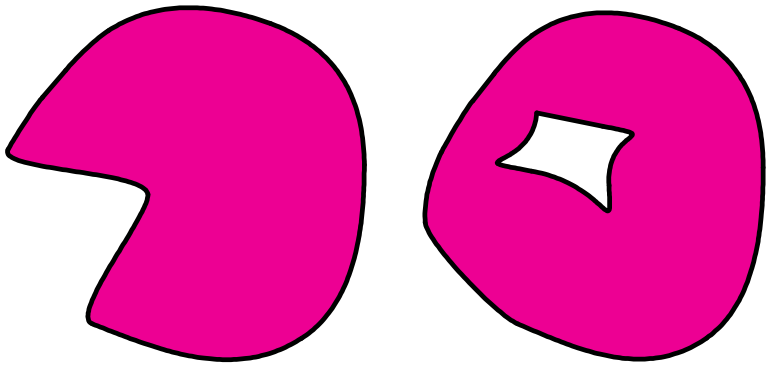
DEFINITION 12. *A subset $C \subseteq \Re^n$ is convex if $\vec{x}, \vec{y} \in C$ and $0 \leq r \leq 1$ implies that $r\vec{x} + (1-r)\,\vec{y} \in C$.*

In other words, a subset C is convex if and only if for every choice of elements $\vec{x}, \vec{y} \in C$ the line segment $conv\,\{\vec{x}, \vec{y}\}$ is contained in C, i.e. $conv\,\{\vec{x}, \vec{y}\} \subseteq C$.

A convex set:



Two sets that are not convex:



THEOREM 20. *Every line segment is a convex set.*

THEOREM 21. *Every half-space in $\Re^n$ is a convex set.*

PROOF. A half space in $\Re^n$ is of the form $A = \left\{\vec{x} \in \Re^n : \vec{x} \cdot \vec{d} \leq c\right\}$, where $\vec{d} \in \Re^n$ is a fixed vector and where $c \in \Re$ is a fixed number. If $\vec{x}, \vec{y} \in A$ and $0 \leq r \leq 1$ are given, then we have to show that . We know that $\vec{x} \cdot \vec{d} \leq c$ and $\vec{y} \cdot \vec{d} \leq c$. We now can compute

$$\begin{aligned} (r\vec{x} + (1-r)\vec{y}) \cdot \vec{d} &= r\left(\vec{x} \cdot \vec{d}\right) + (1-r)\left(\vec{y} \cdot \vec{d}\right) \\ &\leq rc + (1-r)c \\ &= c \end{aligned}$$

Hence $r\vec{x} + (1-r)\vec{y} \in A$. □

THEOREM 22. *The "closed unit ball" $U = \{\vec{x} \in \Re^n : \|\vec{x}\| \leq 1\}$ and the "open unit ball" $U^o = \{\vec{x} \in \Re^n : \|\vec{x}\| < 1\}$ are convex sets.*

PROOF. If $\|\vec{x}\|, \|\vec{y}\| \leq 1$, and if $0 \leq r \leq 1$, then the triangle inequality implies that

$$\begin{aligned} \|r\vec{x} + (1-r)\vec{y}\| &\leq \|r\vec{x}\| + \|(1-r)\vec{y}\| \\ &= |r|\,\|\vec{x}\| + |1-r|\,\|\vec{y}\| \\ &= r\,\|\vec{x}\| + (1-r)\,\|\vec{y}\| \\ &\leq r + (1-r) = 1 \end{aligned}$$

Hence the closed unit ball is a convex set. The proof for the open unit ball is similar. □

THEOREM 23. *If C_1,$C_2 \subseteq \Re^n$ are convex sets, then $C_1 \cap C_2$ is a convex set. Similarly, if $(C_i)_{i\in I}$ is a family of convex subsets of $\Re^n$, then $\bigcap_{i\in I} C_i$ is a convex set.*

PROOF. If $\vec{x}, \vec{y} \in \bigcap_{i\in I} C_i$ and if $0 \leq r \leq 1$, then we have to show that $r\vec{x} + (1-r)\vec{y} \in \bigcap_{i\in I} C_i$. Since $\vec{x}, \vec{y} \in C_i$ for all i and since each set C_i is convex, we find that $r\vec{x} + (1-r)\vec{y} \in C_i$ for each $i \in I$. Hence $r\vec{x} + (1-r)\vec{y} \in \bigcap_{i\in I} C_i$. □

DEFINITION 13. *If $A \subseteq \Re^n$ is a set, then the intersection of all convex subsets $C \subseteq \Re^n$ is called the convex hull of A, denoted by $conv\,(A)$.*

The following propositions allow us to find convex hulls of sets A :

PROPOSITION 1. *Assume that $A \subseteq \Re^n$ is a convex set, that $\vec{x}_1, \ldots, \vec{x}_m \in A$ and that $r_1, \ldots, r_m \geq 0$ are non-negative real numbers so that $\sum_{i=1}^m r_i = 1$. Then $\sum_{i=1}^m r_i \vec{x}_i \in A$.*

PROOF. For $m = 1$, the assumptions reduce to $\vec{x}_1 \in A$ and $r_1 = 1$. In this case, clearly $\sum_{i=1}^m r_i \vec{x}_i = \sum_{i=1}^1 r_i \vec{x}_i = r_1 \vec{x}_1 = \vec{x}_1 \in A$.

For $m = 2$, we are assuming that $\vec{x}_1, \vec{x}_2 \in A$ and $r_1, r_2 \geq 0$, $r_1 + r_2 = 1$. It follows that $r_2 = 1 - r_1$. Hence

$$\begin{aligned} \sum_{i=1}^{2} r_1 \vec{x}_1 &= r_1 \vec{x}_1 + r_2 \vec{x}_2 \\ &= r_1 \vec{x}_1 + (1 - r_1)\, \vec{x}_2 \end{aligned}$$

The definition of convexity yields that $r_1 \vec{x}_1 + (1 - r_1)\, \vec{x}_2 \in A$.

We now proceed with a proof by induction: Assume that we have verified the statement of the proposition for a fixed value of m, and we are trying to prove the same statement for $m + 1$. So we are given vectors $\vec{x}_1, \ldots, \vec{x}_{m+1} \in A$ and non-negative number $r_1, \ldots, r_{m+1} \geq 0$ so that $r_1 + \cdots + r_{m+1} = 1$. We reorder the indices so that r_{m+1} is the minimum of the numbers $r_1, \ldots, r_{m+1}$. If $r_{m+1} = 0$, then $r_1 + \cdots + r_m = 1$ and hence by the induction hypothesis we have

$$\sum_{i=1}^{m+1} r_i \vec{x}_i = \sum_{i=1}^{m} r_i \vec{x}_i \in A$$

We now assume that $r_{m+1} > 0$. It follows that $0 < r_{m+1} \leq r_i$ for all i and hence

$$0 < r_1 + \cdots + r_m = 1 - r_{m+1}$$

This implies

$$\frac{r_1}{1 - r_{m+1}} + \cdots + \frac{r_m}{1 - r_{m+1}} = 1$$

The induction hypothesis implies that

$$\frac{1}{1 - r_{m+1}} \sum_{i=1}^{m} r_i \vec{x}_i = \sum_{i=1}^{m} \frac{r_i}{1 - r_{m+1}} \vec{x}_i \in A$$

Since A is convex, we conclude that

$$\sum_{i=1}^{m+1} r_i \vec{x}_i = (1 - r_{m+1}) \sum_{i=1}^{m} \frac{r_i}{1 - r_{m+1}} \vec{x}_i + r_{m+1} \vec{x}_{m+1} \in A$$

□

PROPOSITION 2. *If $A \subseteq \Re^n$ is any subset, then*

$$conv\,(A) = \left\{ \sum_{i=1}^{m} r_i \vec{x}_i : 0 \leq m \text{ is an integer, } 0 \leq r_1, \ldots, r_m, r_1 + \cdots + r_m = 1 \text{ and } \vec{x}_1, \ldots, \vec{x}_m \in A \right\}$$

PROOF. First, we show that

$$B = \left\{ \sum_{i=1}^{m} r_i \vec{x}_i : 0 \leq m \text{ is an integer, } 0 \leq r_1, \ldots, r_m, r_1 + \cdots + r_m = 1 \text{ and } \vec{x}_1, \ldots, \vec{x}_m \in A \right\}$$

is a convex set. Let $\vec{x}, \vec{y} \in B$ and assume that $0 \leq r \leq 1$. We have to show that $r\vec{x} + (1-r)\,\vec{y} \in B$. By definition, there are integers m_1 and m_2, numbers $0 \leq r_1, \ldots, r_{m_1}$, $0 \leq s_1, \ldots, s_{m_2}$ with $r_1 + \cdots + r_{m_1} = 1 = s_1 + \cdots + s_{m_2}$ and vectors $\vec{x}_1, \ldots, \vec{x}_{m_1}, \vec{y}_1, \ldots, \vec{y}_{m_2} \in A$ so that

$$\begin{aligned} \vec{x} &= \sum_{i=1}^{m_1} r_i \vec{x}_i \\ \vec{y} &= \sum_{i=1}^{m_2} s_i \vec{y}_i \end{aligned}$$

We now let

$$\begin{aligned} t_1 &= r \cdot r_1 \text{ and } \vec{z}_1 = \vec{x}_1 \\ t_2 &= r \cdot r_2 \text{ and } \vec{z}_2 = \vec{x}_2 \\ &\vdots \\ t_{m_1} &= r \cdot r_{m_1} \text{and } \vec{z}_{m_1} = \vec{x}_{m_1} \\ t_{m_1+1} &= (1-r) \cdot s_1 \text{ and } \vec{z}_{m_1+1} = \vec{y}_1 \\ t_{m_1+2} &= (1-r) \cdot s_2 \text{ and } \vec{z}_{m_1+2} = \vec{y}_2 \\ &\vdots \\ t_{m_1+m_2} &= (1-r) \cdot s_{m_2} \text{ and } \vec{z}_{m_1+m_2} = \vec{y}_{m_2} \end{aligned}$$

and compute:

$$\begin{aligned} t_1 + \cdots + t_{m_1+m_2} &= (t_1 + \cdots + t_{m_1}) + (t_{m_1+1} + \cdots + t_{m_1+m_2}) \\ &= (rr_1 + \cdots + rr_{m_1}) + ((1-r)\,s_1 + \cdots + (1-t)\,s_{m_2}) \\ &= r\,(r_1 + \cdots + r_{m_1}) + (1-r)\,(s_1 + \cdots + s_{m_2}) \\ &= r + (1-r) = 1 \end{aligned}$$

and

$$\begin{aligned} r\vec{x} + (1-r)\,\vec{y} &= r\left(\sum_{i=1}^{m_1} r_i \vec{x}_i\right) + (1-r)\sum_{i=1}^{m_2} s_i \vec{y}_i \\ &= \sum_{i=1}^{m_1} r r_i \vec{x}_i + \sum_{i=1}^{m_2} (1-r)\,s_i \vec{y}_i \\ &= \sum_{i=1}^{m_1} t_i \vec{z}_i + \sum_{i=m_1+1}^{m_1+m_2} t_i \vec{z}_i \\ &= \sum_{i=1}^{m_1+m_2} t_i \vec{z}_i \\ &\in B \end{aligned}$$

This shows that B is convex.

The previous proposition confirms that B is contained in any convex subset C so that $A \subseteq C$. Hence B is the smallest convex subset containing A, i.e. $B = conv(A)$. □

EXAMPLE 34. *Find the convex hull of A, if*

(1) $A = \{-1, 1\} \subseteq \Re$.

(2) $A = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\} \subseteq \Re^2$

(3) $A = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\} \subseteq \Re^2$

DEFINITION 14. *We define*

$$\Delta_{n-1} = \{(x_1, \ldots, x_n) \in \Re^n : 0 \leq x_1, \ldots, x_n \text{ and } x_1 + \cdots + x_n = 1\}$$

PROPOSITION 3. *The sets Δ_{n-1} are convex. Moreover, if $E_n = \{\vec{e}_1, \ldots, \vec{e}_n\} \subseteq \Re^n$ (the set of all canonical unit vectors), then $conv(E_n) = \Delta_{n-1}$*

PROOF. First, we show that Δ_{n-1} is convex. Let $\vec{x}, \vec{y} \in \Delta_{n-1}$ and let $0 \leq r \leq 1$ be given. We have to show that $r\vec{x} + (1-r)\vec{y} \in \Delta_{n-1}$. We know that $0 \leq x_1, \ldots, x_n$, $0 \leq y_1, \ldots, y_n$ and $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i = 1$. For each index i, this implies that $rx_i + (1-r)y_i \geq 0$. Moreover,

$$\begin{aligned} \sum_{i=1}^n (rx_i + (1-r)y_i) &= r\sum_{i=1}^n x_i + (1-r)\sum_{i=1}^n y_i \\ &= r \cdot 1 + (1-r) \cdot 1 \\ &= 1 \end{aligned}$$

Hence $r\vec{x} + (1-r)\vec{y} \in \Delta_{n-1}$.

Clearly, the set Δ_{n-1} contains all canonical unit vectors. Therefore Δ_{n-1} is a convex set containing E_n. We have to show that Δ_{n-1} is the smallest convex set containing E :

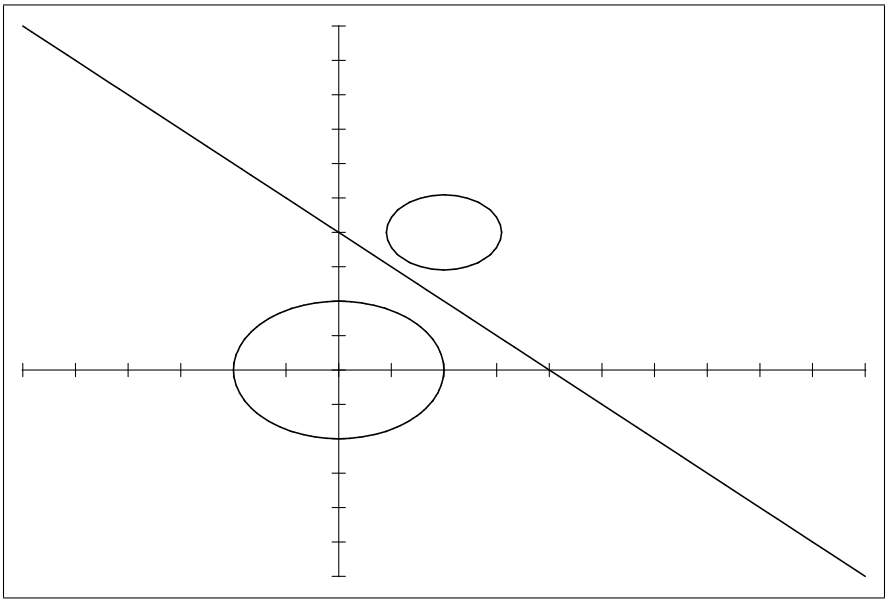
Let A be any convex set with $E_n \subseteq A$. We have to prove that $\Delta_{n-1} \subseteq A$. Let $\vec{x} \in \Delta_{n-1}$. Then, since A is convex, since $\vec{e}_i \in A$ for each i, since $x_1 + \cdots + x_n = 1$ and since $x_1, \ldots, x_n \geq 0$, the last proposition implies that $\vec{x} = x_1\vec{e}_1 + \cdots + x_n\vec{e}_n \in A$. □

DEFINITION 15. *The convex set from the previous example is denoted by Δ_{n-1} and called the $n-1$ dimensional standard simplex.*

EXAMPLE 35. *Draw pictures of the sets Δ_1, Δ_2 and Δ_3.*

14. Interiors of Convex Sets and Separation

In this section, we discuss the interiors of convex sets. We also show investigate conditions under which two convex sets can be separated by a hyperplane.



Two convex sets separated by a line

DEFINITION 16. *Let C be a convex set. The interior of C is defined by*

$$C^o = \{\vec{x} \in C : \textit{the is a number } \varepsilon > 0 \textit{ so that } \vec{y} \in C \textit{ whenever } \|\vec{x} - \vec{y}\| < \varepsilon\}$$

We define the boundary of C by

$$\partial C = \{\vec{x} \in \Re^n : \textit{for each } \varepsilon > 0 \textit{ there is are elements } \vec{y}_1 \in C, \vec{y}_2 \notin C \textit{ so that } \|\vec{x} - \vec{y}_1\|, \|\vec{x} - \vec{y}_2\| < \varepsilon\}$$

PROPOSITION 4. *Let $C \subseteq \Re^n$ be a convex set. Let $\vec{a} \in C$, $\vec{b} \in C^o$, and let $0 \leq r < 1$. Then $r\vec{a} + (1-r)\vec{b} \in C^o$.*

PROOF. This is obvious for $r = 0$. Assume that $0 < r < 1$. We have to find a number $\varepsilon > 0$ so that $\left\|\vec{y} - \left(r\vec{a} + (1-r)\vec{b}\right)\right\| < \varepsilon$ implies that $\vec{y} \in C$. First, pick a number $\varepsilon_1 > 0$ so that $\left\|\vec{y} - \vec{b}\right\| < \varepsilon_1$ implies that $\vec{y} \in C$. Let

$$\varepsilon = (1-r)\,\varepsilon_1$$

Assume that $\vec{y} \in \Re^n$ is given so that $\left\|\vec{y} - \left(r\vec{a} + (1-r)\vec{b}\right)\right\| < \varepsilon$. We have to show that $\vec{y} \in C$. Let

$$\vec{y}_1 = \frac{1}{1-r}\left(\vec{y} - r\vec{a}\right)$$

Then

$$\begin{aligned}\left\|\vec{y}_1-\vec{b}\right\| &= \left\|\frac{1}{1-r}\left(\vec{y}-r\vec{a}\right)-\vec{b}\right\| \\ &= \frac{1}{1-r}\left\|\vec{y}-\left(\left(1-r\right)\vec{b}+r\vec{a}\right)\right\| \\ &< \frac{1}{1-r}\varepsilon=\varepsilon_1\end{aligned}$$

and hence $\vec{y}_1 \in C$ by the choice of ε_1. Now the fact that C is convex yields

$$\vec{y}=\left(1-r\right)\vec{y}_1+r\vec{a}\in C$$

□

THEOREM 24. *Let C be a convex set. Then the interior C^o of C is also convex.*

PROOF. Let $\vec{a},\vec{b}\in C^o$, and let $0<r<1$. Then the last proposition shows that $r\vec{a}+\left(1-r\right)\vec{b}\in C^o$. Hence C^o is convex. □

Topologically, the set $C\cup\partial C$ is the closure of the set C, i.e. it consists of all $\vec{a}\in\Re^n$ so that there is a converging sequence $\left(\vec{a}_n\right)_{n\geq 1}\subseteq C$ with $\vec{a}=\lim_{n\to\infty}\vec{a}_n$. We will use this fact in the proof of the following theorem.

THEOREM 25. *Let C be convex. Then $C\cup\partial C$ is also convex.*

PROOF. Let $\vec{a},\vec{b}\in C\cup\partial C$, and let $0\leq r\leq 1$ Then there are sequences $\left(\vec{a}_n\right)_{n\geq 1},\left(\vec{b}_n\right)_{n\geq 1}\subseteq C$ so that

$$\begin{aligned}\vec{a} &= \lim_{n\to\infty}\vec{a}_n \\ \vec{b} &= \lim_{n\to\infty}\vec{b}_n\end{aligned}$$

Since C is convex, the elements $r\vec{a}_n+\left(1-r\right)\vec{b}_n$ belong to C for all $n\geq 1$. We conclude that

$$\begin{aligned}r\vec{a}+\left(1-r\right)\vec{b} &= r\left(\lim_{n\to\infty}\vec{a}_n\right)+\left(1-r\right)\left(\lim_{n\to\infty}\vec{b}_n\right) \\ &= \lim_{n\to\infty}\left(r\vec{a}_n+\left(1-r\right)\vec{b}_n\right) \\ &\in C\cup\partial C\end{aligned}$$

□

EXAMPLE 36. *Show that the set $C=\left\{[x,y]\in\Re^2:x^2+y^2<1\right\}\cup\left\{[1,0],[0,1]\right\}$ is convex, and find the sets C^o, ∂C, and $C\cup\partial C$.*

DEFINITION 17. *Let C be a convex set so that $C=C^o$. Then C is called an open convex set.*

PROPOSITION 5. *If $T:\Re^n\to\Re^m$ is a linear map, and if C is convex, then $T\left(C\right)$ is convex.*

PROOF. Let $\vec{y}_1,\vec{y}_2\in T\left(C\right)$ and let r be a real number with $0\leq r\leq 1$. We have to show that $r\vec{y}_1+\left(1-r\right)\vec{y}_2\in T\left(C\right)$. We know that there are vectors $\vec{x}_1,\vec{x}_2\in C$ so that $T\left(\vec{x}_1\right)=\vec{y}_1$ and $T\left(\vec{x}_2\right)=\vec{y}_2$. Since C is convex, the vector $r\vec{x}_1+\left(1-r\right)\vec{x}_2$ belongs to C. Hence $r\vec{y}_1+\left(1-r\right)\vec{y}_2=rT\left(\vec{x}_1\right)+\left(1-r\right)T\left(\vec{x}_2\right)=T\left(r\vec{x}_1+\left(1-r\right)\vec{x}_2\right)\in T\left(C\right)$. □

DEFINITION 18. *A linear map $T : \Re^n \to \Re^m$ is surjective, if for each $\vec{y} \in \Re^m$ there is a vector $\vec{x} \in \Re^n$ so that $T(\vec{x}) = \vec{y}$. In other words, a linear map is surjective if the range space of T is $\Re^m$*

If the linear map T is represented by a matrix M, then T is surjective if and only if the columns of M span $\Re^m$.

PROPOSITION 6. *If $T : \Re^n \to \Re^m$ is a surjective linear map, then there is a number $R > 0$ so that for every $\vec{y} \in \Re^m$ with $\|\vec{y}\| \leq 1$ there is a vector $\vec{x} \in \Re^n$ with $\|\vec{x}\| \leq R$ and $T(\vec{x}) = \vec{y}$.*

PROOF. The vectors $T(\vec{e}_1), \ldots, T(\vec{e}_n)$ span $\Re^m$. Hence we can choose a basis from the set $\{T(\vec{e}_1), \ldots, T(\vec{e}_n)\}$. After renumbering the coordinates, we may assume that $T(\vec{e}_1), \ldots, T(\vec{e}_m)$ is a a basis of $\Re^m$. For each index i with $1 \leq i \leq m$ let

$$\vec{a}_i = T(\vec{e}_i)$$

Then the vectors $\vec{a}_1, \ldots, \vec{a}_m$ are linearly independent. Let

$$M = [\vec{a}_i \cdot \vec{a}_j]_{1 \leq i,j \leq m}$$

Then for every $\vec{0} \neq \vec{z} = [z_1, \ldots, z_m] \in \Re^m$ we have

$$\begin{aligned} 0 &< \|z_1\vec{a}_1 + \cdots + z_m\vec{a}_m\|^2 \\ &= (z_1\vec{a}_1 + \ldots + z_m\vec{a}_m) \cdot (z_1\vec{a}_1 + \cdots + z_m\vec{a}_m) \\ &= [z_1, \ldots, z_m] M \begin{bmatrix} z_1 \\ \vdots \\ z_m \end{bmatrix} \end{aligned}$$

Hence the matrix M is positive definite and therefore all eigenvalues of M are strictly positive. Let $\lambda > 0$ be the smallest eigenvalue of M. Then

$$[z_1, \ldots, z_m] M \begin{bmatrix} z_1 \\ \vdots \\ z_m \end{bmatrix} \geq \lambda \|[z_1, \ldots, z_m]\|^2$$

and hence

$$\|z_1\vec{a}_1 + \cdots + z_m\vec{a}_m\| \geq \sqrt{\lambda} \|[z_1, \ldots, z_m]\|$$

We let

$$R = \frac{1}{\sqrt{\lambda}}$$

Let $\vec{y} \in \Re^m$ be given with $\|\vec{y}\| \leq 1$. We have to find a vector $\vec{x} \in \Re^n$ with $T(\vec{x}) = \vec{y}$ and $\|\vec{x}\| \leq R$. Pick numbers $z_1, \ldots, z_m$ so that

$$\vec{y} = z_1\vec{a}_1 + \cdots + z_m\vec{a}_m$$

and define $\vec{x}$ by

$$\vec{x} = [z_1, \ldots, z_m, 0, \ldots, 0]$$

Then

$$T(\vec{x}) = \vec{y}$$

and

$$\begin{aligned}\|\vec{x}\| &= \|[z_1,\ldots,z_m,0\ldots,0]\| \\ &= \|[z_1,\ldots,z_m]\| \\ &\leq \frac{1}{\sqrt{\lambda}}\|z_1\vec{a}_1+\cdots+z_m\vec{a}_m\| \\ &= R\|\vec{y}\| \leq R\end{aligned}$$

□

PROPOSITION 7. *Let $C \subseteq \Re^n$ be an open set, and assume that $T : \Re^n \to \Re^m$ is a surjective linear map. Then $T(C)$ is open.*

PROOF. Let $\vec{y} \in T(C)$. We have to find a number $\varepsilon > 0$ so that $\|\vec{y}_1 - \vec{y}\| < \varepsilon$ implies that $\vec{y}_1 \in C$.

From the last proposition we know that there is a number $C > 0$ so that for every $\vec{v} \in \Re^m$ with $\|\vec{v}\| \leq 1$ there is an element $\vec{u} \in \Re^n$ with $\|\vec{u}\| \leq C$ and $T(\vec{u}) = \vec{v}$.

For the given $\vec{y} \in T(C)$ pick $\vec{x} \in C$ so that $T(\vec{x}) = \vec{y}$. Since C is open, there is a number $\varepsilon_1 > 0$ so that $\|\vec{x} - \vec{x}_1\| < \varepsilon_1$ implies that $\vec{x}_1 \in C$. We define

$$\varepsilon = \frac{\varepsilon_1}{2R}$$

Assume that $\|\vec{y}_1 - \vec{y}\| < \varepsilon$. We have to show that $\vec{y}_1 \in T(C)$.

First, note that

$$\left\|\frac{1}{\varepsilon}(\vec{y}_1 - \vec{y})\right\| < 1$$

Hence there is a vector $\vec{x}_d \in \Re^n$ with $\|\vec{x}_d\| \leq R$ and $T(\vec{x}_d) = \frac{1}{\varepsilon}(\vec{y}_1 - \vec{y})$. It follows that

$$\begin{aligned}T(\vec{x}+\varepsilon\vec{x}_d) &= T(\vec{x}) + \varepsilon T(\vec{x}_d) \\ &= \vec{y} + (\vec{y}_1 - \vec{y}) \\ &= \vec{y}_1\end{aligned}$$

Moreover, since

$$\begin{aligned}\|\vec{x} - (\vec{x}+\varepsilon\vec{x}_d)\| &= \varepsilon\|\vec{x}_d\| \\ &= \frac{\varepsilon_1}{2R}\cdot R \\ &= \frac{\varepsilon_1}{2} < \varepsilon_1\end{aligned}$$

the element $\vec{x} + \varepsilon\vec{x}_d$ belongs to C.Hence $\vec{y}_1 \in T(C)$. □

PROPOSITION 8. *Let $T : \Re^n \to \Re^m$ be a surjective linear map, and let $C \subseteq \Re^n$ be a convex open set. Then $T(C)$ is a convex, open subset of $\Re^m$.*

PROOF. This follows immediately from the previous propositions. □

Here is another consequence of the previous proposition:

THEOREM 26. *The notion of "open convex set" is invariant under the change of coordinate systems.*

PROOF. A change of coordinate systems may the thought of as a linear bijection $T : \Re^n \to \Re^n$. □

PROPOSITION 9. *Let $C \subseteq \Re^2$ be an open, convex set and assume that $[0,0] \notin C$. Then there is a line ℓ passing through the origin so that $\ell \cap C = \emptyset$.*

PROOF. Since $[0,0] \notin C$, we find that $\|\vec{x}\| > 0$ for all $\vec{x} \in C$. Let S_1 be the unit circle:

$$\begin{aligned} S_1 &= \left\{\vec{x} \in \Re^2 : \|\vec{x}\| = 1\right\} \\ &= \left\{[\cos\alpha, \sin\alpha] : 0 \leq \alpha < 2\pi\right\} \end{aligned}$$

We define a map

$$\nu : C \to S_1$$

by

$$\nu\left(\vec{x}\right) = \frac{\vec{x}}{\|\vec{x}\|}$$

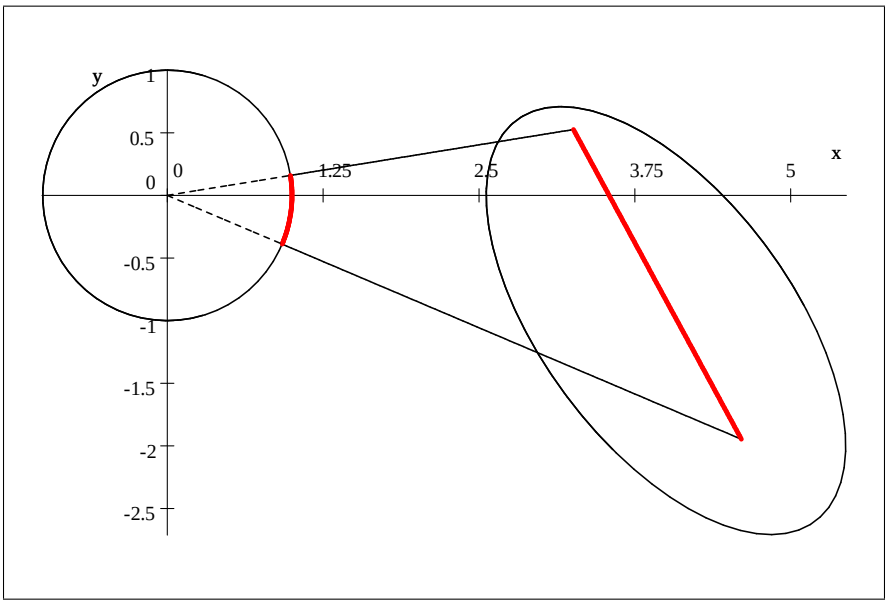
Let $\vec{x}, \vec{y} \in C$, and pick the angles α and β so that

$$\begin{aligned} \nu\left(\vec{x}\right) &= [\cos\alpha, \sin\alpha] \\ \nu\left(\vec{y}\right) &= [\cos\beta, \sin\beta] \end{aligned}$$

The points

$$\nu\left(r\vec{x} + (1-r)\,\vec{y}\right) = \frac{r\vec{x} + (1-r)\,\vec{y}}{\|r\vec{x} + (1-r)\,\vec{y}\|}$$

with $0 \leq r \leq 1$ cover the (short) arc between $\nu\left(\vec{x}\right)$ and $\nu\left(\vec{y}\right)$. The following illustration verifies and explains this statement:



Hence $\nu\left(C\right)$ is an arc in S_1. Since C is open, the endpoints of the arc do not belong to $\nu\left(C\right)$, i.e. $\nu\left(C\right)$ is an open arc.

If the arc $\nu\left(C\right)$ would contain a point $\vec{a}$ and its antipode $-\vec{a}$, then there would be elements $\vec{x}, \vec{y} \in C$ so that

$$\begin{aligned} \nu\left(\vec{x}\right) &= \vec{a} \\ \nu\left(\vec{y}\right) &= -\vec{a} \end{aligned}$$

and hence

$$\frac{\vec{x}}{\|\vec{x}\|} = -\frac{\vec{y}}{\|\vec{y}\|}$$

We find that

$$\begin{aligned} \frac{\vec{x}}{\|\vec{x}\|} + \frac{\vec{y}}{\|\vec{y}\|} &= \vec{0} \\ \|\vec{y}\|\,\vec{x} + \|\vec{x}\|\,\vec{y} &= 0 \\ \frac{\|\vec{y}\|}{\|\vec{x}\| + \|\vec{y}\|}\vec{x} + \frac{\|\vec{x}\|}{\|\vec{x}\| + \|\vec{y}\|}\vec{y} &= \vec{0} \end{aligned}$$

If we let $r = \frac{\|\vec{y}\|}{\|\vec{x}\|+\|\vec{y}\|}$, then $(1-r) = \frac{\|\vec{x}\|}{\|\vec{x}\|+\|\vec{y}\|}$ and hence $r\vec{x} + (1-r)\,\vec{y} = \vec{0} \in C$, contradicting our assumption that $\vec{0} \notin C$.

Since $\nu(C)$ does not contain antipodal points, its arc length is less than or equal to π. Hence the arc length of its complement $S_1 \setminus \nu(C)$ is greater than or equal to π. Since $S_1 \setminus \nu(C)$ is a closed arc, it has to contain both endpoints. Therefore $S_1 \setminus \nu(C)$ contains a point $\vec{p}$ and its antipode $-\vec{p}$. It follows that the line ℓ through $\vec{p}$ cannot intersect C. □

PROPOSITION 10. *Let $C \subseteq \Re^n$ be an open convex set so that $\vec{0} \notin C$. If $n \geq 2$, then there is a line $\ell \subseteq \Re^n$ with $\vec{0} \in \ell$ so that $\ell \cap C = \emptyset$.*

PROOF. Let $\ell_1, \ell_2 \subseteq \Re^n$ be different lines, and let Π be the plane spanned by ℓ_1 and ℓ_2. If Π does not intersect C, then neither ℓ_1 nor ℓ_2 will intersect C, and we have found a line that does not intersect C. Otherwise, since Π can be identified with $\Re^2$, we may apply the last theorem to $\Pi \cap C$ and find a line ℓ in Π that does not intersect C. □

THEOREM 27. *Let $C \subseteq \Re^n$ be an open convex set so that $\vec{0} \notin C$. Then there is a hyperplane $H \subseteq \Re^n$ with $\vec{0} \in H$ so that $H \cap C = \emptyset$.*

PROOF. Let H be a maximal linear subspace of $\Re^n$ that does not intersect C. We can conclude from the last proposition that the dimension of H is at least 1. We would like to show that the dimension of H is $n-1$.

First, we find a basis $\vec{b}_1, \ldots, \vec{b}_n$ of $\Re^n$ so that H is spanned by $\vec{b}_1, \ldots, \vec{b}_k$. After changing coordinate systems, we may assume that

$$H = \{[x_1, \ldots, x_n] \in \Re^n : x_{k+1} = \cdots = x_n = 0\}$$

and we have to show that $k = n-1$.

Assume that $k < n-1$. We define a linear map $T : \Re^n \to \Re^{n-k}$ by

$$T[x_1, \ldots, x_n] = [x_{k+1}, \ldots, x_n]$$

Let

$$\begin{aligned} C_1 &= T(C) \\ &= \left\{[x_{k+1}, \ldots, x_n] \in \Re^{n-k} : \text{there are number } x_1, \ldots, x_k \in \Re \text{ so that } [x_1, \ldots, x_n] \in C\right\} \end{aligned}$$

We know from previous proposition that $C_1 \subseteq \Re^{n-k}$ is an open convex set. Moreover, if $\vec{0}$ would belong to C_1, then there would be number $x_1, \ldots, x_k$ so that $[x_1, \ldots, x_k, 0, \ldots, 0] \in C$. But

$$\begin{aligned} [x_1, \ldots, x_k, 0, \ldots, 0] &\in \{[x_1, \ldots, x_n] \in \Re^n : x_{k+1} = \cdots = x_n = 0\} \\ &= H \end{aligned}$$

and this would lead to the contradiction that $[x_1, \ldots, x_k, 0, \ldots, 0] \in C \cap H = \emptyset$.

Since we are assuming that $k < n-1$, the dimension of $\Re^{n-k}$ is at least 2. The last proposition implies that there is a line $\ell \subseteq \Re^{n-k}$ so that $\ell \cap C_1 = \emptyset$. The line ℓ is of the form.

$$\ell = \{r\,[a_{k+1}, \ldots, a_n] : r \in \Re\}$$

Then the vector $[0, \ldots, 0, a_{k+1}, \ldots, a_n] \in \Re^n$ does not belong to H. Let

$$H_1 = \{[x_1, \ldots, x_k, ra_{k+1}, \ldots, ra_n] : x_1, \ldots, x_k, r \in \Re\}$$

Then H_1 is a linear subspace containing H, and H_1 has dimension $k+1$. The maximality of H implies that H_1 intersects C. Hence there is a vector of the form $[x_1, \ldots, x_k, ra_{k+1}, \ldots, ra_n]$ that belongs to C. It would follow that

$$[ra_{k+1}, \ldots, ra_n] \in C_1 \cap \ell$$

a contradiction. □

The next corollary is actually an equivalent reformulation of the last theorem:

COROLLARY 2. *Let $C \subseteq \Re^n$ be a open convex set. Then there is a vector $\vec{a} \in \Re^n$ so that $\vec{a} \cdot \vec{c} > 0$ for all $\vec{c} \in C$.*

PROOF. By the last theorem, there is a hyperplane H is that $H \cap C = \emptyset$. Let $\vec{n}$ be a vector orthogonal to H. Then

$$H = \{\vec{x} \in \Re^n : \vec{x} \cdot \vec{n} = 0\}$$

Since $C \cap H = \emptyset$, this implies that $\vec{c} \cdot \vec{n} \neq 0$ for all $\vec{c} \in C$. We would like to show that $\vec{c} \cdot \vec{n}$ is either always positive of always negative on C.

Assume that we could find elements $\vec{c}_1, \vec{c}_2 \in C$ so that

$$\begin{aligned} r &= \vec{c}_1 \cdot \vec{n} < 0 \\ s &= \vec{c}_2 \cdot \vec{n} > 0 \end{aligned}$$

Then $-r > 0$. Let

$$\lambda = \frac{s}{s-r}$$

Then

$$0 < \lambda < 1$$

and

$$1 - \lambda = 1 - \frac{s}{s-r} = \frac{-r}{s-r}$$

Hence

$$\vec{c} = \lambda \vec{c}_1 + (1-\lambda)\,\vec{c}_2 \in C$$

This leads to the contradiction

$$\begin{aligned} \vec{c} \cdot \vec{n} &= \lambda \vec{c}_1 \cdot \vec{n} + (1-\lambda)\,\vec{c}_2 \cdot \vec{n} \\ &= \frac{s}{s-r}\,(\vec{c}_1 \cdot \vec{n}) + \left(\frac{-r}{s-r}\right)(\vec{c}_2 \cdot \vec{n}) \\ &= \frac{s}{s-r} r + \left(\frac{-r}{s-r}\right) s \\ &= 0 \end{aligned}$$

If $\vec{c} \cdot \vec{n}$ is always positive, we pick $\vec{a} = \vec{n}$. Otherwise, if $\vec{c} \cdot \vec{n}$ is always negative, we let $\vec{a} = -\vec{n}$. □

In order to show that two open convex sets can be separated by a hyperplane, we need yet another proposition:

PROPOSITION 11. *Let $C, D \subseteq \Re^n$ be two convex sets, and assume that C is open. Then*

$$C - D = \left\{ \vec{c} - \vec{d} : \vec{c} \in C, \vec{d} \in D \right\}$$

is an open and convex subset.

PROOF. First, we show that $C - D$ is convex: Let $\vec{x}, \vec{y} \in C - D$ and let $0 \leq r \leq 1$. We have to show that $r\vec{x} + (1-r)\,\vec{y} \in C - D$. We write

$$\begin{aligned} \vec{x} &= \vec{c}_1 - \vec{d}_1 \\ \vec{y} &= \vec{c}_2 - \vec{d}_2 \end{aligned}$$

with $\vec{c}_1, \vec{c}_2 \in C$ and $\vec{d}_1, \vec{d}_2 \in D$. Then $r\vec{c}_1 + (1-r)\,\vec{c}_2 \in C$ and $r\vec{d}_1 + (1-r)\,\vec{d}_2 \in D$. We conclude that

$$\begin{aligned} r\vec{x} + (1-r)\,\vec{y} &= r\left(\vec{c}_1 - \vec{d}_1\right) + (1-r)\left(\vec{c}_2 - \vec{d}_2\right) \\ &= \left(r\vec{c}_1 + (1-r)\,\vec{c}_2\right) - \left(r\vec{d}_1 + (1-r)\,\vec{d}_2\right) \\ &\in C - D \end{aligned}$$

It remains to show that $C - D$ is open. Let $\vec{x} \in C - D$. We have to find a number $\varepsilon > 0$ so that $\|\vec{x} - \vec{y}\| < \varepsilon$ implies that $\vec{y} \in C - D$. We know that

$$\vec{x} = \vec{c} - \vec{d}$$

with $\vec{c} \in C$ and $\vec{d} \in D$. Since C is open, we can pick a number $\varepsilon > 0$ so that $\|\vec{c} - \vec{u}\| < \varepsilon$ implies $\vec{u} \in C$.

Now assume that $\|\vec{x} - \vec{y}\| < \varepsilon$. Then

$$\vec{y} = \vec{a} - \vec{b}$$

with $\vec{a} \in C$ and $\vec{b} \in D$. Hence

$$\begin{aligned} \left\| \vec{c} - \left(\vec{d} + \vec{a} - \vec{b}\right) \right\| &= \left\| \vec{c} - \vec{d} - \left(\vec{a} - \vec{b}\right) \right\| \\ &= \|\vec{x} - \vec{y}\| \\ &< \varepsilon \end{aligned}$$

We conclude that $\vec{d} + \vec{a} - \vec{b} \in C$. Hence

$$\begin{aligned} \vec{y} &= \vec{a} - \vec{b} \\ &= \left(\vec{d} + \vec{a} - \vec{b}\right) - \vec{d} \\ &\in C - D \end{aligned}$$

□

THEOREM 28 (Separation Theorem). *Let $C, D \subseteq \Re^n$ be two convex set. Assume that the interior of C is non-empty and that $C^o \cap D = \emptyset$. Then there is a hyperplane H of $\Re^n$ that separates C and D in the following sense: There is a vector $\vec{a} \in \Re^n$ and a number b so that*

$$\begin{aligned} \vec{a} \cdot \vec{c} &\geq b \text{ for all } \vec{c} \in C \text{ and} \\ \vec{a} \cdot \vec{d} &\leq b \text{ for all } \vec{d} \in D \end{aligned}$$

Moreover $\vec{a} \cdot \vec{c} > b$ for all $\vec{c} \in C^o$.

PROOF. Since C^o and D are disjoint, $\vec{0} \notin C^o - D$. Hence $C^o - D$ is an open convex set not containing the origin. Therefore we can find a vector $\vec{a}$ so that $\vec{a} \cdot \left(\vec{c} - \vec{d}\right) > 0$ for all $\vec{c} - \vec{d} \in C^o - D$. We now show that this implies that

$$\vec{a} \cdot \left(\vec{c} - \vec{d}\right) \geq 0 \text{ for all } \vec{c} \in C \text{ and all } \vec{d} \in D$$

Assume that this were not the case. Then we can find elements $\vec{c}_0 \in C$ and $\vec{d}_0 \in D$ so that $\vec{a} \cdot \left(\vec{c}_0 - \vec{d}_0\right) < 0$. Let $\vec{c}_1 \in C^o$ and $\vec{d}_1 \in D$ be fixed. We know that for each number λ with $0 < \lambda < 1$ the element $\lambda \vec{c}_1 + (1 - \lambda)\, \vec{c}_0$ belongs to the interior of C. Hence $\lambda \left(\vec{c}_1 - \vec{d}_1\right) + (1 - \lambda) \left(\vec{c}_0 - \vec{d}_0\right) \in C^o - D$. and therefore

$$\left(\lambda \left(\vec{c}_1 - \vec{d}_1\right) + (1 - \lambda) \left(\vec{c}_0 - \vec{d}_0\right)\right) \cdot \vec{a} > 0$$

for each value of λ with $0 < \lambda < 1$. This leads to the contradiction

$$\begin{aligned} 0 &\leq \lim_{\lambda \to 0} \left(\lambda \left(\vec{c}_1 - \vec{d}_1\right) + (1 - \lambda) \left(\vec{c}_0 - \vec{d}_0\right)\right) \cdot \vec{a} \\ &= \vec{a} \cdot \left(\vec{c}_0 - \vec{d}_0\right) < 0 \end{aligned}$$

We conclude that

$$\vec{a} \cdot \vec{c} \geq \vec{a} \cdot \vec{d}$$

whenever $\vec{c} \in C$ and $\vec{d} \in D$ Let

$$b = \sup \left\{\vec{a} \cdot \vec{d} : \vec{d} \in D\right\}$$

i.e. b is the least upper bound of the set $\left\{\vec{a} \cdot \vec{d} : \vec{d} \in D\right\}$. Then, by definition $\vec{a} \cdot \vec{d} \leq b$ for all $\vec{d} \in D$. Moreover, since each number $\vec{a} \cdot \vec{c}$ is an upper bound of $\left\{\vec{a} \cdot \vec{d} : \vec{d} \in D\right\}$, it follows that each number $\vec{a} \cdot \vec{c}$ is at least as large as the least upper bound:

$$\vec{a} \cdot \vec{c} \geq b \text{ for all } \vec{c} \in C$$

This proves the first part of the theorem.

Assume now that - contrary to the statement of the theorem - equality holds for at least one $\vec{c}_0 \in C$:

$$\vec{a} \cdot \vec{c}_0 = b$$

Since C is open, there is an $\varepsilon > 0$ so that $\vec{y} \in C$ whenever $\|\vec{c}_0 - \vec{y}\| < \varepsilon$. Let

$$\vec{c}_1 = \vec{c}_0 + \frac{\varepsilon}{2 \|\vec{a}\|} \vec{a}$$

Then

$$\|\vec{c}_0 - \vec{c}_1\| = \frac{\varepsilon}{2} < \varepsilon$$

and therefore $\vec{c}_1 \in C$. However, this would lead to the contradiction that

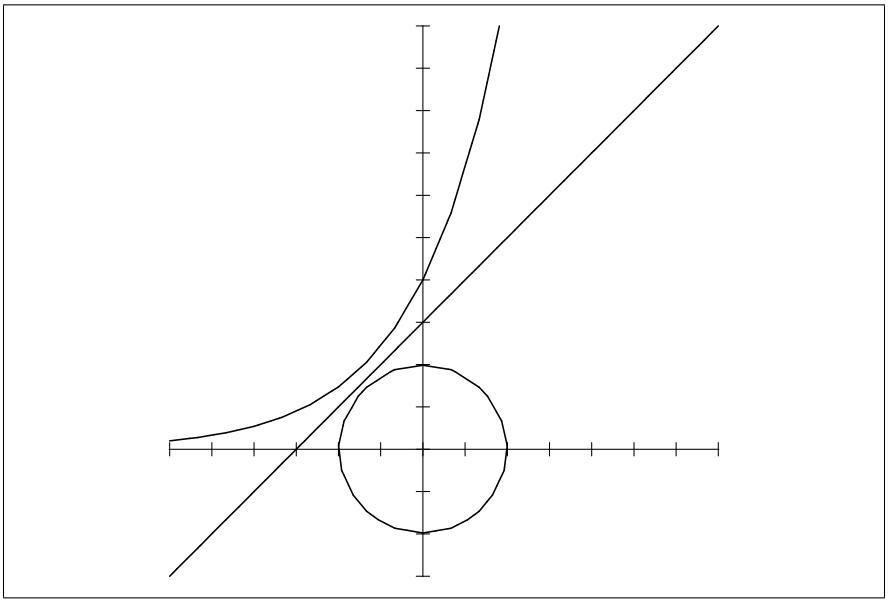
$$\begin{aligned} \vec{c}_1 \cdot \vec{a} &= \left(\vec{c}_0 + \frac{\varepsilon}{2 \|\vec{a}\|} \vec{a}\right) \cdot \vec{a} \\ &= \vec{c}_0 \cdot \vec{a} + \frac{\varepsilon}{2 \|\vec{a}\|} (\vec{a} \cdot \vec{a}) \\ &= b + \frac{\varepsilon \|\vec{a}\|}{2} \\ &> b \end{aligned}$$

□

The next example should illustrate that separation of convex sets is closely related to the problem of finding extreme values of certain function.

EXAMPLE 37. *Show that the sets* $C = \left\{[x, y] : x^2 + y^2 \leq 1\right\}$ *and* $D = \{[x, y] : y \geq 2e^x\}$ *are convex and disjoint. Find a line in* $\Re^2$ *separating those two sets.*

We know that C is convex. To show that D is convex without any further theoretical preparation seems to be a challenge. We delay this until later. For now, we plot both sets:



It seems that the line $y = x + \frac{3}{2}$ separates both convex sets. Indeed, if we would like to prove that $2e^x > x + \frac{3}{2}$, we have to find the minimum value of $2e^x - x - \frac{3}{2}$. The critical points of $2e^x - x - \frac{3}{2}$ are the solutions of

$$2e^x = 1$$

It follows that

$$x = -\ln 2$$

We have to show that at this critical point x_0 we have $2e^{x_0} > x_0 + \frac{3}{2}$. Starting with

$$\ln 4 > \ln e = 1$$

we find that

$$\begin{aligned} 2\ln 2 &> 1 \\ \ln 2 &> \frac{1}{2} \\ \ln 2 - \frac{1}{2} &> 0 \\ 2e^{-\ln 2} + \ln 2 - \frac{3}{2} &= \ln 2 - \frac{1}{2} > 0 \end{aligned}$$

Hence indeed

$$2e^{x_0} - x_0 - \frac{3}{2} > 0$$

Next, we have to show that $x^2 + y^2 \leq 1$ implies that $y < x + \frac{3}{2}$. Assuming that $x^2 + y^2 \leq 1$, we have $y \leq \sqrt{1 - x^2}$. Hence we have to show that $\sqrt{1 - x^2} < x + \frac{3}{2}$. So we have to show that $0 < x + \frac{3}{2} - \sqrt{1 - x^2}$ for all values of x with $-1 \leq x \leq 1$. Again, we have to find a minimum value. The critical points are the solutions of

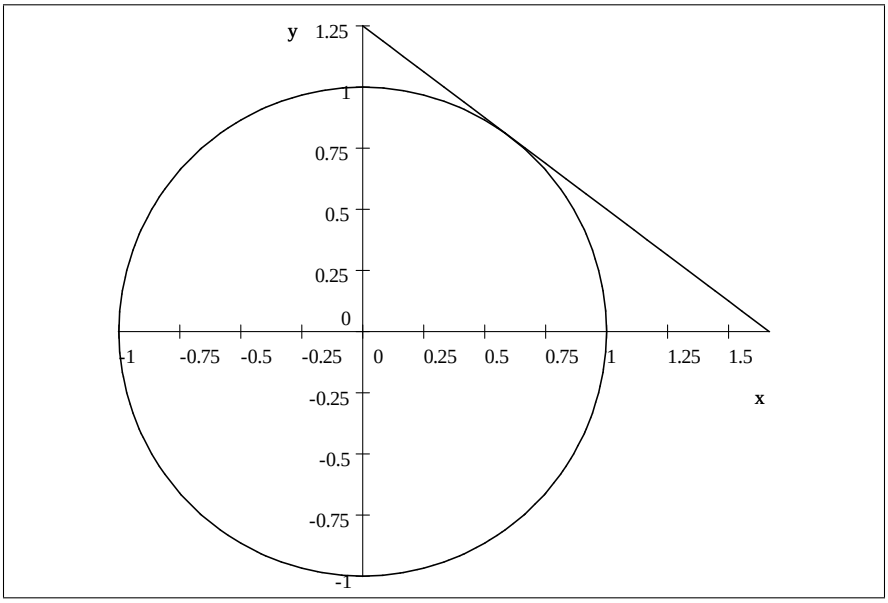
$$\begin{aligned} 1 + \frac{2x}{2\sqrt{1 - x^2}} &= 0 \\ x &= -\sqrt{1 - x^2} \\ x^2 &= 1 - x^2 \\ x &= \pm\frac{1}{2}\sqrt{2} \end{aligned}$$

Since $x = -\sqrt{1 - x^2} \leq 0$, we find that $x = -\frac{1}{2}\sqrt{2}$. Therefore we obtain the minimum value of $x + \frac{3}{2} - \sqrt{1 - x^2}$ is equal to $-\frac{1}{2}\sqrt{2} + \frac{3}{2} - \frac{1}{2}\sqrt{2} = \frac{3}{2} - \sqrt{2} > 0$.

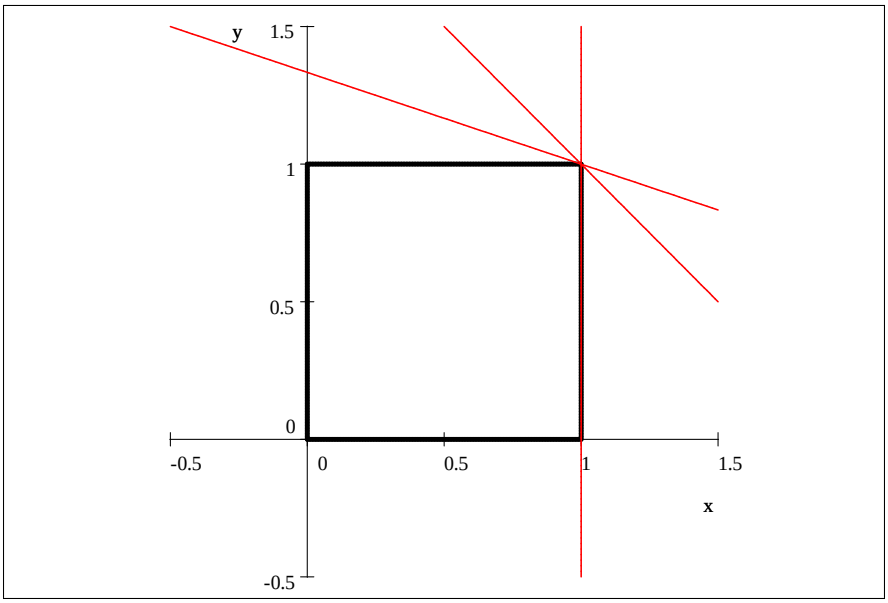
15. Supporting Hyperplanes and Extreme Points

DEFINITION 19. *Let $C \subseteq \Re^n$ be a convex set. A supporting hyperplane of C is an affine hyperplane $H \subseteq \Re^n$ so that $C \cap H \neq \emptyset$ and so that C is contained in one of the two half-spaces determined by H. If $\vec{x}_0 \in C \cap H$, then we say that H supports C at $\vec{x}_0$.*

For example, a tangent line to a circle given by $x^2 + y^2 = 1$ would be a supporting hyperplane of the disk:



EXAMPLE 38. *Find some supporting hyperplanes of the square* $\{[x, y] : 0 \leq x, y \leq 1\}$.



Square (black) and Supporting Lines at $[1,1]$ *(red)*

The last example shows that there can be several hyperplanes supporting a convex set C at a point $\vec{x}_0$, and it also shows that a hyperplane can support C and more than one $\vec{x} \in C$.

In general, a supporting hyperplane H is determined by a vector $\vec{a}$ and a number d so that

(1) $\vec{x} \cdot \vec{a} \leq d$ for all $\vec{x} \in C$.
(2) $\vec{x}_0 \cdot \vec{a} = d$ for at least one vector $\vec{x}_0 \in C$.

Hyperplanes support C only at boundary points:

THEOREM 29. *If C is a convex set, and if the hyperplane H supports C at $\vec{x}_0$, then $\vec{x}_0 \in \partial C$.*

PROOF. Let H be a hyperplane supporting C at $\vec{x}_0$. Then there is a vector $\vec{a}$ and a number d so that
(1) $\vec{x} \cdot \vec{a} \leq d$ for all $\vec{x} \in C$ and
(2) $\vec{x}_0 \cdot \vec{a} = d$ for at least one vector $\vec{x}_0 \in C$.
If $\varepsilon > 0$ is given, we define

$$\vec{y} = \vec{x}_0 + \frac{\varepsilon}{2\left\|\vec{a}\right\|}\vec{a}$$

Then

$$\begin{aligned} \vec{a} \cdot \vec{y} &= \vec{a} \cdot \vec{x}_0 + \frac{\varepsilon}{2\left\|\vec{a}\right\|}\left\|\vec{a}\right\|^2 \\ &= d + \frac{\left\|\vec{a}\right\|}{2}\varepsilon \\ &> d \end{aligned}$$

and hence $\vec{y} \notin C$. Moreover

$$\begin{aligned} \left\|\vec{x}_0 - \vec{y}\right\| &= \left\|\vec{x}_0 - \left(\vec{x}_0 + \frac{\varepsilon}{2\left\|\vec{a}\right\|}\vec{a}\right)\right\| \\ &= \left\|\frac{\varepsilon}{2\left\|\vec{a}\right\|}\vec{a}\right\| \\ &= \frac{\varepsilon}{2} < \varepsilon \end{aligned}$$

It follows from the definition of ∂C that $\vec{x}_0 \in \partial C$. □

The converse of the last theorem is also true. We start the proof of this fact with a proposition:

PROPOSITION 12. *Let $C \subseteq \Re^n$ be a convex set, and assume that $\vec{0} \in C$. Then C contains a basis of $\Re^n$ if and only if C contains an interior point.*

PROOF. Assume that $\vec{c}_1, \ldots, \vec{c}_n \in C$ are a basis of $\Re^n$. After introducing a new coordinate system, we may assume that $\vec{c}_i = \vec{e}_i$, the i^{th} unit vector. Every element of the form $\vec{x} = [x_1, \ldots, x_n]$ with $0 \leq x_1, \ldots, x_n$ and $x_1 + \cdots + x_n \leq 1$ belongs to C, because it is the convex combination of $\vec{e}_1, \ldots, \vec{e}_n$ and $\vec{0}$:

$$\vec{x} = x_1\vec{e}_1 + \cdots + x_n\vec{e}_n + (1 - x_1 - \cdots - x_n)\,\vec{0}$$

Hence the open set

$$U = \{[x_1, \ldots, x_n] : 0 < x_1, \ldots, x_n \text{ and } x_1 + \cdots + x_n < 1\}$$

is a subset of C, i.e. C contains an interior point.

Conversely, assume that C contains an interior point $\vec{x}$. Then there is a number $\varepsilon > 0$ so that $\|\vec{x} - \vec{y}\| < \varepsilon$ implies that $\vec{y} \in C$. The vectors $\vec{x} + \frac{\varepsilon}{2}\vec{e}_i$ have distance less than ε from $\vec{x}$ and therefore belong to C. Since every vector $\vec{v} = [v_1, \ldots, v_n]^T \in \Re^n$ can be written in the form

$$\begin{aligned} \vec{v} &= \frac{2}{\varepsilon} v_1 \left(\frac{\varepsilon}{2}\vec{e}_1\right) + \cdots + \frac{2}{\varepsilon} v_n \left(\frac{\varepsilon}{2}\vec{e}_n\right) \\ &= \frac{2}{\varepsilon} v_1 \left(\frac{\varepsilon}{2}\vec{e}_1 + \vec{x} - \vec{x}\right) + \cdots + \frac{2}{\varepsilon} v_n \left(\frac{\varepsilon}{2}\vec{e}_n + x - \vec{x}\right) \\ &= \frac{2}{\varepsilon} v_1 \left(\frac{\varepsilon}{2}\vec{e}_1 + \vec{x}\right) + \cdots + \frac{2}{\varepsilon} v_n \left(\frac{\varepsilon}{2}\vec{e}_n + x\right) - \left(\frac{2}{\varepsilon} v_1 + \cdots + \frac{2}{\varepsilon} v_n\right) \vec{x} \end{aligned}$$

the vectors $\vec{x} + \frac{\varepsilon}{2}\vec{e}_1, \ldots, \vec{x} + \frac{\varepsilon}{2}\vec{e}_n, \vec{x} \in C$ span $\Re^n$. Using the basis selection theorem, we can pick a basis of $\Re^n$ from those vectors, and hence C contains a basis of $\Re^n$. □

THEOREM 30. *If C contains an interior point $\vec{x}_0$, then every $\vec{x} \in C$ is a limit of a sequence of vectors $\vec{x}_n$ in the interior of C.*

PROOF. If $\vec{x}_0$ is in the interior of C, if $\vec{x} \in C$, and if $0 \leq \lambda < 1$, then we know that $\lambda\vec{x} + (1 - \lambda)\vec{x}_0 \in C^o$. Hence the vectors $\vec{x}_n = \left(1 - \frac{1}{n}\right)\vec{x} + \frac{1}{n}\vec{x}_0$ belong to the interior of C. Since $\lim_{n\to\infty} \vec{x}_n = \vec{x}$, every $\vec{x}$ is the limit of elements in the interior of C. □

THEOREM 31. *Let $C \subseteq \Re^n$ be a closed, bounded convex set and let $\vec{x}_0 \in C$ be given so that $\vec{x}_0 \in \partial C$. Then there is a hyperplane H that supports C at $\vec{x}_0$.*

PROOF. Let $\vec{x}_0 \in \partial C$.

If C does not contain an interior point, then $C - \{\vec{x}_0\}$ does not contain an interior point either. Hence by the previous proposition, $C - \{\vec{x}_0\}$ does not contain a basis of $\Re^n$. Let V be the linear span of $C - \{\vec{x}_0\}$. Then $\dim V < n$. If necessary, add linearly independent elements to V so that V together with those added elements span an hyperplane H_1.Then $C - \{\vec{x}_0\} \subseteq H_1$. Let

$$H = \vec{x}_0 + H_1$$

It follows that $C \subseteq H_1$. In particular, C is contained in a half-space determined by H_1. Moreover,

$$\vec{x}_0 \in C \cap H_1$$

and therefore H_1 is a hyperplane supporting C at $\vec{x}_0$.

Assume that C contains an interior point. Then $C^o \neq \emptyset$ and $\{\vec{x}_0\} \cap C^o = \emptyset$. The Separation Theorem, yields a hyperplane H separating the convex set C^o from the convex set $\{\vec{x}_0\}$. There is a vector $\vec{a}$ and a number b so that

$$H = \{\vec{v} : \vec{a} \cdot \vec{v} = b\}$$

After multiplying $\vec{a}$ and b by (-1), if necessary, we find that

$$\begin{aligned} \vec{a} \cdot \vec{c} &\geq b \text{ for all } \vec{c} \in C^o \\ a \cdot \vec{x}_0 &\leq b \end{aligned}$$

Hence C^o is contained in one of the half-spaces determined by H. We have to show that C is also contained in the same half-space. If $\vec{x} \in C$ is arbitrary, then there is a sequence $\{\vec{x}_n\} \subseteq C^o$ so that $\lim_{n\to\infty} \vec{x}_n = \vec{x}$. It follows that $\vec{a}\cdot\vec{x} = \lim_{n\to\infty} \vec{a}\cdot\vec{x}_n \geq b$. Hence C is contained in one of the half-spaces determined by H. Moreover, since $\vec{a}\cdot\vec{x}_0 \leq b$ and since $\vec{x}_0 \in C$ also implies that $\vec{a}\cdot\vec{x}_0 \geq b$, it follows that $\vec{a}\cdot\vec{x}_0 = b$, i.e. $\vec{x}_0 \in H$. Hence $\vec{x}_0 \in H \cap C$. □

DEFINITION 20. *If C is a convex set, and if $F \subseteq C$ is a subset of C such that the following two conditions are satisfied:*

(1) *F is convex*
(2) *If $\vec{a}, \vec{b} \in C$ and $0 < \lambda < 1$ are given so that $\lambda\vec{a} + (1-\lambda)\vec{b} \in F$, then $\vec{a}, \vec{b} \in F$.*

then F is called a face of C.

If $F = \{\vec{p}\}$, i.e. if F consists of only one point, then $\vec{p}$ is called an extreme point of $\vec{p}$.

EXAMPLE 39. *Find all faces and all extreme points of the following convex sets:*

(1) *The interval* $[0,1]$.
(2) *The unit disk* $\left\{[x,y] : x^2 + y^2 \leq 1\right\}$
(3) *The square* $\{[x,y] : -1 \leq x, y \leq 1\}$
(4) *The simplices* Δ_n *for* $n = 1, 2, 3, 4$.

THEOREM 32. *If F is a face of C, and if G is a face of F, then G is a face of C.*

PROOF. Assume that vectors $\vec{a}, \vec{b} \in C$ and number $0 < \lambda < 1$ are given so that $\lambda\vec{a} + (1-\lambda)\vec{b} \in G$. We have to show that $\vec{a}, \vec{b} \in G$. First, $G \subseteq F$ implies that $\lambda\vec{a} + (1-\lambda)\vec{b} \in F$. Since F is a face of C, we conclude that $\vec{a}, \vec{b} \in F$. Now we use the property that G is a face of F to conclude that $\vec{a}, \vec{b} \in G$. □

COROLLARY 3. *If F is a face of C and if $\vec{p}$ is an extreme point of F, then $\vec{p}$ is an extreme point of C.*

PROOF. If $\vec{p}$ is an extreme point of F, then $\{\vec{p}\}$ is a face of F. Hence, by the last proposition, $\{\vec{p}\}$ is a face of C, i.e. $\vec{p}$ is an extreme point of C. □

THEOREM 33. *If C is a convex set, and if H is a supporting hyperplane, then $F_H = H \cap C$is a face of C.*

PROOF. Since both C and H are convex, the set $F_H = H \cap C$ is also convex. Pick a vector $\vec{n}$ perpendicular to H and a number c so that

$$\begin{aligned} H &= \{\vec{x} : \vec{n}\cdot\vec{x} = c\} \\ C &= \{\vec{x} : \vec{n}\cdot\vec{x} \leq c\} \end{aligned}$$

Assume that vectors $\vec{a}, \vec{b} \in C$ and a number $0 < \lambda < 1$ are given so that $\lambda\vec{a} + (1-\lambda)\vec{b} \in F_H$. We have to show that $\vec{a}, \vec{b} \in F_H$. First, $\lambda\vec{a} + (1-\lambda)\vec{b} \in H$ implies that

$$\begin{aligned} \vec{a} \cdot \vec{n} &\leq c \\ \vec{b} \cdot \vec{n} &\leq c \\ \left[\lambda\vec{a} + (1-\lambda)\vec{b}\right] \cdot \vec{n} &= c \end{aligned}$$

If $\vec{a} \cdot \vec{n}$ would be less than c, then we would arrive at the contradiction

$$\begin{aligned} c &= \left[\lambda\vec{a} + (1-\lambda)\vec{b}\right] \cdot \vec{n} \\ &= \lambda\left(\vec{a} \cdot \vec{n}\right) + (1-\lambda)\left(\vec{b} \cdot \vec{n}\right) \\ &< \lambda c + (1-\lambda)\left(\vec{b} \cdot \vec{n}\right) \\ &\leq \lambda c + (1-\lambda)c \\ &= c \end{aligned}$$

Hence $\vec{a} \cdot \vec{n} = c$ and similarly $\vec{b} \cdot \vec{n} = c$. We conclude that $\vec{a}, \vec{b} \in C \cap H = F_H$. □

THEOREM 34 (Krein-Milman). *If $C \subseteq \Re^n$ is a closed and bounded convex set, and if $\vec{x} \in C$, then there are extreme points $\vec{p}_1, \ldots, \vec{p}_k \in C$ and positive numbers $\lambda_1, \ldots, \lambda_k$ with $\lambda_1 + \cdots + \lambda_k = 1$ so that $\vec{x} = \lambda_1\vec{p}_1 + \cdots + \lambda_k\vec{p}_k$.*("Every point in a convex set C is a convex combination of extreme points of C.")

PROOF. We prove the theorem by induction on the dimension n of $\Re^n$. If $n = 1$, then C is a closed interval $[a, b]$. The extreme points of $[a, b]$ are the endpoints a and b, and every $x \in [a, b]$ can be written down as a convex combination of a and b. Indeed, $x = \frac{b-x}{b-a}a + \left(1 - \frac{b-x}{b-a}\right)b$, so we can pick $\lambda = \frac{b-x}{b-a}$.

Assume now that we have verified the theorem for all closed and bounded convex set $D \subseteq \Re^k$ with $k < n$, and that $C \subseteq \Re^n$ is a closed and bounded convex set. Let $\vec{x} \in C$. We define

$$\begin{aligned} m &= \min\{r : r\vec{x} \in C\} \\ M &= \max\{r : r\vec{x} \in C\} \end{aligned}$$

Since C is closed and bounded, we conclude that $m\vec{x}, M\vec{x} \in C$. By definition, since $\vec{x} \in C$, we find that $m \leq 1 \leq M$. We let

$$\lambda = \frac{M-1}{M-m}$$

Then

$$0 \leq \lambda \leq 1$$

and

$$\begin{aligned}
\lambda(m\vec{x}) + (1-\lambda)(M\vec{x}) &= (\lambda m + (1-\lambda)M)\vec{x} \\
&= \left(\frac{M-1}{M-m}m + \left(1 - \frac{M-1}{M-m}\right)M\right)\vec{x} \\
&= \frac{(M-1)m + ((M-m)-(M-1))M}{M-m}\vec{x} \\
&= \frac{(M-1)m + (1-m)M}{M-m}\vec{x} \\
&= \frac{Mm - m + M - mM}{M-m}\vec{x} \\
&= \vec{x}
\end{aligned}$$

Hence $\vec{x}$ is a convex combination of $m\vec{x}$ and $M\vec{x}$. We now have to show that $m\vec{x}$ and $M\vec{x}$ are convex combinations of extreme points, because then $\vec{x}$ would also be a convex combination of all the extreme points used in a convex combination of $m\vec{x}$ and $M\vec{x}$, respectively.

Note that by definition $m\vec{x}$ cannot belong to the interior of C, because otherwise $(m-\varepsilon)\vec{x}$ would also belong to C for some $\varepsilon > 0$, contradicting the fact that m is the minimal real number r so that $r\vec{x} \in C$. Hence $m\vec{x}$ belongs to the boundary of C. Similarly, $M\vec{x}$ belongs also to ∂C.

It remains to show that every vector $\vec{v} \in \partial C$ is a convex combination of extreme points. Pick a hyperplane H that supports C at $\vec{v}$. Then $F = C \cap H$ is a face of C, hence every extreme point of F is also an extreme point of C.

We now show that $\vec{v}$ is a convex combination of extreme points of F. Since F is contained in the affine hyperplane H, the convex set $F - \vec{v}$ is contained in the hyperplane $H - \vec{v}$. Since $H - \vec{v}$ contains $0 = \vec{v} - \vec{v}$, $H - \vec{v}$ is a linear subspace, hence $H - \vec{v} \equiv \Re^{n-1}$. Therefore, $F - \vec{v}$ can be thought of as a subset of an $n-1$ dimensional space. We therefore may assume that$F - \vec{v} \subseteq \Re^{n-1}$, i.e.

$$\vec{v} \in F \subseteq \Re^{n-1}$$

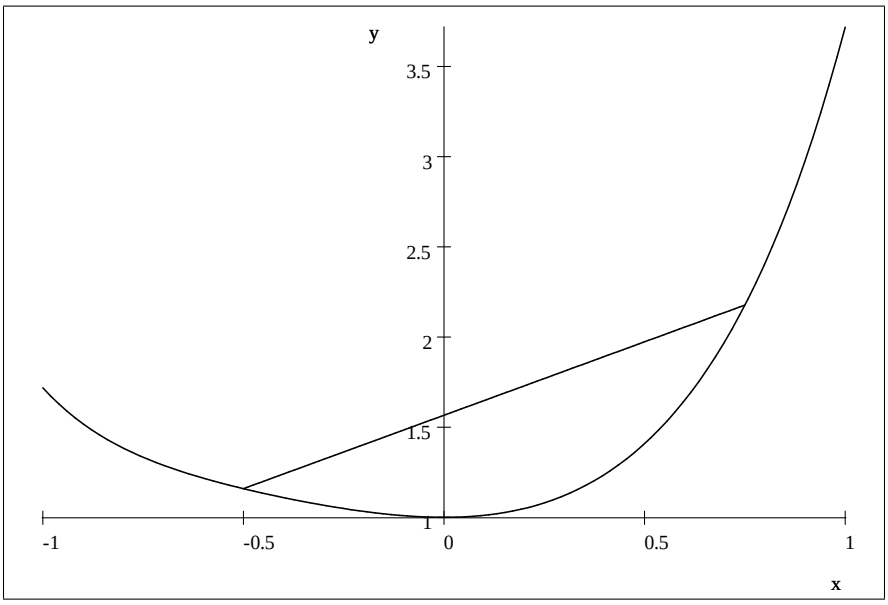
By induction hypothesis, $\vec{v}$ is a linear combination of extreme points of F, and this completes the proof. □

COROLLARY 4. *Let C be a closed and bounded convex subset of $\Re^n$. Then C contains at least one extreme point.*

16. Extreme Values of Convex Functions

DEFINITION 21. *Let $C \subseteq \Re^n$ be a convex set. A continuous real valued function $f : C \to \Re$ is called convex, if $\vec{x}, \vec{y} \in C$ and $0 \leq \lambda \leq 1$ imply $f(\lambda\vec{x} + (1-\lambda)\vec{y}) \leq \lambda f(\vec{x}) + (1-\lambda) f(\vec{y})$.*

In a graphical representation of f, convex functions are characterized by the property that the secant line between two points is above the graph of the function:



PROPOSITION 13. *The function $\vec{x} \to \|\vec{x}\| : \Re^n \to \Re$ is convex.*

PROOF. This statement follows immediately from the triangle inequality: If $\vec{x}, \vec{y} \in \Re^n$ and if $0 \leq \lambda \leq 1$, then

$$\begin{aligned} \|\lambda\vec{x} + (1-\lambda)\vec{y}\| &\leq \|\lambda\vec{x}\| + \|(1-\lambda)\vec{y}\| \\ &= |\lambda| \|\vec{x}\| + |1-\lambda| \|\vec{y}\| \\ &= \lambda \|\vec{x}\| + (1-\lambda) \|\vec{y}\| \end{aligned}$$

□

Convex function were already discussed in a course on Calculus, maybe under a different name and with a different definition. They were the functions $f : [a, b] \to \Re$ so that $f'(x)$ is increasing, or, equivalently, $f''(x) \geq 0$. We first need to relate those two definitions. The next example illustrates that in general our definition will be more general than the one given in Calculus using derivatives:

EXAMPLE 40. *Not every convex function is differentiable. For instance, $f(x) = |x|$ is convex by the last proposition (take $n = 1$), but this function is not differentiable.*

We show that differentiable convex functions have increasing first derivatives. Since the term "increasing" is used differently by different people, we better repeat the definition:

DEFINITION 22. *Let $f:[a,b]\to\Re$ be a function. If $a\leq x<y\leq b$ implies that $f(x)\leq f(y)$, then f is called an increasing function. If $a\leq x<y\leq b$ implies that $f(x)<f(y)$, then f is called strictly increasing. The notions of decreasing functions and strictly decreasing functions are defined accordingly.*

In particular, constant functions are increasing and decreasing. However, they are not strictly increasing and not strictly decreasing.

PROPOSITION 14. *If $f:[a,b]\to\Re$ is a differentiable function, and if $f'(x)$ is decreasing, then the minimum of f is obtained at one of the endpoints. Moreover, the critical points form an interval, and the function f has a global maximum at every critical point.*

PROOF. The critical points are the solutions of $f'(x)=0$. If x_0 and x_2 are critical points, and if $x_0\leq x_c\leq x_1$, then $0=f(x_0)\leq f(x_c)\leq f(x_1)=0$, hence $f(x_c)$ is also a critical point. Therefore the critical points form an interval.

The Fundamental Theorem of Calculus implies that

$$f(x)=f(a)+\int_a^x f'(t)\,dt$$

Assume that c is a critical point of f. Then $f'(c)=0$. Since f' is decreasing, $f'(t)\geq 0$ for $t\leq c$ and $f'(t)\leq 0$ for $t\geq c$. Hence $x\leq c$ implies that

$$\begin{aligned} f(x) &= f(a)+\int_a^x f'(t)\,dt \\ &\leq f(a)+\int_a^x f'(t)\,dt+\int_x^c f'(t)\,dt \\ &= f(a)+\int_a^c f'(t)\,dt=f(c) \end{aligned}$$

and $x\geq c$ implies that

$$\begin{aligned} f(x) &= f(a)+\int_a^x f'(t)\,dt \\ &= f(a)+\int_a^c f'(t)\,dt+\int_c^x f'(t)\,dt \\ &\leq f(c) \end{aligned}$$

It follows that f has a global maximum at c.

Since f has a global maximum at every critical point, the minimum has to be obtained at one of the endpoints. □

PROPOSITION 15. *If f is a convex function, defined on an interval $[a,b]$, then $a\leq u<v<w\leq b$ implies*

$$\frac{f(v)-f(u)}{v-u}\leq\frac{f(w)-f(u)}{w-u}\leq\frac{f(w)-f(v)}{w-v}$$

PROOF. We let

$$\lambda=\frac{w-v}{w-u}$$

Then

$$\begin{aligned} 1-\lambda &= 1-\frac{w-v}{w-u} \\ &= \frac{(w-u)-(w-v)}{w-u} \\ &= \frac{v-u}{w-u} \end{aligned}$$

and

$$\begin{aligned} \lambda u+(1-\lambda)w &= \frac{w-v}{w-u}u+\frac{v-u}{w-u}w \\ &= \frac{wu-vu+vw-uw}{w-u} \\ &= \frac{v(w-u)}{w-u} \\ &= v \end{aligned}$$

Since f is convex, we conclude that

$$\begin{aligned} f(v) &= f(\lambda u+(1-\lambda)w) \\ &\leq \lambda f(u)+(1-\lambda)f(w) \\ &= \lambda(f(u)-f(w))+f(w) \end{aligned}$$

It follows that

$$\begin{aligned} f(v)-f(w) &\leq \frac{w-v}{w-u}(f(u)-f(w)) \\ f(w)-f(v) &\geq (w-v)\frac{f(w)-f(u)}{w-u} \end{aligned}$$

and hence

$$\frac{f(w)-f(u)}{w-u} \leq \frac{f(w)-f(v)}{w-v}$$

Similarly, we compute from $f(v) \leq \lambda f(u)+(1-\lambda)f(w)$ that

$$\begin{aligned} f(v) &\leq \lambda f(u)+(1-\lambda)f(w) \\ &= (\lambda-1)f(u)+(1-\lambda)f(w)+f(u) \\ f(v)-f(u) &\leq (1-\lambda)(f(w)-f(u)) \\ &= \frac{v-u}{w-u}(f(w)-f(u)) \end{aligned}$$

and therefore

$$\frac{f(v)-f(u)}{v-u} \leq \frac{f(w)-f(u)}{w-u}$$

□

PROPOSITION 16. *If f is a convex function, defined on an interval $[a,b]$, if $a \leq x_0 < x_1 \leq b$, and if $0 < h < x_1 - x_0$, then*

$$\frac{f(x_0+h)-f(x_0)}{h} \leq \frac{f(x_1)-f(x_1-h)}{h}$$

PROOF. We let

$$\begin{aligned} u &= x_0 \\ v &= x_0 + h \\ w &= x_1 \end{aligned}$$

Then $v = x_0 + h < x_0 + (x_1 - x_0) = w$ and hence the last proposition implies that

$$\begin{aligned} \frac{f(x_0 + h) - f(x_0)}{h} &= \frac{f(v) - f(u)}{v - u} \\ &\leq \frac{f(w) - f(u)}{w - u} \end{aligned}$$

Similarly, if we let

$$\begin{aligned} u &= x_0 \\ v &= x_1 - h \\ w &= x_1 \end{aligned}$$

then $h < x_1 - x_0$ implies $v = x_1 - h > x_1 - (x_1 - x_0) = x_0$ and therefore

$$\begin{aligned} \frac{f(w) - f(u)}{w - u} &\leq \frac{f(w) - f(v)}{w - v} \\ &= \frac{f(x_1) - f(x_1 - h)}{x_1 - (x_1 - h)} \\ &= \frac{f(x_1) - f(x_1 - h)}{h} \end{aligned}$$

Both inequalities together yield the statement of the proposition. □

PROPOSITION 17. *If $f : [a, b] \to \Re$ is a differentiable convex function, then $f'(x)$ is an increasing function. Especially, if the second derivative exists, then $f''(x) \geq 0$.*

PROOF. Assume that $x_0 < x_1$. Then we apply proposition and obtain

$$\begin{aligned} f'(x_0) &= \lim_{h \to 0^+} \frac{f(x_0 + h) - f(x_0)}{h} \\ &\leq \lim_{h \to 0^+} \frac{f(x_1) - f(x_1 - h)}{h} \\ &= f'(x_1) \end{aligned}$$

Hence $f'(x)$ is an increasing function. □

We now show that the converse is also true.

PROPOSITION 18. *If $f : [a, b] \to \Re$ is a differentiable function, and if $f'(x)$ is increasing, then f is convex. This is especially the case if $f''(x) \geq 0$ for all $x \in [a, b]$.*

PROOF. Let $x, y \in [a, b]$ be given, and let $0 \leq \lambda \leq 1$. We have to prove that

$$f(\lambda x + (1 - \lambda) y) \leq \lambda f(x) + (1 - \lambda) f(y)$$

or, equivalently, that

$$0 \leq \lambda f(x) + (1 - \lambda) f(y) - f(\lambda x + (1 - \lambda) y)$$

After swapping x and y, if necessary, we may assume without loss of generality that $x < y$. For each $\lambda \in [0,1]$ let

$$g(\lambda) = \lambda f(x) + (1-\lambda) f(y) - f(\lambda x + (1-\lambda) y)$$

We have to show that $g(\lambda) \geq 0$ for all values of λ. Since $f'(x)$ is increasing

$$\begin{aligned} g'(\lambda) &= f(x) - f(y) - f'(\lambda x + (1-\lambda) y)(x-y) \\ &= f(x) - f(y) + (y-x) f'(\lambda x + (1-\lambda) y) \end{aligned}$$

As λ increases from 0 to 1, the value of $\lambda x + (1 - \lambda y)$ decreases from y to x. Since $y > x$, the function values $(y-x) f'(\lambda x + (1-\lambda) y)$ also decrease. It follows that $g'(\lambda)$ is a decreasing function. We conclude that g obtains its minimum at the one of the endpoints. Since $g(0) = g(1) = 0$, the minimum of g is 0. This concludes the proof. □

THEOREM 35. *Let $f : [a,b] \to \Re$ be a differentiable function. Then f is convex if and only if $f'(x)$ is an increasing function. If the second derivative also exists, then this is the case if and only if $f''(x) \geq 0$ for all $x \in [a,b]$.*

THEOREM 36. *If $f : [a,b] \to \Re$ is a convex function, then f obtains its maximum at one of the endpoint. Moreover, every local minimum is a global minimum, and the points where the global minimum is obtained form an interval.*

PROOF. This would follow from the previous propositions if f would be differentiable. However, this is not always the case.

First, let $M = \max\{f(a), f(b)\}$. We have to show that M is the maximum of f. Let $x \in [a,b]$. Then there is a number $0 \leq \lambda \leq 1$ so that $x = \lambda a + (1-\lambda) b$. We conclude that

$$\begin{aligned} f(x) &= f(\lambda a + (1-\lambda) b) \\ &\leq \lambda f(a) + (1-\lambda) f(b) \\ &\leq \lambda M + (1-\lambda) M \\ &= M \end{aligned}$$

Hence M is indeed the maximum.

Now let m be the minimum of f, and assume that m is obtained at c, i.e.

$$m = f(c)$$

Assume that f has a local minimum at d. We show that the assumption $f(d) > m$ would lead to a contradiction, hereby verifying that f has indeed assumes the global minimum at d. For values x of the form $x = \lambda c + (1-\lambda) d$ with $0 < \lambda \leq 1$ we find that

$$\begin{aligned} f(x) &\leq \lambda f(c) + (1-\lambda) f(d) \\ &= \lambda m + (1-\lambda) f(d) \\ &< \lambda f(d) + (1-\lambda) f(d) = f(d) \end{aligned}$$

If λ approaches 0, the values of x approach d. Hence f has function values less that $f(d)$ close to d, and therefore f cannot have local minimum at d. It follows that $f(d) = m$.

Finally, let c and d be two points where the global minimum is obtained, and let x be a point between c and d. We have to show that $f(x) = m$. Pick a number $0 \leq \lambda \leq 1$ so that $x = \lambda c + (1-\lambda) d$. Then

$$\begin{aligned} m &\leq f(x) \\ &= f(\lambda c + (1-\lambda) d) \\ &\leq \lambda f(c) + (1-\lambda) f(d) \\ &= \lambda m + (1-\lambda) m = m \end{aligned}$$

We find that also $f(x) = m$ □

We now would like to generalize the last two theorem to higher dimensions.

PROPOSITION 19. *Let $C \subseteq \Re^n$ be a convex set, and let $f : C \to \Re$ be a real-valued function. Then f is convex, if and only if for each pair of points $\vec{a}, \vec{b} \in C$ the function $g : [0,1] \to \Re$ defined by*

$$g(t) = f\left(\vec{a} + t\left(\vec{b} - \vec{a}\right)\right)$$

is convex.

PROOF. First, assume that f is convex, and that $s, t \in [0,1]$ are given. Then for each $0 \leq \lambda \leq 1$ we have

$$\begin{aligned} g(\lambda s + (1-\lambda) t) &= f\left(\vec{a} + (\lambda s + (1-\lambda) t)\left(\vec{b} - \vec{a}\right)\right) \\ &= f\left(\lambda\left(\vec{a} + s\left(\vec{b} - \vec{a}\right)\right) + (1-\lambda)\left(\vec{a} + t\left(\vec{b} - \vec{a}\right)\right)\right) \\ &\leq \lambda f\left(\left(\vec{a} + s\left(\vec{b} - \vec{a}\right)\right)\right) + (1-\lambda) f\left(\vec{a} + t\left(\vec{b} - \vec{a}\right)\right) \\ &= \lambda g(s) + (1-\lambda) g(t) \end{aligned}$$

Hence g is convex for each choice of $\vec{a}, \vec{b} \in C$.

Conversely, assume that g is convex for each choice of $\vec{a}, \vec{b} \in C$. Fix vectors $\vec{x}, \vec{y} \in C$ and a number $0 \leq \lambda \leq 1$. We have to show that $f(\lambda \vec{x} + (1-\lambda)\vec{y}) \leq \lambda f(\vec{x}) + (1-\lambda) f(\vec{y})$. Pick $\vec{a} = \vec{y}$ and $\vec{b} = \vec{x}$. Then the function $g(t) = f\left(\vec{a} + t\left(\vec{b} - \vec{a}\right)\right) = f(\vec{y} + \lambda(\vec{x} - \vec{y}))$ is convex. We now compute

$$\begin{aligned} f(\lambda \vec{x} + (1-\lambda)\vec{y}) &= f(\vec{y} + \lambda(\vec{x} - \vec{y})) \\ &= g(\lambda) \\ &= g(\lambda \cdot 1 + (1-\lambda) \cdot 0) \\ &\leq \lambda g(1) + (1-\lambda) g(0) \\ &= \lambda f(\vec{y} + 1 \cdot (\vec{x} - \vec{y})) + (1-\lambda) f(\vec{y} + 0 \cdot (\vec{x} - \vec{y})) \\ &= \lambda f(\vec{x}) + (1-\lambda) f(\vec{y}) \end{aligned}$$

It follows that f is a convex function. □

Recall that a symmetric square matrix M is positive semi-definite, if $\vec{x}^T M \vec{x} \geq 0$ for all choices of vectors $\vec{x}$. A matrix M is positive semi-definite, if and only if all eigenvalues of M are non-negative.

THEOREM 37. *Let $C \subseteq \Re^n$ is an open convex set and let $f : C \to \Re$ a twice differentiable function. Then f is convex if and only if the Hessian matrix $H(f)(\vec{x})$ is positive semi-definite for all values of $\vec{x} \in C$.*

PROOF. We would like to use the last proposition. For given values $\vec{a}, \vec{b} \in C$ we have to investigate the function

$$g(t) = f\left(\vec{a} + t\left(\vec{b} - \vec{a}\right)\right), 0 \leq t \leq 1$$

This function will be positive definite if and only if $g''(t) \geq 0$ for all values of t. Using the chain rule, we compute:

$$\begin{aligned} g'(t) &= \nabla f\left(\vec{a} + t\left(\vec{b} - \vec{a}\right)\right) \cdot \left(\vec{b} - \vec{a}\right) \\ &= \left(\vec{b} - \vec{a}\right)^T \begin{bmatrix} \frac{\partial f}{\partial x_1}\left(\vec{a} + t\left(\vec{b} - \vec{a}\right)\right) \\ \vdots \\ \frac{\partial f}{\partial x_n}\left(\vec{a} + t\left(\vec{b} - \vec{a}\right)\right) \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned} g''(t) &= \left(\vec{b} - \vec{a}\right)^T \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}\left(\vec{a} + t\left(\vec{b} - \vec{a}\right)\right) & \frac{\partial^2 f}{\partial x_1 \partial x_n}\left(\vec{a} + t\left(\vec{b} - \vec{a}\right)\right) \\ \frac{\partial^2 f}{\partial x_n \partial x_1}\left(\vec{a} + t\left(\vec{b} - \vec{a}\right)\right) & \frac{\partial^2 f}{\partial x_n \partial x_n}\left(\vec{a} + t\left(\vec{b} - \vec{a}\right)\right) \end{bmatrix} \left(\vec{b} - \vec{a}\right) \\ &= \left(\vec{b} - \vec{a}\right)^T H(f)\left(\vec{a} + t\left(\vec{b} - \vec{a}\right)\right)\left(\vec{b} - \vec{a}\right) \end{aligned}$$

Hence $g''(t)$ will be non-negative for all values of $\vec{a}, \vec{b} \in C$ and t if and only if $\left(\vec{b} - \vec{a}\right)^T H(f)\left(\vec{a} + t\left(\vec{b} - \vec{a}\right)\right)\left(\vec{b} - \vec{a}\right) \geq 0$ for all such choices. This is clearly the case if $H(f)\left(\vec{a} + t\left(\vec{b} - \vec{a}\right)\right)$ is always positive semi-definite. Conversely, if $\left(\vec{b} - \vec{a}\right)^T H(f)\left(\vec{a} + t\left(\vec{b} - \vec{a}\right)\right)\left(\vec{b} - \vec{a}\right)$ is non-negative for all choices of $\vec{a}, \vec{b} \in C$, then, letting $t = 0$, we find that $\left(\vec{b} - \vec{a}\right)^T H(f)(\vec{a})\left(\vec{b} - \vec{a}\right)$ is always non-negative. If $\vec{x}$ is arbitrary, then since C is open, we can pick $\varepsilon > 0$ so that $\vec{b} = \vec{a} + \varepsilon\vec{x} \in C$. Hence $\vec{b} - \vec{a} = \varepsilon\vec{x}$, and therefore $0 \leq (\varepsilon\vec{x})^T H(f)(\vec{a})(\varepsilon\vec{x}) = \varepsilon^2\left(\vec{x}^T H(f)(a)\vec{x}\right)$. It follows that $H(f)(\vec{a})$ is positive semi-definite for each $\vec{a} \in C$.

We conclude that $g''(t) \geq 0$ for all values of $\vec{a}, \vec{b} \in C$ if and only if $H(f)(\vec{a})$ is positive semi-definite for all values of $\vec{a} \in C$. This proves the theorem. □

EXAMPLE 41. *The function* $f(x, y, z) = x^2 + y^2 + z^2 + xy + xz + yz - 7x - 8y - 9z$ *is convex.*

To verify this statement, we have to form the Hessian matrix of f :

$$H(f) = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

The eigenvalues of $H(f)$ are 1 (twice) and 4. Hence all eigenvalues are positive and therefore $H(f)$ is positive definite. It follows that f is convex.

PROPOSITION 20. *The matrix*

$$A = \begin{bmatrix} a & c \\ c & b \end{bmatrix}$$

is positive semi-definite if and only if $ab \geq c^2$ *and* $a + b \geq 0$.

PROOF. Let λ_1 and λ_2 be the two eigenvalues of A. Then

$$\begin{aligned}\det\begin{bmatrix} a-\lambda & c \\ c & b-\lambda \end{bmatrix} &= \lambda^2-(a+b)\lambda+ab-c^2 \\ &= (\lambda-\lambda_1)(\lambda-\lambda_2)\end{aligned}$$

and hence

$$\begin{aligned}ab-c^2 &= \lambda_1\lambda_2 \\ a+b &= \lambda_1+\lambda_2\end{aligned}$$

If both eigenvalues are non-negative, then $ab-c^2\geq 0$ and $a+b\geq 0$. Conversely, assume $ab-c^2\geq 0$ and $a+b\geq 0$. Since $0\leq ab-c^2=\lambda_1\lambda_2$, the eigenvalues λ_1 and λ_2 have to have the same sign. The inequality $\lambda_1+\lambda_2=a+b\geq 0$ implies that they are both have to be positive. □

EXAMPLE 42. *Find a convex set on which $f(x,y)=x^2+y^2-x^3-xy$, a convex function.*

Again, we find the Hessian matrix:

$$H(f)=\begin{bmatrix} 2-6x & -1 \\ -1 & 2 \end{bmatrix}$$

This matrix is positive definite if and only if $\det M=2(2-6x)-1\geq 0$, and $tr(M)=(2-6x)+2\geq 0$. This leads to the equations $x\leq\frac{1}{4}$ and $x\leq\frac{2}{3}$. Hence the largest convex set for which the function is convex is $\left\{[x,y]:x\leq\frac{1}{4}\right\}$.

THEOREM 38. *Let $C\subseteq\Re^n$ be a bounded, closed convex set and let $f:C\to\Re$ a continuous real-valued function. Then the maximum of f is obtained at an extreme point of f.*

PROOF. Let M be the maximum of f, and let $\vec{x}\in C$ be a point so that $f(\vec{x})=M$. Then there are extreme points $\vec{p}_1,\ldots,\vec{p}_n$ and positive number $\lambda_1,\ldots,\lambda_n$ with $\sum_{i=1}^n\lambda_i=1$ so that $\vec{x}=\sum_{i=1}^n\lambda_i\vec{p}_i$. We show that $f(\vec{p}_i)=M$ holds for at least one index i. Assume not. Then, since M is the maximum of f on C, we conclude that $f(\vec{p}_i)<M$ for all indices i. It follows that

$$\begin{aligned}M &= f(\vec{x}) \\ &= f\left(\sum_{i=1}^n\lambda_i\vec{p}_i\right) \\ &\leq \sum_{i=1}^n\lambda_i f(\vec{p}_i) \\ &< \sum_{i=1}^n(\lambda_i M) \\ &= \left(\sum_{i=1}^n\lambda_i\right)M \\ &= M\end{aligned}$$

a contradiction. □

THEOREM 39. *Let $C \subseteq \Re^n$ be an open convex set, and let $f : C \to \Re$ be a continuous convex function. Then every local minimum of f is a global minimum, and the points where the global minimum is obtained (if there is one) is a convex set.*

PROOF. Assume that f has a local minimum at $\vec{d} \in C$. We would like to show that f has a global minimum at $\vec{d}$. We show that the assumption $f\left(\vec{d}\right) > f(\vec{c})$ for some $\vec{c} \in C$ would lead to a contradiction, hereby verifying that f has indeed assumes the global minimum at $\vec{d}$. For values $\vec{x}$ of the form $\vec{x} = \lambda c + (1-\lambda) d$ with $0 < \lambda \leq 1$ we find that

$$\begin{aligned} f(\vec{x}) &\leq \lambda f(\vec{c}) + (1-\lambda) f\left(\vec{d}\right) \\ &< \lambda f\left(\vec{d}\right) + (1-\lambda) f\left(\vec{d}\right) = f\left(\vec{d}\right) \end{aligned}$$

If λ approaches 0, the values of $\vec{x}$ approach $\vec{d}$. Hence f has function values less that $f\left(\vec{d}\right)$ close to $\vec{d}$, and therefore f cannot have local minimum at $\vec{d}$, a contraction.

Now let $\vec{c}$ and $\vec{d}$ be two points where the global minimum is obtained, and let $0 \leq \lambda \leq 1$. If $\vec{x} = \lambda\vec{c} + (1-\lambda)\vec{d}$. then

$$\begin{aligned} m &\leq f(\vec{x}) \\ &= f\left(\lambda\vec{c} + (1-\lambda)\vec{d}\right) \\ &\leq \lambda f(\vec{c}) + (1-\lambda) f\left(\vec{d}\right) \\ &= \lambda m + (1-\lambda) m = m \end{aligned}$$

We find that also $f(\vec{x}) = m$ Hence the set $\{\vec{x} : f(\vec{x}) = m\}$ is indeed a convex set. □

EXAMPLE 43. *Let $C = \left\{[x,y] : |x| + |y| \leq \frac{\pi}{2}\right\}$ and let $f(x,y) = \cos x + \cos y$. Show that f is convex on C and find the global maximum and the global minimum of f.*

The Hessian matrix of f is given by

$$H(f) = \begin{bmatrix} -\cos x & 0 \\ 0 & -\cos y \end{bmatrix}$$

This matrix is not positive semi-definite, because both eigenvalues are negative. But the Hessian matrix of $-f$ is positive semi-definite:

$$H(-f) = \begin{bmatrix} \cos x & 0 \\ 0 & \cos y \end{bmatrix}$$

The eigenvalues are $\cos x$ and $\cos y$, and both are non-negative for $|x| \leq \frac{\pi}{2}$ and $|y| \leq \frac{\pi}{2}$. We proceed by finding the global maximum and the global minimum for $-f$. The global maximum of $-f$ is obtained at an extreme point of C. The extreme points are $\left[\pm\frac{\pi}{2}, 0\right]$ and $\left[0, \pm\frac{\pi}{2}\right]$. Since $-f\left(\pm\frac{\pi}{2}, 0\right) = -\cos\left(\frac{\pi}{2}\right) - \cos(0) = -1$, and $-f\left(0, \pm\frac{\pi}{2}\right) = -1$, the maximum of $-f$ is -1.

The function $-f$ has only one critical point in the interior of C :

$$\nabla(-f) = [\sin x, \sin y]$$

The solution of $\sin x = 0$ and $\sin y = 0$ on C^o is $[x, y] = [0, 0]$. Hence $-f$ has a local minimum at $[0, 0]$. This value then also has to be the global minimum. So the global minimum of $-f$ on the open set C^o is -2. Since $-f$ is continuous, this is also the global minimum of $-f$ on the closed convex set C.

It follows that f has a global maximum at $[0, 0]$, and the global maximum is 2. The global minimum of f is obtained at each of the points $[\pm 1, 0]$ and $[0, \pm 1]$ and has the value 1.

REMARK 1. *If a convex function has no critical point in the interior of its domain, then the global minimum will be obtained on the boundary, and we are back to Lagrange multipliers!*

CHAPTER 2

Linear Programming

1. Elementary Examples of Linear Programming Problems

We now consider a special case of regional constraints. In the problem

$$\begin{array}{rcl} & & \text{Maximize (minimize) } f(\vec{x}) \\ \text{subject to } g_1(\vec{x}) & \leq & 0 \\ g_2(\vec{x}) & \leq & 0 \\ & & \vdots \\ g_n(\vec{x}) & \leq & 0 \end{array}$$

we will assume that the function $f(\vec{x})$ is a linear function, and that $g_1(\vec{x}), \ldots, g_n(\vec{x})$ are affine functions. Then each of the sets $\{\vec{x} : g_i(\vec{x}) \leq 0\}$ is a closed half space, and therefore the feasible set is a closed polytope, i.e. a closed convex set. Since every linear function is a convex function, we know in principle how to solve the problem: Find all extreme points of the polytope - the maximum (minimum) will be obtained at one of those. Of course, the last argument is missing one important step. Which one?

Before we discuss the theory any further, we will give some examples:

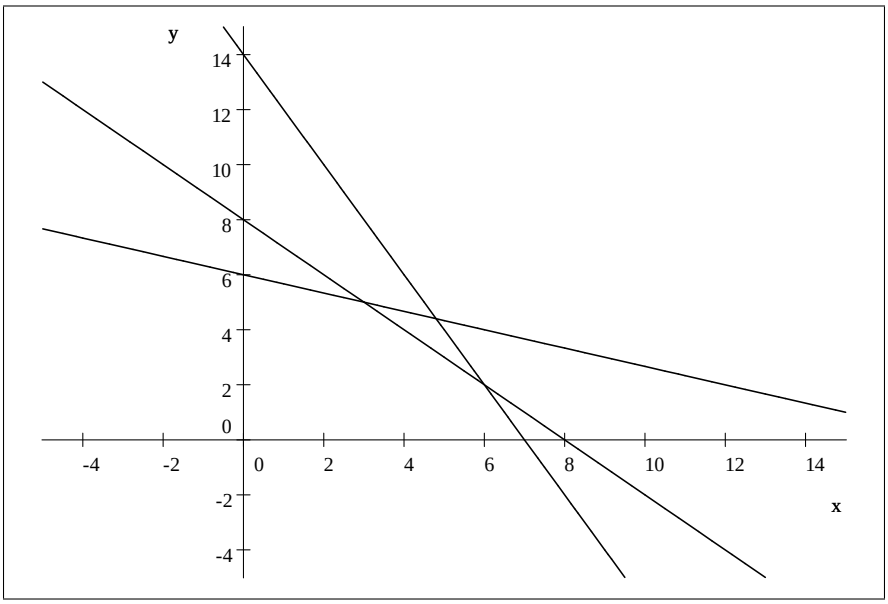
EXAMPLE 44. *Find the minimum and maximum of*

$$f(x, y) = 4x + 5y$$

subject to the constraints

$$\begin{array}{rcl} x + y & \leq & 8 \\ x + 3y & \leq & 18 \\ 2x + y & \leq & 14 \end{array}$$

We plot the feasible region. It is the region bound by the three line $x + y = 8$, $x + 3y = 18$, and $2x + y = 14$ that also contains the origin:



The extreme points of the feasible set can be obtained as intersections of lines:

$$\begin{aligned} x+y &= 8 \\ x+3y &= 18 \\ (x,y) &= (3,5) \text{ satisfies the third inequality, part of region} \end{aligned}$$

$$\begin{aligned} x+y &= 8 \\ 2x+y &= 14 \\ (x,y) &= (6,2) \text{ satisfies the second inequality, part of region} \end{aligned}$$

$$\begin{aligned} x+3y &= 18 \\ 2x+y &= 14 \\ (x,y) &= \left(\frac{24}{5},\frac{22}{5}\right) \text{ does not satisfy the first inequality, not part of region.} \end{aligned}$$

We also have to consider the origin as well as the intersections of the $y-$axis and the x-axis with some of the lines. This leads to the extreme points $(3,5)$, $(6,2)$, $(7,0)$, $(0,6)$ and $(0,0)$. The maximum in the feasible region is 37, obtained at $(3,5)$.

EXAMPLE 45. *A brewer brews two different kinds of beer: Pilsner and Export. He makes a profit of* 100 *ducats for a barrel of Pilsner and* 150 *ducats for a barrel of Export. The Pilsner requires* 10 *stones of hops and* 15 *stones of malt per barrel, whereas the Export requires only* 5 *stones of hops, but* 17 *stones of malt. However, only* 1000 *stones of hops and* 1600 *stones of malt are available each year.*

(1) *How many barrels of each type should she brew so as to maximize profit.*
(2) *Repeat the problem, if she earns as much for the Pilsner as she does for the Export.*
(3) *What happens if a law passed by the local royalties requires her to sell only completely filled barrels?*

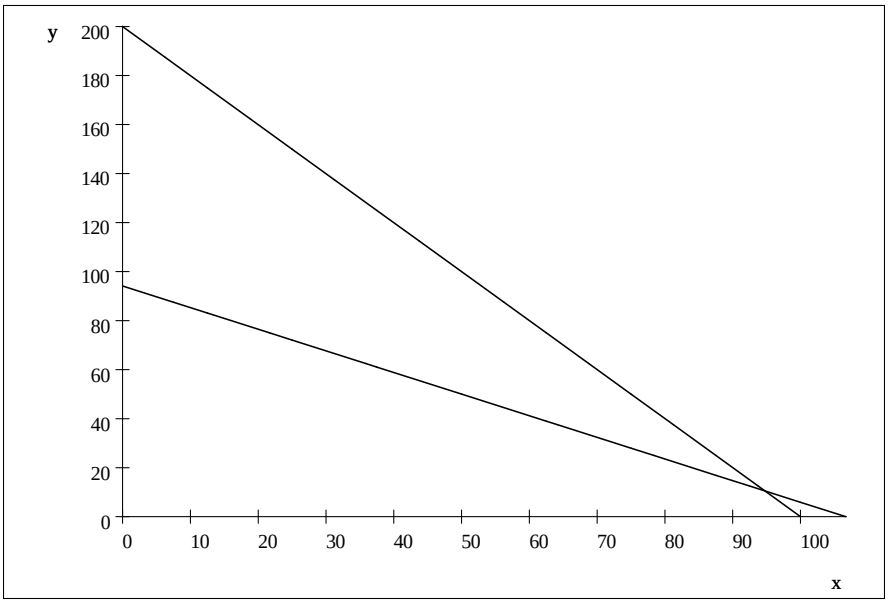
Let x be the number of barrels of Pilsner produced by the brewer, and let y be the number of barrels of Export. The profit can be computed as

$$f(x,y) = 100x + 150y$$

The constraints are

$$\begin{aligned} 10x + 5y &\leq 1000 \\ 15x + 17y &\leq 1600 \\ x &\geq 0 \\ y &\geq 0 \end{aligned}$$

We plot the feasible region:



The extreme points are $\left(0, \frac{1600}{17}\right)$, $(100, 0)$ and the solution of

$$\begin{aligned} 10x + 5y &= 1000 \\ 15x + 17y &= 1600 \end{aligned}$$

which is $(x,y) = \left(\frac{1800}{19}, \frac{200}{19}\right)$. Since $f(100,0) = 10\,000$, $f\left(0, \frac{1600}{17}\right) = \frac{240\,000}{17} = 14118.$ and $f\left(\frac{1800}{19}, \frac{200}{19}\right) = \frac{210\,000}{19} = 11053$, the maximum is obtained at $\left(0, \frac{1600}{17}\right)$, i.e. her profit is maximal if shew brews only Export.

If we change the profit function so that she earns as much on Pilsner as on Export, then we obtain a function of the form $f(x,y) = a(x+y)$. In this case, $f(100,0) = 100a$, $f\left(0, \frac{1600}{17}\right) = \frac{1600}{17}a = 94.118a$ and $f\left(\frac{1800}{19}, \frac{200}{19}\right) = \frac{2000}{19}a = 105.26a$, and she would try to make $\frac{1800}{19} = 94.737$ barrels of Pilsner and $\frac{200}{19} = 10.526$ barrels of Export.

If she has to sell only whole barrels, then this method cannot be applied - we have to use "Integer Programming".

EXAMPLE 46. *A car manufacturer produces three types of engines. An economy engine with a low gas mileage of* 45 *miles/gallon, a mid-sized engine with a gas mileage of* 25 *miles/gallon, and engines for SUVs with a gas mileage of* 12 *miles / gallon. The company makes a profit of* 125 *knusels for the economy engine,* 400

knusels for the mid-sized engine and 2500 *knusels for the SUV engines. They need one worker for an economy engine, two workers for a mid-sized engine and five workers for a SUV engine. They have a work force of* 1000 *workers, and they have no problems in letting some people working only part time. In order to deal with the gas prizes and the impact of cars on the environment, the state passes a law restricting the average mileage of all engines manufactured by a company to be above* 28 *miles/gallon. Assuming that the factory intends to follow the law, how many engines should they produce of each type?*

If x, y and z are the numbers of small, mid-sized and SUV engines, respectively, then the profit is given by

$$f(x, y, z) = 125x + 400y + 2500z$$

The constraints by the work force yields

$$x + 2y + 5z \leq 1000$$

and the state law would lead to

$$\frac{45x + 25y + 12z}{x + y + z} \geq 28$$

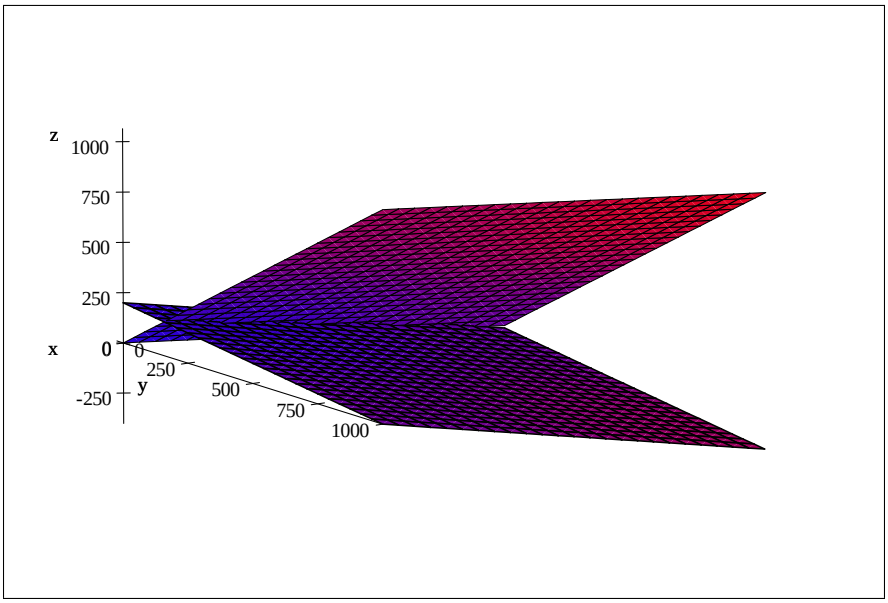
The second inequality can be written as

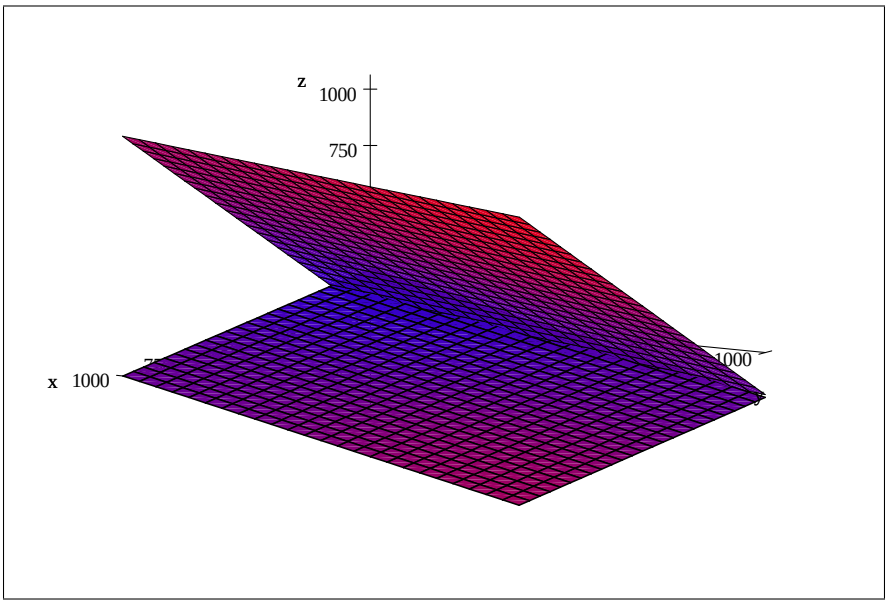
$$17x - 3y - 16z \geq 0$$

Of course, we also have

$$\begin{aligned} x &\geq 0 \\ y &\geq 0 \\ z &\geq 0 \end{aligned}$$

Plotting the two constraints in 3 dimensions yields the following graph:





There are four extreme points, namely $(0,0,0)$, $(1000,0,0)$, $\left(\frac{3000}{37}, \frac{17\,000}{37}, 0\right)$, and $\left(\frac{16\,000}{101}, 0, \frac{17\,000}{101}\right)$. (The last two solutions are obtained by solving the equations

$$\begin{aligned} x + 2y + 5z &= 1000 \\ 17x - 3y - 16z &= 0 \end{aligned}$$

simultaneous , which leads to $\left[x = \frac{17}{37}z + \frac{3000}{37}, y = \frac{17\,000}{37} - \frac{101}{37}z\right]$, and the observing that $x, y, z \geq 0$.) We have to find the maximum value of $f(x, yz,)$ on those four points:

$$\begin{aligned} f(0,0,0) &= 0 \\ f(1000,0,0) &= 1.25 \times 10^5 \\ f\left(\frac{3000}{37}, \frac{17\,000}{37}, 0\right) &= 1.939\,2 \times 10^5 \\ f\left(\frac{16\,000}{101}, 0, \frac{17\,000}{101}\right) &= 4.405\,9 \times 10^5 \end{aligned}$$

Clearly the last possibility gives the highest profit. So, no mid-sized cars will be build, just $\frac{16\,000}{101} = 158.42$ economy cars. and lots of SUVs $\left(\frac{17\,000}{101} = 168.32\right)$. This was probably not the intention of the law.

2. Standard Forms

We will be discussing problems of the form:
Maximize

$$f(\vec{x}) = \vec{a} \cdot \vec{x}$$

subject to the conditions

$$\begin{aligned} \vec{b}_1 \cdot \vec{x} &\leq r_1 \\ &\vdots \\ \vec{b}_k \cdot \vec{x} &\leq r_n \\ x_1 &\geq 0 \\ &\vdots \\ x_n &\geq 0 \end{aligned}$$

where $\vec{x} = (x_1, \ldots, x_n), \vec{a}, \vec{b}_1, \ldots, \vec{b}_k \in \Re^n$.

The function $f(\vec{x})$ is linear and therefore convex. The constraint describes a closed convex set. If this set is also bounded, then we know that convex functions defined on convex sets obtain their maximum values at extreme points So the strategy would be to identify all extreme points of this polytope, evaluate the linear function at those extreme points, and then take the point that yields the maximum value.

2.0.5. *Dealing the problem of finding minimum values:* If we are asked to minimize $f(\vec{x}) = \vec{a} \cdot \vec{x}$, we might as well maximize $-f(\vec{x}) = (-\vec{a}) \cdot \vec{x}$.

2.0.6. *Dealing with* $\geq$:. If one of the constraints is of the form

$$\vec{b} \cdot \vec{x} \geq r$$

we multiply the inequality by -1, and obtain

$$\left(-\vec{b}\right) \cdot \vec{x} \leq -r$$

2.0.7. *Replacing inequalities by equalities:* Every inequality

$$\vec{b} \cdot \vec{x} \leq r$$

is replaced by

$$\begin{aligned} \vec{b} \cdot \vec{x} + y &= r \\ y &\geq 0 \end{aligned}$$

2.0.8. *Dealing with negative right-hand-sides:* Each equation of the form

$$\vec{b} \cdot \vec{x} = r$$

where $r < 0$ is replaced by

$$\left(-\vec{b}\right) \cdot \vec{x} = -r$$

In this way, we produce a *standard form* of a linear programming problem:

Maximize the objective function

$$f(\vec{x}) = \vec{a} \cdot \vec{x}$$

subject to the constraints

$$\begin{aligned} B\vec{x} &= \vec{b} \\ \vec{x} &\geq 0 \end{aligned}$$

where

$$\vec{b} \geq \vec{0}$$

The features of a problem in standard form are:

(1) The objective function is to be maximized.
(2) All constraints except the nonnegativity conditions are strict equations.
(3) The independent variables are all nonnegative.
(4) The constant to the right of each equality sign in each constraint is nonnegative.

Later, when we discuss duality, we will also use different standarized forms of linear programming problems.

EXAMPLE 47. *Consider the problem:*

Minimize

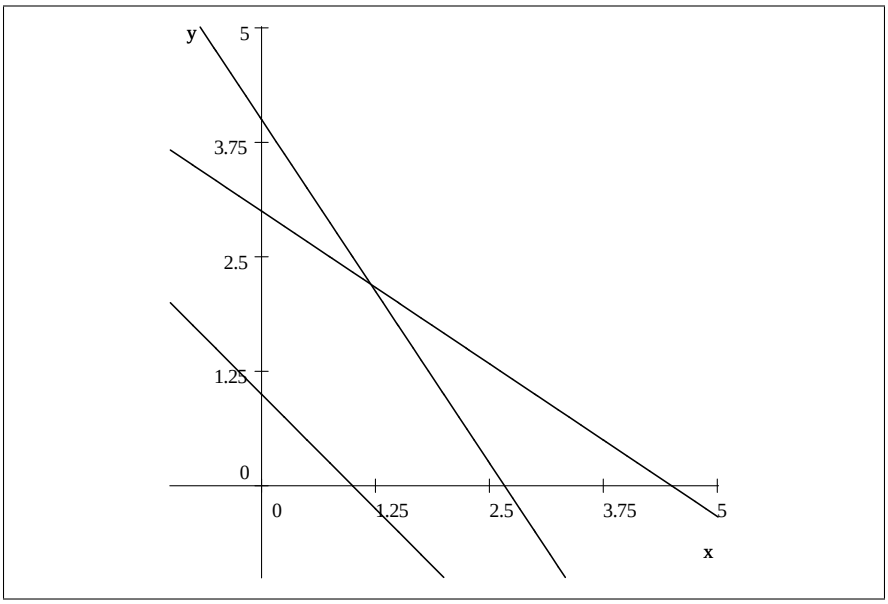
$$f(x, y) = 4x - 2y$$

subject to

$$\begin{aligned} 3x + 2y &\leq 8 \\ 2x + 3y &\leq 9 \\ x + y &\geq 1 \\ x &\geq 0 \\ y &\geq 0 \end{aligned}$$

Find the corresponding standard form for this problem.

Even though this is not part of the problem, we first plot the feasible region:



First, we would like to find the corresponding problem for finding maximum values:
Maximize

$$f(x,y) = -4x + 2y$$

subject to

$$\begin{aligned} 3x + 2y &\leq 8 \\ 2x + 3y &\leq 9 \\ x + y &\geq 1 \\ x &\geq 0 \\ y &\geq 0 \end{aligned}$$

Then we convert $x + y \geq 1$ into $-x - y \leq -1$:
Maximize

$$f(x,y) = -4x + 2y$$

subject to

$$\begin{aligned} 3x + 2y &\leq 8 \\ 2x + 3y &\leq 9 \\ -x - y &\leq -1 \\ x &\geq 0 \\ y &\geq 0 \end{aligned}$$

Next, we we add slack variables:
Maximize

$$f(x,y) = -4x + 2y$$

subject to

$$\begin{array}{rcrcrcrcrcl} 3x & + & 2y & + & u & & & & & = & 8 \\ 2x & + & 3y & & & + & v & & & = & 9 \\ -x & - & y & & & & & + & w & = & -1 \\ & & & & & & & & x & \geq & 0 \\ & & & & & & & & y & \geq & 0 \\ & & & & & & & & u & \geq & 0 \\ & & & & & & & & v & \geq & 0 \\ & & & & & & & & w & \geq & 0 \end{array}$$

The third equation need to be multiplied by -1:
Maximize

$$f(x,y) = -4x + 2y$$

subject to

$$\begin{array}{rcrcrcrcrcl} 3x & + & 2y & + & u & & & & & = & 8 \\ 2x & + & 3y & & & + & v & & & = & 9 \\ x & + & y & & & & & - & w & = & 1 \\ & & & & & & & & x & \geq & 0 \\ & & & & & & & & y & \geq & 0 \\ & & & & & & & & u & \geq & 0 \\ & & & & & & & & v & \geq & 0 \\ & & & & & & & & w & \geq & 0 \end{array}$$

Finally, the problem is written in matrix form:

Maximize

$$f(\vec{x}) = (-4, 2, 0, 0, 0) \begin{pmatrix} x \\ y \\ u \\ v \\ w \end{pmatrix}$$

subject to

$$\begin{pmatrix} 3 & 2 & 1 & 0 & 0 \\ 2 & 3 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ u \\ v \\ w \end{pmatrix} = \begin{pmatrix} 8 \\ 9 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} x \\ y \\ u \\ v \\ w \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

or:

Maximize

$$f(\vec{x}) = (-4, 2, 0, 0, 0) \cdot \vec{x}$$

subject to

$$\begin{pmatrix} 3 & 2 & 1 & 0 & 0 \\ 2 & 3 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & -1 \end{pmatrix} \vec{x} = \begin{pmatrix} 8 \\ 9 \\ 1 \end{pmatrix}$$
$$\vec{x} \geq \vec{0}$$

Note that the equations

$$\begin{pmatrix} 3 & 2 & 1 & 0 & 0 \\ 2 & 3 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & -1 \end{pmatrix} \vec{x} = \begin{pmatrix} 8 \\ 9 \\ 1 \end{pmatrix}$$
$$\vec{x} \geq \vec{0}$$

and

$$\begin{aligned} 3x + 2y &\leq 8 \\ 2x + 3y &\leq 9 \\ x + y &\geq 1 \\ x &\geq 0 \\ y &\geq 0 \end{aligned}$$

have exactly the same solutions. A bijection between both sets is given by the map

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \\ 8 - 3x - 2y \\ 9 - 2x - 3y \\ x + y - 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 8 \\ 9 \\ -1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -3 & -2 \\ -2 & -3 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

This map is the sum of a linear map and a constant, i.e. an affine map. Bijective affine maps preserve convex sets and their extreme points. Hence it does not matter whether we look for the extreme points of the original convex set or the extreme points of the new convex set - in our concrete example, the extreme points of the solutions of

$$\begin{pmatrix} 3 & 2 & 1 & 0 & 0 \\ 2 & 3 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & -1 \end{pmatrix} \vec{x} = \begin{pmatrix} 8 \\ 9 \\ 1 \end{pmatrix}$$
$$\vec{x} \geq \vec{0}$$

3. Extreme Points of Feasible Sets and Basic Solutions

We start again with the problem

$$\begin{aligned} \text{Maximize } f(\vec{x}) &= \vec{a} \cdot \vec{x} \\ \text{Subject to } A\vec{x} &= \vec{b} \\ \vec{x} &\geq \vec{0} \end{aligned}$$

where $\vec{b} \geq 0$. The set of all solutions of

$$\begin{aligned} A\vec{x} &= \vec{b} \\ \vec{x} &\geq 0 \end{aligned}$$

is called the feasible set. The feasible set is a closed convex set (a polytope), and the objective function $f(\vec{x}) = \vec{a} \cdot \vec{x}$ is linear and therefore convex. If the feasible set is also bounded, then maximum of $f(\vec{x})$ will be obtained at an extreme point. Actually, we will show that this remains true even without the assumption of boundedness.

In order to find the extreme points, we find basic solutions. They are defined as follows.

First, we make sure that the rows of A are linearly independent by deleting equations in $A\vec{x} = \vec{b}$ that are implied by other equations. After we have done this, the matrix A is a matrix with m rows and n columns. Hence $\vec{x} \in \Re^n$ and $\vec{b} \in \Re^m$. Moreover, since the rows of A are now linearly independent, it follows that $m \leq n$.

DEFINITION 23. *Let A be a $m \times n$ matrix, assume that the rows of A are linearly independent, and assume that $\vec{b} \geq \vec{0}$.*

1. *We say that B is a basic submatrix of A, if B is invertible, and B is obtained from A by deleting $n - m$ columns.*
2. *Assume that we deleted the columns $j_1, \ldots, j_{n-m}$ from A to obtain B. Then x_i is a basic variable of $\vec{x} = (x_1, \ldots, x_n)$, if i is not one of the indices $j_1, \ldots, j_{n-m}$.*
3. *Deleting all non-basic variables from $\vec{x}$ and keeping only the basic variables leads to a vector $\vec{x}_B \in \Re^m$.*
4. *We say that $\vec{x}$ is a basic solution, if there is a basic submatrix B of A so that*
   (a) *$\vec{x}_B = B^{-1}\vec{b}$, and*
   (b) *$x_i = 0$ for all non-basic variables.*
5. *If $\vec{x}$ is a feasible solution of $A\vec{x} = \vec{b}$, and if $\vec{x}$ is also a basic solution for some choice of a basic submatrix of A, then $\vec{x}$ is called a basic feasible solution.*

PROPOSITION 21. *Let A be a matrix with m linearly independent rows and n columns. Assume that $\vec{b} \geq 0$ The columns of A are denoted by $\vec{a}_1, \ldots, \vec{a}_n$. A solution $\vec{x}_0$ of $A\vec{x} = \vec{b}$ is a basic solution, if and only if*

$$\{\vec{a}_i : x_i \neq 0\}$$

is a linearly independent set of vectors.

PROOF. Indeed, if $\vec{x}_0 = (x_1, \ldots, x_n)$ is a basic solution, then there are indices $j_1, \ldots, j_m$ so that the matrix $B = (\vec{a}_{j_1}, \ldots, \vec{a}_{j_m})$ is invertible and $x_i = 0$ for all non-basic variables. Since

$$\{\vec{a}_i : x_i \neq 0\} \subseteq \{\vec{a}_{j_1}, \ldots, \vec{a}_{j_m}\}$$

it follows that $\{\vec{a}_i : x_i \neq 0\}$ is a linearly independent set of vectors.

Conversely, assume $A\vec{x}_0 = \vec{b}$ and that $\{\vec{a}_i : x_i \neq 0\}$ is a linearly independent set of vectors. We would like to show that there are columns $j_1, \ldots j_m$ so that $B = (\vec{a}_{j_1}, \ldots, \vec{a}_{j_m})$ is an invertible matrix and that $x_i = 0$ if i is a non-basic variable. We know that the m rows of A are linearly independent. Hence A has also m linearly independent columns. It follows that the can extend the set $\{\vec{a}_i : x_i \neq 0\}$ to a basis of the column space of A (i.e. of all linear combinations of columns of A, a.k.a. the range space of A), and we can even achieve this by using only columns of A. Hence we find m columns $\vec{a}_{j_1}, \ldots, \vec{a}_{j_m}$ of A so that $\{\vec{a}_i : x_i \neq 0\} \subseteq \{\vec{a}_{j_1}, \ldots, \vec{a}_{j_m}\}$. We conclude that $x_k = 0$ whenever x_k is a non-basic variable. Hence $\vec{x}_0$ is a basic solution. □

EXAMPLE 48. *Consider the set of all solutions of*

$$\begin{aligned} 2x + 3y + 4z &= 1 \\ x + y + z &= 2 \\ 3x + 4y + 5z &= 3 \\ x, y, z &\geq 0 \end{aligned}$$

Find all basic feasible solutions.

The last equation is redundant. We can delete it from the system:

$$\begin{aligned} 2x + 3y + 4z &= 1 \\ x + y + z &= 2 \\ x, y, z &\geq 0 \end{aligned}$$

There are three possible basic submatrices:

$$\begin{aligned} B_1 &= \begin{pmatrix} 2 & 3 \\ 1 & 1 \end{pmatrix} \\ B_2 &= \begin{pmatrix} 2 & 4 \\ 1 & 1 \end{pmatrix} \\ B_3 &= \begin{pmatrix} 3 & 4 \\ 1 & 1 \end{pmatrix} \end{aligned}$$

For B_1, the basic solution would satisfy

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 & 3 \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 5 \\ -3 \end{pmatrix}$$

This will not lead to a feasible solution, because one of the basic variables is negative.
- Repeat the same procedure with B_2 :

$$\begin{pmatrix} x \\ z \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{7}{2} \\ -\frac{3}{2} \end{pmatrix}$$

This solution is also not feasible.
For B_3, we find

$$\begin{pmatrix} y \\ z \end{pmatrix} = \begin{pmatrix} 7 \\ -5 \end{pmatrix}$$

Again, this is not feasible. There are no feasible basic solutions.

EXAMPLE 49. *Consider the set of all solutions of*

$$\begin{aligned} 2x + 3y + 4z + 3w &= 1 \\ x + y + z + 9w &= 2 \\ 3x + 4y + 5z + 12w &= 3 \\ x, y, z, w &\geq 0 \end{aligned}$$

Find all basic feasible solutions.

Again, one equation is redundant:

$$\begin{aligned} 2x + 3y + 4z + 3w &= 1 \\ x + y + z + 9w &= 2 \\ x, y, z, w &\geq 0 \end{aligned}$$

The matrix A is given by

$$A = \begin{pmatrix} 2 & 3 & 4 & 3 \\ 1 & 1 & 1 & 9 \end{pmatrix}$$

There are six possible basic submatrices. Only three lead to basic solutions:

$$B_1 = \begin{pmatrix} 2 & 3 \\ 1 & 9 \end{pmatrix} \text{ (delete columns 2 and 3)}$$

$$\begin{pmatrix} x \\ w \end{pmatrix} = \begin{pmatrix} 2 & 3 \\ 1 & 9 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{1}{5} \\ \frac{1}{5} \end{pmatrix}$$

The corresponding basic solution is

$$\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} \frac{1}{5} \\ 0 \\ 0 \\ \frac{1}{5} \end{pmatrix}$$

$$B_2 = \begin{pmatrix} 3 & 3 \\ 1 & 9 \end{pmatrix} \text{ (delete columns 1 and 3)}$$

$$\begin{pmatrix} y \\ w \end{pmatrix} = \begin{pmatrix} 3 & 3 \\ 1 & 9 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{1}{8} \\ \frac{5}{24} \end{pmatrix}$$

The second basic solution is

$$\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{1}{8} \\ 0 \\ \frac{5}{24} \end{pmatrix}$$

Finally,

$$B_2 = \begin{pmatrix} 4 & 3 \\ 1 & 9 \end{pmatrix} \text{ (delete columns 1 and 2)}$$

$$\begin{pmatrix} z \\ w \end{pmatrix} = \begin{pmatrix} 4 & 3 \\ 1 & 9 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{1}{11} \\ \frac{7}{33} \end{pmatrix}$$

The last basis solution is

$$\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \frac{1}{11} \\ \frac{7}{33} \end{pmatrix}$$

THEOREM 40. *Let A be a matrix with m rows and n columns, assume that the of A are linearly independent and that $\vec{b} \geq \vec{0}$. If Ω is the convex set of all solutions of*

$$\begin{aligned} A\vec{x} &= \vec{b} \\ \vec{x} &\geq 0 \end{aligned}$$

then $\vec{x}$ is an extreme point of Ω if and only if $\vec{x}$ is a feasible basic solution.

PROOF. Let

$$\Omega = \left\{ \vec{x} \in \Re^n : A\vec{x} = \vec{b} \text{ and } \vec{x} \geq 0 \right\}$$

First, we assume that $\vec{x}$ is an extreme point of Ω. We have to show that $\vec{x}$ is a basic solution.

Reorder the coordinates of $\vec{x} = (x_1, \ldots, x_n)$ so that the first k coordinates are all strictly positive whereas $x_i = 0$ for $i > k$:

$$\vec{x} = (x_1, \ldots, x_k, 0, \ldots, 0)$$

This corresponds to a renumbering of the columns of A. The i^{th} column of A will be denoted by $\vec{a}_i$. We now show that the rows $\vec{a}_1, \ldots, \vec{a}_k$ are linearly independent. Indeed, assume that $y_1\vec{a}_1 + \cdots + y_k\vec{a}_k = \vec{0}$. Then we have to verify that $y_1 = \cdots = y_k = 0$. Before we do this, we extend the $y_i's$ to a vector in $\Re^n$ by adding $0's$ in the remaining coordinates:

$$\vec{y} = (y_1, \ldots, y_k, 0, \ldots, 0)$$

Then

$$\begin{aligned} A\vec{y} &= y_1\vec{a}_1 + \cdots + y_k\vec{a}_k + 0 \cdot \vec{a}_{k+1} + \cdots + 0 \cdot \vec{a}_n \\ &= (y_1\vec{a}_1 + \cdots + y_k\vec{a}_k) + 0 \cdot (\vec{a}_{k+1} + \cdots + \vec{a}_n) \\ &= \vec{0} + \vec{0} \\ &= \vec{0} \end{aligned}$$

We now assume that $y_i \neq 0$ for at least one index i.Under this assumption, we define

$$\varepsilon = \min \left\{ \frac{x_i}{|y_i|} : y_i \neq 0 \right\}$$

Then $\varepsilon > 0$, since $y_i \neq 0$ implies that $i \leq k$ and hence $x_i > 0$. Moreover, we have

$$\varepsilon \leq \frac{x_i}{|y_i|}$$

i.e.

$$\varepsilon |y_i| \leq x_i$$

for all i with $y_i \neq 0$. Clearly, this inequality is also valid for $y_i = 0$, and we obtain

$$\pm \varepsilon y_i \leq x_i$$

for all indices i. This implies

$$0 \leq x_i \pm \varepsilon y_i$$

for all indices i, or

$$\vec{0} \leq \vec{x} \pm \varepsilon \vec{y}$$

We know that $A\vec{y} = 0$, which leads to

$$\begin{aligned} A(\vec{x} \pm \varepsilon \vec{y}) &= A\vec{x} \pm \varepsilon A\vec{y} \\ &= \vec{b} \end{aligned}$$

and therefore to $\vec{x} \pm \varepsilon \vec{y} \in \Omega$. Since

$$\vec{x} = \frac{1}{2}(\vec{x} + \varepsilon \vec{y}) + \frac{1}{2}(\vec{x} - \varepsilon \vec{y})$$

and since $\vec{x}$ is an extreme point of A, we find that $\vec{x} = \vec{x} + \varepsilon \vec{y} = \vec{x} - \varepsilon \vec{y}$. This allows us to conclude that $\varepsilon \vec{y} = \vec{0}$, contradicting the assumption that $\varepsilon \neq 0$ and $y_i \neq 0$ for at least one index i.

We now have shown that the columns $\vec{a}_1, \ldots, \vec{a}_k$ of A are linearly independent. The preceding proposition shows that $\vec{x}$ is a basic feasible solution.

Conversely, assume that $\vec{x}$ is a basic feasible solution of $A\vec{x} = \vec{b}$, $\vec{x} \geq 0$. Then there is an invertible submatrix B of A so that all non-basic variables of $\vec{x} = (x_1, \ldots, x_n)$ are equal to 0. Again, we renumber the coordinates of $\vec{x}$ in such a way that $x_{m+1} = \cdots = x_m = 0$. Then

$$A\vec{x} = B \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} = \vec{b}$$

We would like to show that $\vec{x}$ is an extreme point of Ω. Assume that there are vectors $\vec{y}, \vec{z} \in \Omega$ and a number λ with $0 < \lambda < 1$ so that

$$\vec{x} = \lambda \vec{y} + (1 - \lambda)\vec{z}$$

We have to show that $\vec{y} = \vec{z}$. Writing the equation $\vec{x} = \lambda \vec{y} + (1 - \lambda)\vec{z}$ coodinate-wise yields

$$x_i = \lambda y_i + (1 - \lambda) z_i$$

Since all coordinates of $\vec{y}$ and $\vec{z}$ are positive, and since $x_i = 0$ for $i \geq m + 1$, we find that also $y_i = z_i = 0$ for $i \geq m + 1$. Hence

$$\begin{aligned} \vec{b} &= A\vec{y} \\ &= B \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \end{aligned}$$

and similarly

$$\vec{b} = B \begin{pmatrix} z_1 \\ \vdots \\ z_m \end{pmatrix}$$

This implies that

$$\begin{aligned} B\begin{pmatrix} y_1 - z_1 \\ \vdots \\ y_m - z_m \end{pmatrix} &= B\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} - B\begin{pmatrix} z_1 \\ \vdots \\ z_m \end{pmatrix} \\ &= \vec{b} - \vec{b} = \vec{0} \end{aligned}$$

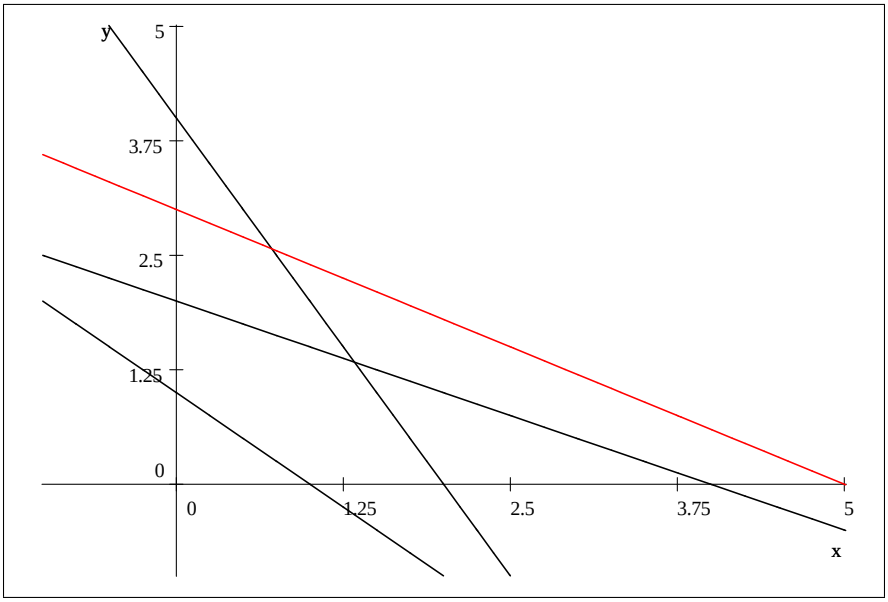
Since B is invertible, we conclude that

$$\begin{pmatrix} y_1 - z_1 \\ \vdots \\ y_m - z_m \end{pmatrix} = B^{-1}\vec{0} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

and therefore $y_i = z_i$ for $1 \leq i \leq m$. Since we already remarked earlier that $y_i = z_i = 0$ for $i \geq m+1$, it follows that $y_i = z_i$ for all indices i, i.e. $\vec{y} = \vec{z}$. □

EXAMPLE 50. *Find all extreme points of the set given by*

$$\begin{aligned} x + y &\geq 1 \\ x + 2y &\leq 4 \\ 2x + y &\leq 4 \\ 3x + 5y &\leq 15 \\ x, y &\geq 0 \end{aligned}$$



$3x + 5y = 15$ *in red*

First, we bring the problem into standard form

$$\begin{pmatrix} 1 & 1 & -1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 & 0 \\ 2 & 1 & 0 & 0 & 1 & 0 \\ 3 & 5 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \\ 4 \\ 15 \end{pmatrix}$$

One basic submatrix can be obtained by deleting columns 3 and 6 from the original matrix. We now have to solve the equation

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 2 & 1 & 0 & 1 \\ 3 & 5 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ b \\ c \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \\ 4 \\ 15 \end{pmatrix}$$

and find

$$\begin{pmatrix} x \\ y \\ b \\ c \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 2 & 1 & 0 & 1 \\ 3 & 5 & 0 & 0 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 4 \\ 4 \\ 15 \end{pmatrix} = \begin{pmatrix} -5 \\ 6 \\ -3 \\ 8 \end{pmatrix}$$

This would not lead to a feasible basic solutions.

Deleting columns 4 and 5 from the original matrix leads to the equation

$$\begin{pmatrix} 1 & 1 & -1 & 0 \\ 1 & 2 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 5 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ a \\ d \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \\ 4 \\ 15 \end{pmatrix}$$

with the solutions

$$\begin{pmatrix} x \\ y \\ a \\ d \end{pmatrix} = \begin{pmatrix} 1 & 1 & -1 & 0 \\ 1 & 2 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 5 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 4 \\ 4 \\ 15 \end{pmatrix} = \begin{pmatrix} \frac{4}{3} \\ \frac{4}{3} \\ \frac{5}{3} \\ \frac{13}{3} \end{pmatrix}$$

The corresponding basic feasible solution is

$$\begin{pmatrix} x \\ y \\ a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} \frac{4}{3} \\ \frac{4}{3} \\ \frac{5}{3} \\ 0 \\ 0 \\ \frac{13}{3} \end{pmatrix}$$

So the corresponding extreme point is $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{4}{3} \\ \frac{4}{3} \end{pmatrix}$. Since the slack variables b and c are equal to 0, this solution represents the solution of the system

$$\begin{aligned} x + 2y &= 4 \\ 2x + y &= 4 \end{aligned}$$

This method needs to be continued until all extreme points are found.

4. Basic Solutions and Maxima of Convex Functions

THEOREM 41. *Let A be a matrix with m rows and n columns, assume that the rows of A are linearly independent, and assume that $\vec{b} \geq \vec{0}$ Let $f(\vec{x})$ be a convex function, defined on the set $\left\{\vec{x} \in \Re^n : \vec{x} \geq \vec{0} \text{ and } A\vec{x} = \vec{b}\right\}$ If the problem*

$$\text{Maximize } f(\vec{x})$$

with respect to

$$\begin{aligned} A\vec{x} &= \vec{b} \\ \vec{x} &\geq 0 \end{aligned}$$

has a solution, then it has also a feasible basic solution.

PROOF. Let $\vec{x}_0 = [x_1, \ldots, x_n]^T$ be any solution of $A\vec{x} = \vec{b}$ satisfying $\vec{x}_0 \geq 0$ that maximizes $f(x)$ We have to modify $\vec{x}_0$ to obtain a basic feasible solution. Again, the columns of A will be denoted by $\vec{a}_1, \ldots, \vec{a}_n$. Consider

$$\{\vec{a}_i : x_i \neq 0\}$$

If this set is linearly independent, then, according to the previous proposition, we found a basic solution. On the other hand, if this set is not linearly independent, then we will modify $\vec{x}_0$, obtaining a feasible solution $\vec{y}_0 = [y_1, \ldots, y_n]^T$ so that

$$\{\vec{a}_i : y_i \neq 0\} \subseteq \{\vec{a}_i : x_i \neq 0, i \neq i_0\}$$

where i_0 is a suitable index with $x_{i_0} \neq 0$. If the set $\{\vec{a}_i : y_i \neq 0\}$ is still not independent, we repeat this construction, removing more and more vectors from $\{\vec{a}_i : x_i \neq 0\}$ until we find an independent set.

Here is the construction: If the set $\{\vec{a}_i : x_i \neq 0\}$ is not linearly independent, then we can find numbers s_i not all 0 so that

$$\sum_{i \text{ is an index so that } x_i \neq 0} s_i \vec{a}_i = \vec{0}$$

If $x_i = 0$, we let $s_i = 0$. Then

$$\sum_{i=1}^{n} s_i \vec{a}_i = \vec{0}$$

i.e. $\vec{s} = s\,[1, \ldots, s_n]^T$ satisfies

$$A\vec{s} = \vec{0}$$

Furthermore

$$s_i \neq 0 \implies x_i \neq 0$$

or, equivalently,

$$x_i = 0 \implies s_i = 0$$

For each number r we have

$$\begin{aligned} A(\vec{x}_0 - r\vec{s}) &= A\vec{x}_0 - rA\vec{s} \\ &= \vec{b} + \vec{0} \\ &= \vec{b} \end{aligned}$$

We now would like to choose $r > 0$ in such a way that $\vec{x}_0 - r\vec{s} \geq \vec{0}$. For the coordinates, this means that

$$x_i - rs_i \geq 0$$

If we only allow values of r for which $r \geq 0$, this equation is obvious if $s_i \leq 0$ (recall that $\vec{x} \geq 0$). If $s_i > 0$, then we have to choose r so that

$$\begin{array}{rcl} x_i & \geq & rs_i \\ \dfrac{x_i}{s_i} & \geq & r \end{array}$$

We can use any r so that

$$0 \leq r \leq \min \left\{ \frac{x_i}{s_i} : s_i > 0 \right\}$$

This leads us to the definition

$$r = \min \left\{ \frac{x_i}{s_i} : s_i > 0 \right\}$$

and

$$\vec{y} = \vec{x}_0 - r\vec{s}$$

Then all coordinates of $\vec{y}$ are positive, and hence $\vec{y}$ is a feasible solution. If $x_i = 0$, then $s_i = 0$ and therefore $y_i = 0$. Hence,

$$\{\vec{a}_i : y_i \neq 0\} \subseteq \{\vec{a}_i : x_i \neq 0\}$$

Let i_0 be an index for which $r = \frac{x_{i_0}}{s_{i_0}}$ with $s_{i_0} > 0$. Then $x_{i_0} - rs_{i_0} = 0$, hence $y_{i_0} = 0$. Since $s_{i_0} \neq 0$ implies that $x_{i_0} \neq 0$, it follows that $i_0 \in \{i : x_i \neq 0\}$, and

$$\{\vec{a}_i : y_i \neq 0\} \subseteq \{\vec{a}_i : x_i \neq 0, i \neq i_0\}$$

Similarly, we replace $\vec{s}$ by $-\vec{s}$ and repeat the previous steps: Let

$$t = \min \left\{ \frac{x_i}{-s_i} : -s_i > 0 \right\}$$

and

$$\begin{array}{rcl} \vec{z} & = & \vec{x}_0 - t\left(-\vec{s}\right) \\ & = & \vec{x}_0 + t\vec{s} \end{array}$$

Then all coordinates of $\vec{z}$ are positive, and hence $\vec{z}$ is also a feasible solution. Let i_1 be an index for which $\frac{x_{i_1}}{-s_{i_1}} = t$. Then

$$\{\vec{a}_i : z_i \neq 0\} \subseteq \{\vec{a}_i : x_i \neq 0, i \neq i_1\}$$

It remains to show that either $\vec{y}$ or $\vec{z}$ also maximizes the function $f\left(\vec{x}\right)$.Let

$$\lambda = \frac{t}{t+r}$$

Then $0 \leq \lambda \leq 1$ and

$$\begin{array}{rcl} \lambda\vec{y} + (1-\lambda)\,\vec{z} & = & \dfrac{t}{t+r}\left(\vec{x}_0 - r\vec{s}\right) + \dfrac{r}{t+r}\left(\vec{x}_0 + t\vec{s}\right) \\ & = & \vec{x}_0 \end{array}$$

It follows that

$$f\left(\vec{x}_0\right) \leq \lambda f\left(\vec{y}\right) + (1-\lambda)\, f\left(\vec{z}\right)$$

If both $f(\vec{y})$ and $f(\vec{y})$ are strictly less then $f(\vec{x}_0)$, then we would obtain the contradiction

$$f(\vec{x}_0) < \lambda f(\vec{x}_0) + (1-\lambda) f(\vec{x}_0) = f(\vec{x}_0)$$

Hence either $f(\vec{x}_0) \leq f(\vec{y})$ or $f(\vec{x}_0) \leq f(\vec{z})$ Since $f(\vec{x}_0)$ was assumed to be the maximum of all values $f(\vec{x})$, we conclude that either $f(\vec{x}_0) = f(\vec{y})$ or $f(\vec{x}_0) = f(\vec{z})$. Hence f takes it maximum either at $\vec{y}$ or at $\vec{z}$. □

If $f(\vec{x}) = 0$ in the previous theorem, then every feasible solution maximizes $f(\vec{x})$. Hence

COROLLARY 5. *Let A be a matrix with m rows and n columns, and assume that the of A are linearly independent. If the system*

$$\begin{aligned} A\vec{x} &= \vec{b} \\ \vec{x} &\geq 0 \end{aligned}$$

has a solution, then it has also a feasible basic solution.

EXAMPLE 51. *Consider the system of linear equations*

$$\begin{aligned} 2x + 3y - z - w &= -1 \\ 4x + y + z - 2w &= 9 \end{aligned}$$

Find

(1) *the general solution;*
(2) *a feasible solution;*
(3) *a feasible basic solution.*

We first find the general solution of

$$\begin{aligned} 2x + 3y - z - w &= -1 \\ 4x + y + z - 2w &= 9 \end{aligned}$$

Subtracting the first equation twice from the second equation gives

$$\begin{aligned} 2x + 3y - z - w &= -1 \\ -5y + 3z &= 11 \end{aligned}$$

So we can solve for x and y in terms of z and w :

$$\begin{aligned} x &= \frac{1}{2}w - \frac{2}{5}z + \frac{14}{5} \\ y &= \frac{3}{5}z - \frac{11}{5} \end{aligned}$$

or

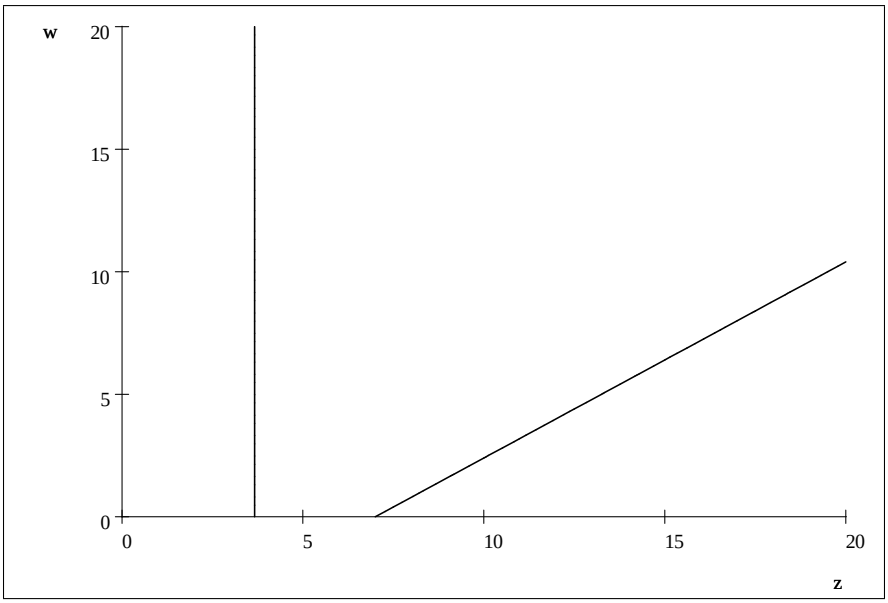
$$\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} \frac{1}{2}w - \frac{2}{5}z + \frac{14}{5} \\ \frac{3}{5}z - \frac{11}{5} \\ z \\ w \end{pmatrix}$$

If we would like to find all solutions satisfying $x, y, z, w \geq 0$, we are led to

$$\begin{aligned} \frac{1}{2}w - \frac{2}{5}z + \frac{14}{5} &\geq 0 \\ \frac{3}{5}z - \frac{11}{5} &\geq 0 \\ z &\geq 0 \\ w &\geq 0 \end{aligned}$$

or

$$\begin{aligned} 4z - 5w &\leq 28 \\ z &\geq \frac{11}{3} \end{aligned}$$



We see that there are two extreme points, namely

$$\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} \frac{4}{3} \\ 0 \\ \frac{11}{3} \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ 7 \\ 0 \end{pmatrix}$$

We can also use parts of the previous proof to construct at least one of those extreme point. If we start the iteration of the previous proof with $(z, w) = (5, 2)$, then,

$$\vec{x} = \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} \frac{9}{5} \\ \frac{4}{5} \\ 5 \\ 2 \end{pmatrix}$$

is a feasible solution. We use this feasible solution as a start point to construct a basic feasible solution. The matrix A is given by

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

The columns are

$$\begin{aligned} \vec{a}_1 &= \begin{pmatrix} 2 \\ 4 \end{pmatrix} \\ \vec{a}_2 &= \begin{pmatrix} 3 \\ 1 \end{pmatrix} \\ \vec{a}_3 &= \begin{pmatrix} -1 \\ 1 \end{pmatrix} \\ \vec{a}_4 &= \begin{pmatrix} -1 \\ -2 \end{pmatrix} \end{aligned}$$

The set $\{\vec{a}_i : x_i \neq 0\}$ consists of all four columns. They are dependent. For example,

$$-2\vec{a}_1 + 3\vec{a}_2 + 5\vec{a}_3 = \vec{0}$$

So, in the proof of the theorem we can pick

$$\begin{aligned} \vec{x} &= \left(\frac{9}{5}, \frac{4}{5}, 5, 2\right) \\ \vec{s} &= (-2, 3, 5, 0) \end{aligned}$$

We find

$$\begin{aligned} r &= \min\left\{\frac{x_i}{s_i} : s_i > 0\right\} \\ &= \min\left\{\frac{\frac{4}{5}}{3}, \frac{5}{5}\right\} \\ &= \frac{4}{15} \end{aligned}$$

Hence

$$\begin{aligned} \vec{y} &= \vec{x} - r\vec{s} \\ &= \left(\frac{9}{5}, \frac{4}{5}, 5, 2\right) - \frac{4}{15}(-2, 3, 5, 0) \\ &= \left(\frac{7}{3}, 0, \frac{11}{3}, 2\right) \end{aligned}$$

We now repeat this procedure with

$$\vec{x} = \left(\frac{7}{3}, 0, \frac{11}{3}, 2\right)$$

In this case, $\{\vec{a}_i : x_i \neq 0\} = \{\vec{a}_1, \vec{a}_3, \vec{a}_4\}$. These three columns are again dependent. For example,

$$\vec{a}_1 + 0 \cdot \vec{a}_2 + 2\vec{a}_4 = 0$$

In this case, we pick

$$\begin{aligned}\vec{x} &= \left(\frac{7}{3}, 0, \frac{11}{3}, 2\right)\\ \vec{s} &= (1, 0, 0, 2)\end{aligned}$$

and

$$\begin{aligned}r &= \min\left\{\frac{x_i}{s_i} : s_i > 0\right\}\\ &= \min\left\{\frac{\frac{7}{3}}{1}, \frac{2}{2}\right\}\\ &= 1\end{aligned}$$

and

$$\begin{aligned}\vec{y} &= \vec{x} - r\vec{s}\\ &= \left(\frac{7}{3}, 0, \frac{11}{3}, 2\right) - (1, 0, 0, 2)\\ &= \left(\frac{4}{3}, 0, \frac{11}{3}, 0\right)\end{aligned}$$

We repeat the steps with

$$\vec{x} = \left(\frac{4}{3}, 0, \frac{11}{3}, 0\right)$$

In this case $\{\vec{a}_i : x_i \neq 0\} = \{\vec{a}_1, \vec{a}_3\}$. These columns are indeed independent, and we found a feasible basic solution.

5. The Simplex Algorithm: An Example

We start with an example
Maximize

$$f(x_1, x_2) = 4x_1 + 3x_2$$

subject to

$$\begin{aligned} 3x_1 + 4x_2 &\leq 12 \\ 3x_1 + 3x_2 &\leq 10 \\ 4x_1 + 2x_2 &\leq 8 \end{aligned}$$

:

Convert to standard form:
Maximize

$$\alpha = f(x_1, x_2) = 4x_1 + 3x_2$$

subject to

$$\begin{aligned} 3x_1 + 4x_2 + x_3 &= 12 \\ 3x_1 + 3x_2 + x_4 &= 10 \\ 4x_1 + 2x_2 + x_5 &= 8 \\ x_i &\geq 0 \text{ for } i = 1, 2, \ldots, 5 \end{aligned}$$

We write the problem in matrix form, adding the function $f(x_1, x_2)$ as last row:

$$\left(\begin{array}{ccccc|c} 3 & 4 & 1 & 0 & 0 & 12 \\ 3 & 3 & 0 & 1 & 0 & 10 \\ 4 & 2 & 0 & 0 & 1 & 8 \\ 4 & 3 & 0 & 0 & 0 & \alpha \end{array}\right)$$

It is easy to identify a basic feasible solution. The last three columns form a unit matrix and are therefore linearly independent. This leads to the basic feasible solution

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 12 \\ 10 \\ 8 \end{pmatrix}$$

Except for the last row, adding multiples of one row to another row (i.e. a step in the Gauss-Jordan algorithm) does not change feasible solution. What happens if we add a multiple of one row to the last row? The last row represents

$$\alpha = 4x_1 + 3x_2$$

The first row represents the equation

$$\begin{aligned} 3x_1 + 4x_2 + x_3 &= 12 \\ 3x_1 + 4x_2 + x_3 - 12 &= 0 \end{aligned}$$

Hence

$$\begin{aligned} f(x_1, x_2) &= 4x_1 + 3x_2 \\ &= 4x_1 + 3x_2 + 0 \\ &= 4x_1 + 3x_2 + (3x_1 + 4x_2 + x_3 - 12) \\ &= 7x_1 + 7x_2 + x_3 - 12 \end{aligned}$$

So, up to an additive constant of -12, the functions $4x_1 + 3x_2$ and $7x_1 + 7x_2 + x_3$ have the same maximum on the feasible set, and the maximum of the new function is by 12 units larger as the maximum of the old function. - In the matrix, adding row No. 1 to the last row gives

$$\left(\begin{array}{ccccc|c} 3 & 4 & 1 & 0 & 0 & 12 \\ 3 & 3 & 0 & 1 & 0 & 10 \\ 4 & 2 & 0 & 0 & 1 & 8 \\ 7 & 7 & 1 & 0 & 0 & \alpha + 12 \end{array}\right)$$

The new last row represents the new function we are trying to maximize.

It is the objective of the simplex algorithm to use Gauss-Jordan steps to

(1) make sure the basic submatrix consists of canonical unit vectors,
(2) keep positive entries in the last column (except for the last entry representing the answer for the maximum).
(3) make all entries in the last row 0 or less (except possibly for the last entry representing the offset).

Starting with

$$\left(\begin{array}{ccccc|c} 3 & 4 & 1 & 0 & 0 & 12 \\ 3 & 3 & 0 & 1 & 0 & 10 \\ 4 & 2 & 0 & 0 & 1 & 8 \\ 4 & 3 & 0 & 0 & 0 & \alpha \end{array}\right)$$

we can use Gauss-Jordan steps to change the first column into a unit vector column:

$$\left(\begin{array}{ccccc|c} 3 & 4 & 1 & 0 & 0 & 12 \\ 3 & 3 & 0 & 1 & 0 & 10 \\ 4 & 2 & 0 & 0 & 1 & 8 \\ 4 & 3 & 0 & 0 & 0 & 0 \end{array}\right) \rightarrow \left(\begin{array}{ccccc|c} 0 & \frac{5}{2} & 1 & 0 & -\frac{3}{4} & 6 \\ 0 & \frac{3}{2} & 0 & 1 & -\frac{3}{4} & 4 \\ 1 & \frac{1}{2} & 0 & 0 & \frac{1}{4} & 2 \\ 0 & 1 & 0 & 0 & -1 & \alpha - 8 \end{array}\right)$$

Next, we turn the second column into the first unit vector column:

$$\left(\begin{array}{ccccc|c} 0 & \frac{5}{2} & 1 & 0 & -\frac{3}{4} & 6 \\ 0 & \frac{3}{2} & 0 & 1 & -\frac{3}{4} & 4 \\ 1 & \frac{1}{2} & 0 & 0 & \frac{1}{4} & 2 \\ 0 & 1 & 0 & 0 & -1 & \alpha - 8 \end{array}\right) \rightarrow \left(\begin{array}{ccccc|c} 0 & 1 & \frac{2}{5} & 0 & -\frac{3}{10} & \frac{12}{5} \\ 0 & 0 & -\frac{3}{5} & 1 & -\frac{3}{10} & \frac{2}{5} \\ 1 & 0 & -\frac{1}{5} & 0 & \frac{2}{5} & \frac{4}{5} \\ 0 & 0 & -\frac{2}{5} & 0 & -\frac{7}{10} & \alpha - \frac{52}{5} \end{array}\right)$$

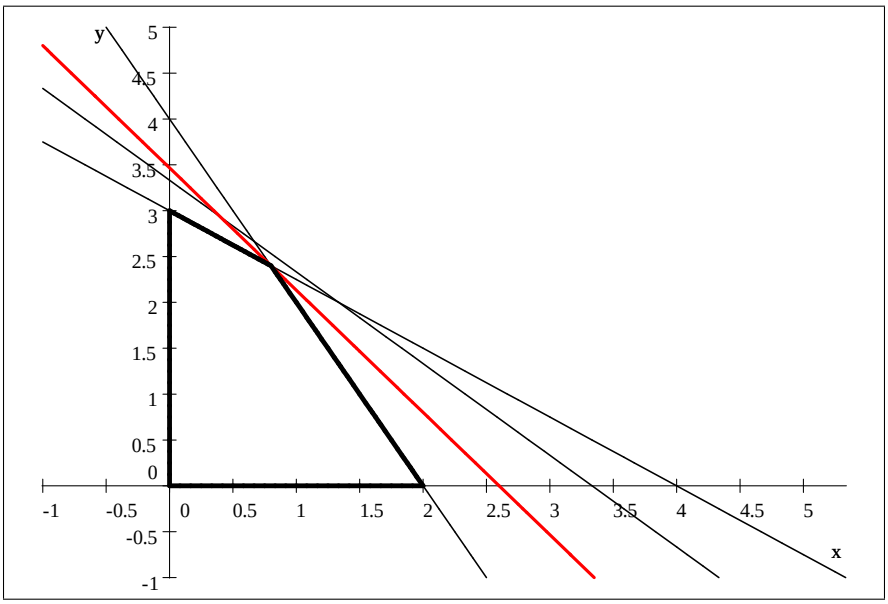
Now all entries in the last row are 0 or less, and the algorithm stops. The new function we have to maximize is

$$\alpha - \frac{52}{5} = -\frac{2}{5}x_3 - \frac{7}{10}x_4$$

Each positive solution $(x_1, x_2, x_3, x_4, x_5) \geq (0, 0, 0, 0, 0)$ satisfies $\alpha - \frac{52}{5} = -\frac{2}{5}x_3 - \frac{7}{10}x_4 \leq 0$, hence 0 is an upper bound on the feasible set. This upper bound is reached for the feasible solution $(x_1, x_2, x_3, x_4, x_5) = \left(\frac{4}{5}, \frac{12}{5}, 0, \frac{2}{5}, 0\right)$ (obtained from the basic columns) It follows that 0 is the maximum on the feasible set. The

maximum of the original function is obtained from the equation $\alpha - \frac{52}{5} = 0$. Hence the maximum of the original function on the feasible set is equal to $\frac{52}{5}$, obtained at $(x_1, x_2) = \left(\frac{4}{5}, \frac{12}{5}\right)$.

The following graph represents the feasible set (black) and the maximized objective function (red).



6. The Simplex Algorithm

Let us consider again the problem

$$\text{Maximize } f(\vec{x}) = \vec{a} \cdot \vec{x}$$

subject to

$$\begin{aligned} A\vec{x} &= \vec{b} \\ \vec{x} &\geq \vec{0} \end{aligned}$$

where A has m linearly independent rows and n columns, and where $\vec{b} \geq 0$. We also assume that we already know a basic feasible solution $\vec{x}_0$- we will discuss later how to construct such solution.

If the original solution does not maximize $f(\vec{x}) = \vec{a} \cdot \vec{x}$, then the simplex algorithm is a method to find a basic feasible solution $\vec{x}_1$ with $f(\vec{x}_1) > f(\vec{x}_0)$.

Here is a description of the basic step of the simplex algorithm. This basic step will also serve to decide whether or not $\vec{x}_0$ maximizes $f(\vec{x}) = \vec{a} \cdot \vec{x}$ subject to the constraints. Parallel to describing these steps abstractly, we will write a short version of the algorithm (taking the training wheels off) and work through an example.

STEP 1 It is convenient to maximize

$$f(\vec{x}) = \vec{a} \cdot \vec{x} + z_{offset}$$

instead of $f(\vec{x}) = \vec{a} \cdot \vec{x}$, where z_{offset} is any real number. In the beginning, when we enter the algorithm,

$$z_{offset} = 0$$

STEP 2 In order to simplify notation, we will assume that the basic variables in $\vec{x}_0$ are listed first:

$$\vec{x}_0 = (x_1, \ldots, x_m, 0, \ldots 0)$$

All the number x_i are non-negative, even so some of them might still be equal to 0.Performing the Gauss-Jordan elimination process on the augmented matrix $\left(A, \vec{b}\right)$ will transform the system

$$A\vec{x} = \vec{b}$$

into

$$A_0\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$$

Why is the new right side equal to $\vec{x}_0$? Since the first m columns are linearly independent, Gauss Jordan will change $\left(A, \vec{b}\right)$ into a matrix of the form where

$$(A_0, \vec{y}) = \begin{pmatrix} 1 & 0 & \ldots & 0 & a_{1,m+1} & \ldots & a_{1,n} & y_1 \\ 0 & 1 & \ddots & \vdots & a_{2,m+1} & \ldots & a_{2,n} & y_2 \\ \vdots & \ddots & \ddots & 0 & \vdots & & \vdots & \vdots \\ 0 & \ldots & 0 & 1 & a_{m,m+1} & \ldots & a_{n,m} & y_m \end{pmatrix}$$

Since $\vec{x}_0 = (x_1, \ldots, x_m, 0, \ldots, 0)$ was a solution of $A\vec{x} = \vec{b}$, and since the Gauss-Jordan method does not alter solutions, we find that

$$\begin{aligned} A_0\vec{x}_0 &= \vec{y}_0 \\ \vec{y}_0 &= \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \end{aligned}$$

However, since A_1 starts with the unit matrix,

$$A_0\vec{x}_0 = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$$

Hence

$$\vec{y}_0 = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$$

We now find that

$$T_0 = (A_0, \vec{y}_0) = \begin{pmatrix} 1 & 0 & \ldots & 0 & a_{1,m+1} & \ldots & a_{1,n} & y_1 \\ 0 & 1 & \ddots & \vdots & a_{2,m+1} & \ldots & a_{2,n} & y_2 \\ \vdots & \ddots & \ddots & 0 & \vdots & & \vdots & \vdots \\ 0 & \ldots & 0 & 1 & a_{m,m+1} & \ldots & a_{n,m} & y_m \end{pmatrix}$$

Since $\vec{x}_0$ is feasible and not degenerate, the new right side $(x_1, \ldots, x_m) = (y_1, \ldots, y_m)$ still has only positive entries. For the next step, we assume that we are starting with this augmented matrix

STEP 3 We now "reduce" the function $f(\vec{x}) = \vec{a} \cdot \vec{x}$. Let

$$\vec{a} = (\alpha_1, \ldots, \alpha_m, \alpha_{m+1}, \ldots, \alpha_n)$$

Every feasible solution $\vec{x}$ will satisfy

$$A_0\vec{x} = \vec{y}_0$$

If $\vec{r}_i$ is the i^{th} row of A_0, then

$$\vec{r}_i \cdot \vec{x} = y_i$$

Multiplying this equation by α_i leads to

$$(\alpha_i\vec{r}_i \cdot \vec{x}) = \alpha_i y_i = \alpha_i x_i$$

Adding those equations for $i = 1, \ldots, m$ yields

$$(\alpha_1\vec{r}_1 \cdot \vec{x}) + \cdots + (\alpha_m\vec{r}_m \cdot \vec{x}) = \alpha_1 x_1 + \cdots + \alpha_m x_m$$

Since $x_{m+1} = \cdots = x_n = 0$, we find

$$\begin{aligned} (\alpha_1\vec{r}_1 \cdot \vec{x}) + \cdots + (\alpha_m\vec{r}_m \cdot \vec{x}) &= \alpha_1 x_1 + \cdots + \alpha_m x_m + \alpha_{m+1} x_{m+1} + \cdots + \alpha_n x_n \\ &= \vec{a} \cdot \vec{x}_0 \end{aligned}$$

or

$$\begin{aligned} (\alpha_1\vec{r}_1 \cdot \vec{x}) + \cdots + (\alpha_m\vec{r}_m \cdot \vec{x}) - \vec{a} \cdot \vec{x}_0 &= 0 \\ (\alpha_1\vec{r}_1 + \cdots + \alpha_m\vec{r}_m) \cdot \vec{x} - \vec{a} \cdot \vec{x}_0 &= 0 \end{aligned}$$

Subtracting this equation from $f(\vec{x}) = \vec{a} \cdot \vec{x}$ leads to

$$\begin{aligned} f(\vec{x}) &= \vec{a} \cdot \vec{x} - (\alpha_1 \vec{r}_1 + \cdots + \alpha_m \vec{r}_m) \cdot \vec{x} + \vec{a} \cdot \vec{x}_0 \\ &= (\vec{a} - (\alpha_1 \vec{r}_1 + \cdots + \alpha_m \vec{r}_m)) \vec{x} + \vec{a} \cdot \vec{x}_0 \end{aligned}$$

Let

$$\vec{a}_0 = \vec{a} - (\alpha_1 \vec{r}_1 + \cdots + \alpha_m \vec{r}_m)$$

Then

$$f(\vec{x}) = \vec{a}_0 \vec{x} + \vec{a} \cdot \vec{x}_0$$

The first component of $\alpha_1 \vec{r}_1$ is equal to α_1 (note that $\vec{r}_1$ is the first row of T_0), whereas the first components of all the other summands $\alpha_i \vec{r}_i$ is equal to 0. Since $\vec{a}$ has also α_1 as first component, the first components cancel, and hence

$$\vec{a}_0 = (0, \ldots, 0, a_{m+1}, \ldots, a_n)$$

Practically, the computation of $\vec{a}_0$ and $\vec{a} \cdot \vec{x}_0$ can be achieved by performing Gauss-Jordan steps after adding the row $(\alpha_1, \ldots, \alpha_n, 0)$ to the matrix T. So instead of maximizing

$$f(\vec{x}) = \vec{a} \cdot \vec{x}$$

we might as well maximize

$$f_0(\vec{x}) = \vec{a}_0 \cdot \vec{x}$$

and then increment the new maximum by

$$z_0 = \vec{a} \cdot \vec{x}_0$$

Define the new z_{offset} by

$$z_{offset} \leftarrow z_{offset} + z_0$$

Since $\vec{a}_0 \cdot \vec{x}_0 = 0$, the maximum of $f_0(\vec{x})$ is at least 0. So $\vec{x}_0$ does not maximize the original function $f(\vec{x}) = \vec{a} \cdot \vec{x}$ if and only if there is a feasible point $\vec{x}$ so that $f_0(\vec{x}) = \vec{a}_0 \cdot \vec{x} > 0$.

STEP 4 We now have achieved the following:

(a) The augmented matrix describing the constraint has the form

$$T_0 = \begin{pmatrix} 1 & 0 & \ldots & 0 & a_{1,m+1} & \ldots & a_{1,n} & y_1 \\ 0 & 1 & \ddots & \vdots & a_{2,m+1} & \ldots & a_{2,n} & y_2 \\ \vdots & \ddots & \ddots & 0 & \vdots & & \vdots & \vdots \\ 0 & \ldots & 0 & 1 & a_{m,m+1} & \ldots & a_{n,m} & y_m \end{pmatrix}$$

where all the numbers y_i are strictly positive.

(b) A basic solution has the form

$$\vec{x}_0 = (x_1, \ldots, x_m, 0, \ldots, 0)$$

(c) We have maximize

$$f_0(\vec{x}) = \vec{a}_0 \cdot \vec{x}$$

where

$$\vec{a}_0 = (0, \ldots, 0, a_{m+1}, \ldots, a_n)$$

After we have done this, the maximum needs to be increased by z_{offset} to solve the original problem.
Since $\vec{a}_0 \cdot \vec{x}_0 = 0$, we have $f_0(\vec{x}_0) = 0$.

If all the number $a_{m+1}, \ldots, a_n$ are less than or equal to 0, then, since each feasible solution $\vec{x}$ has only non-negative entries, it follows that

$$f_0(\vec{x}) = \vec{a}_0 \cdot \vec{x} \leq 0$$

and 0 is indeed the maximum of $f_0(\vec{x})$ on the feasible set, obtained at $\vec{x}_0$. In this case, the problem is solved. and the algorithm stops. The solution is $\vec{x}_0$, and the maximum is z_{offset}.

STEP 5 Since the algorithm did not stop, one of the coordinates of

$$\vec{a}_0 = (0, \ldots, 0, a_{m+1}, \ldots, a_n)$$

We reorder the variables so that $a_{m+1} > 0$. Now consider the $(m+1)^{st}$ column of T

$$_0 = \begin{pmatrix} 1 & 0 & \ldots & 0 & a_{1,m+1} & \ldots & a_{1,n} & y_1 \\ 0 & 1 & \ddots & \vdots & a_{2,m+1} & \ldots & a_{2,n} & y_2 \\ \vdots & \ddots & \ddots & 0 & \vdots & & \vdots & \vdots \\ 0 & \ldots & 0 & 1 & a_{m,m+1} & \ldots & a_{n,m} & y_m \end{pmatrix}$$

If all entries are 0 or less, i.e. if $a_{1,m+1}, \ldots, a_{m,m+1} \leq 0$, then

$$\vec{x} = \begin{pmatrix} y_1 - r \cdot a_{1,m+1} \\ \cdots \\ y_m - r \cdot a_{m,m+1} \\ r \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

is a feasible solution for each positive $r > 0$. Moreover,

$$\begin{aligned} f_0(x) &= \vec{a}_0 \cdot \vec{x} \\ &= r a_{m+1} > 0 \end{aligned}$$

Since $\lim_{r \to \infty} r a_{m+1} = \infty$, the problem is unbounded and hence the problem has no solution. The algorithm stops.

STEP 6 Now we know that at least one of the entries $a_{1,m+1}, \ldots, a_{m,m+1}$ is strictly positive. With this knowledge, we are trying to find a feasible value of $\vec{x}$ so that $f_0(\vec{x}) = \vec{a}_0 \vec{x} > 0$. We are again trying to define $\vec{x}_1$ as

$$\vec{x}_1 = \begin{pmatrix} y_1 - r \cdot a_{1,m+1} \\ \cdots \\ y_m - r \cdot a_{m,m+1} \\ r \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

This $\vec{x}_1$ will still satisfy the constraint $A\vec{x} = \vec{b}$, but in addition, we also have to make sure that all entries are positive. So we have to pick r so that $r \geq 0$ and $y - r \cdot a_{i,m+1} \geq 0$ for all values of i between 1 and m. This

will follow automatically for those values of i for which $a_{i,m+1} \leq 0$. For the values of i for which $a_{i,m+1} > 0$ we end up with the condition that

$$\begin{aligned} y_i - r \cdot a_{i,m+1} &\geq 0 \\ y_i &\geq r \cdot a_{i,m+1} \\ r &\leq \frac{y_i}{a_{i,m+1}} \end{aligned}$$

So we can define

$$r = \min \left\{ \frac{y_i}{a_{i,m+1}} : a_{i,m+1} > 0 \right\}$$

Since all entries y_i are positive, we find that

$$r \geq 0.$$

If $r = 0$, then there is at least one index i so that $a_{i,m+1} > 0$ and $y_i = 0$.This is not desirable, because it might lead to "cycling". We have to fix this: Go to Step 7 below!. We no can assume that $a_{i,m+1} > 0$ implies that $y_i > 0$. In this case,

$$r > 0$$

$\vec{x}_1$ is a feasible point, and

$$\begin{aligned} f_0(\vec{x}_1) &= a_0 \cdot \vec{x}_1 \\ &= a_{m+1} \cdot r \\ &= \min \left\{ \frac{y_i}{a_{i,m+1}} a_{m+1} : a_{i,m+1} > 0 \right\} > 0 \end{aligned}$$

We will now verify that $\vec{x}_1$ is a basic solution. Let j be the coordinate for which the minimum in $r = \min \left\{ \frac{y_i}{a_{i,m+1}} : a_{i,m+1} > 0 \right\}$ is obtained. Then

$$\begin{aligned} r &= \frac{y_j}{a_{j,m+1}} \\ y_j - r a_{j,m+1} &= 0 \end{aligned}$$

Hence only coordinates of $\vec{x}_1$ that are possibly different from 0, are the coordinates $1, 2, \ldots, j-1, j+1, \ldots, m, m+1, 0 \ldots, 0$. The columns with numbers $1, \ldots, j-1, j+1, \ldots, m, m+1$ are linearly independent, because the $(m+1)^{st}$ column has a non-zero entry in the j^{th} component, namely $a_{j,m+1} > 0$. So $\vec{x}_1$ is a basic feasible solution, with $f(\vec{x}_1) > f(\vec{x}_0)$, and we continue with STEP 2. with the value of $\vec{x}_1$ replacing $\vec{x}_0$, the value of $\vec{a}_0$ replacing $\vec{a}$.

STEP 7 We now are in the situation that the number r formed in step 6 vanishes:

$$0 = \min \left\{ \frac{y_i}{a_{i,m+1}} : a_{i,m+1} > 0 \right\}$$

Hence there is an index i_0 so that $a_{i_0,m+1} > 0$ and $y_{i_0} = 0$.We only consider the rows with indices i for which $y_i = 0$ and $a_{i,m+1} > 0$ and use only those as constraints, keeping the same objective function. All other rows are ignored for

now. To make notations easier, we reshuffle rows and variables so that the tableaux has the following shape:

$$\begin{array}{cccccccc|c} 1 & 0 & \dots & \dots & 0 & a_{1,m+1} & \dots & a_{1,n} & 0 \\ & \ddots & & \dots & 0 & \vdots & & \vdots & 0 \\ 0 & \dots & 1 & \dots & 0 & a_{k,m+1} & \dots & a_{k,n} & 0 \\ 0 & \dots & \dots & \dots & 0 & a_{m+1} & \dots & a_n & z_{offset} \end{array}$$

We replace the last column by a new column:

$$\begin{array}{cccccccc|c} 1 & 0 & \dots & \dots & 0 & a_{1,m+1} & \dots & a_{1,n} & 1 \\ & \ddots & & \dots & 0 & \vdots & & \vdots & 1 \\ 0 & \dots & 1 & \dots & 0 & a_{k,m+1} & \dots & a_{k,n} & 1 \\ 0 & \dots & \dots & \dots & 0 & a_{m+1} & \dots & a_n & 0 \end{array}$$

and run the algorithm with the first columns (containing the canonical unit vectors) as basic variables (i.e. start at STEP 2 with this subproblem). This is a recursive call. But if not all constants on the right side are equal to 0, then each subsequent recursive call will work with a subproblem that has a strictly smaller number of constraints, eventually ending with a problem that has at most one constraint. You should convince yourself that problems with only one constraint and a non-vanishing right side are really solved by the algorithm.

The algorithm will stop on this subproblem - possibly using STEP 7 more than once. There are two ways the subproblem can stop:

(1) (a) The algorithm stops in STEP 4. All the coefficients in the objective function are now negative. If we would not use the $1's$ in the last column but only $0's$, then neither of the entries in the last column nor the entry for z_{offset} would never change during the Gauss-Jordan steps in the Simplex algorithm. We now add the other rows (the rows for which the corresponding $y_i > 0$) This means: Given the constraints, the objective function $f_0(\vec{x}) = \vec{a}_0 \cdot \vec{x}$ is equivalent to an objective function with only negative coefficients, hence the maximum cannot be larger than 0. So the maximum was already found before we entered this recursive call.

(b) The algorithm stops in STEP 5 above. So there is one column - (with column index k) that - with the exception in the row for the objective function - contains only negative entries. If we would not use the $1's$ in the last column but only $0's$, then neither of the entries in the last column nor the entry for z_{offset} would ever change during the Gauss-Jordan steps in the Simplex algorithm. We now add the other rows (the rows for which the corresponding $y_i > 0$) Since only Gauss-Jordan steps were used, this will lead to a problem that is equivalent to the original problem, Reorder the variables so that the basic variables are now in the first columns and so the column with index k becomes the $(m+1)^{st}$ column. Go back to STEP 5. If STEP 6 is reached, then - since all the entries in column $m+1$ that correspond to $y_i = 0$ are negative - the number r will be strictly positive, and the algorithm will continue.

Here is the algorithm without training wheels. Indeed, we do not have to reorder the variables at each step, and we certainly do not have to carry the explanations along.

STEP 1 Set

$$z_{offset} = 0$$

.

STEP 2 Identify the basic variables and the basic submatrix B of A. Performing the Gauss-Jordan elimination process on the augmented matrix $\left(A, \vec{b}\right)$ will transforming the basic submatrix B into the identity matrix. The result is

$$T_0 = (A_0, \vec{y}_0)$$

STEP 3 "Reduce" the function $f(\vec{x}) = \vec{a} \cdot \vec{x}$ to obtain

$$f_0(\vec{x}) = \vec{a}_0 \cdot \vec{x}$$

Let

$$z_0 = \vec{a} \cdot \vec{x}_0$$

where $\vec{x}_0$ is obtained from $\vec{y}_0$ by adding $0's$ in the non-basic variables. Let

$$\begin{aligned} \vec{a} &\leftarrow \vec{a}_0 \\ z_{offset} &\leftarrow z_{offset} + z_0 \end{aligned}$$

STEP 4 If all the number $a_{m+1}, \ldots, a_n$ are less than or equal to 0, then the problem is solved. and the algorithm stops. The solution is $\vec{x}_0$, and the maximum is z_{offset}.

STEP 5 If for each strictly positive a_k, all entries in the k^{th} row of T_0 are 0 or less, i.e. if $a_{1,k1}, \ldots, a_{m,k} \leq 0$, then the problem is unbounded and hence the problem has no solution. The algorithm stops.

STEP 6 Let k be an index so that $a_k > 0$ and $a_{i,k} > 0$ for at least one index i. Let

$$r = \min\left\{\frac{y_i}{a_{i,k}} : a_{i,k} > 0\right\}$$

If $r = 0$,go to step 7 below**!.**If $r > 0$, let j be the coordinate for which the minimum in $r = \min\left\{\frac{y_i}{a_{i,k}} : a_{i,k} > 0\right\}$ is obtained. The new basic submatrix B is obtained from removing column j and adding column k.Continue with STEP 2. with T_0 replacing the augmented matrix $\left(A, \vec{b}\right)$:

$$\begin{aligned} \left(A, \vec{b}\right) &\leftarrow T_0 \\ B &\leftarrow B \text{ (remove column } j, \text{ add column } k) \end{aligned}$$

STEP 7 We only consider the rows with index i for which $y_i = 0$ and $a_{i,k} > 0$ and use only those as constraints, keeping the same objective function. All other rows are ignored for now. We replace the last column by a new column containing only $1's$. Let B the largest submatrix of the new problem that contains only canonical unit vectors. Start the simplex algorithm on this new problem. There are two ways the subproblem can stop:

(a) The algorithm stops in STEP 4. All the coefficients in the objective function are now negative. In this case, the maximum was already found before we entered the recursive call.

(b) The algorithm stops in STEP 5. So there is one column -(with column index k_0) that - with the exception in the row for the objective function - contains only negative entries. Replace the entries of the last columns again by 0, except for the entry containing the offset, which is again set to z_{offset}.We now add the other rows (the rows for which the corresponding y_i is strictly positive.) Go back to STEP 5. If STEP 6 is reached, then make sure that the index k is set to k_0.

7. More Examples for the Simplex Algorithm

We give some additional examples. The numbering of the steps correspond to the numbers of the last section.

EXAMPLE 52. *Maximize*

$$4x_1 + 3x_2$$

subject to

$$\begin{aligned} 3x_1 + 4x_2 &\leq 12 \\ 3x_1 + 3x_2 &\leq 10 \\ 4x_1 + 2x_2 &\leq 8 \\ x_i &\geq 0 \text{ for } i = 1, 2 \end{aligned}$$

We **Convert to standard form:**
Maximize

$$4x_1 + 3x_2$$

subject to

$$\begin{aligned} 3x_1 + 4x_2 + x_3 &= 12 \\ 3x_1 + 3x_2 + x_4 &= 10 \\ 4x_1 + 2x_2 + x_5 &= 8 \\ x_i &\geq 0 \text{ for } i = 1, 2, \ldots, 5 \end{aligned}$$

Identify a basic feasible solution:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 12 \\ 10 \\ 8 \end{pmatrix}$$

STEP 1 Set

$$z_{offset} = 0$$

STEP 2 Identify the basic variables and the basic submatrix B of A : Performing the Gauss-Jordan elimination process on the augmented matrix $\left(A, \vec{b}\right)$ will transforming the basic submatrix B into the identity matrix. - (In practical examples, we also add the coefficients of the function $f(\vec{x}) = \vec{a} \cdot \vec{x}$ (i.e. the row $\vec{a}$, augmented by $\vec{a} \cdot \vec{x}_0$) as a last row to $\left(A, \vec{b}\right)$. As a result, STEP 3 becomes a part of STEP 2.) In this case, the basic submatrix consists of columns $3, 4, 5$.and the result of the Gauss-Jordan process does not change anything:

$$T_0 = (A_0, \vec{y}_0)$$

$$\left(A, \vec{b}\right) = \left(\begin{array}{ccccc|c} 3 & 4 & 1 & 0 & 0 & 12 \\ 3 & 3 & 0 & 1 & 0 & 10 \\ 4 & 2 & 0 & 0 & 1 & 8 \\ 4 & 3 & 0 & 0 & 0 & 0 \end{array}\right)$$

In this case, T_0 also equals $\left(A, \vec{b}\right)$.

STEP 3 "Reduce" the function $f(\vec{x}) = \vec{a} \cdot \vec{x}$ to obtain

$$f_0(\vec{x}) = \vec{a}_0 \cdot \vec{x}$$

Let

$$z_0 = \vec{a} \cdot \vec{x}_0$$

where $\vec{x}_0$ is obtained from $\vec{y}_0$ by adding $0's$ in the non-basic variables. Let

$$\begin{aligned} \vec{a} &\leftarrow \vec{a}_0 \\ z_{offset} &\leftarrow z_{offset} + z_0 \end{aligned}$$

In this case, nothing changes!

STEP 4 If all the number $a_{m+1}, \ldots, a_n$ are less than or equal to 0, then the problem is solved. and the algorithm stops. The solution is $\vec{x}_0$, and the maximum is z_{offset}. - In this case, the last row has strictly positive entries in the row representing $\vec{a}$, and we have to continue.

STEP 5 If for each strictly positive a_k, all entries in the k^{th} row of T_0 are 0 or less, i.e. if $a_{1,k1}, \ldots, a_{m,k} \leq 0$, then the problem is unbounded and hence the problem has no solution. The algorithm stops. - In this case, we have to continue.

STEP 6 Let k be an index so that $a_k > 0$ and $a_{i,k} > 0$ for at least one index i. Let

$$r = \min\left\{\frac{y_i}{a_{i,k}} : a_{i,k} > 0\right\}$$

If $r = 0$, then go to step 7**.** Otherwise, let j be the coordinate for which the minimum in $r = \min\left\{\frac{y_i}{a_{i,k}} : a_{i,k} > 0\right\}$ is obtained. The new basic submatrix B is obtained from removing column j and adding column k.Continue with STEP 1. with T_0 replacing the augmented matrix $\left(A, \vec{b}\right)$:

$$\begin{aligned} \left(A, \vec{b}\right) &\leftarrow T_0 \\ B &\leftarrow B \text{ (remove column } j, \text{ add column } k) \end{aligned}$$

- In this case, we can pick $k = 1$ and

$$\begin{aligned} r &= \min\left\{\frac{12}{3}, \frac{10}{3}, \frac{8}{4}\right\} \\ &= \min\left\{4, \frac{10}{3}, 2\right\} \end{aligned}$$

So the minimum occurs in the first row. That means that the column having a 1 in this row (the last column) leaves the basic submatrix, and the first column is added. The basic submatrix now contains columns $1, 3$ and 4.- Go back to step 2:

STEP 2 / STEP 3 The basic submatrix consists of columns $1, 3$ and 4. Gauss Jordan elimination leads to

$$\left(A, \vec{b}\right) = \left(\begin{array}{ccccc|c} 3 & 4 & 1 & 0 & 0 & 12 \\ 3 & 3 & 0 & 1 & 0 & 10 \\ 4 & 2 & 0 & 0 & 1 & 8 \\ 4 & 3 & 0 & 0 & 0 & 0 \end{array}\right) \rightarrow \left(\begin{array}{ccccc|c} 0 & \frac{5}{2} & 1 & 0 & -\frac{3}{4} & 6 \\ 0 & \frac{3}{2} & 0 & 1 & -\frac{3}{4} & 4 \\ 1 & \frac{1}{2} & 0 & 0 & \frac{1}{4} & 2 \\ 0 & 1 & 0 & 0 & -1 & -8 \end{array}\right)$$

In this case,

$$T_0 = \left(\begin{array}{ccccc|c} 0 & \frac{5}{2} & 1 & 0 & -\frac{3}{4} & 6 \\ 0 & \frac{3}{2} & 0 & 1 & -\frac{3}{4} & 4 \\ 1 & \frac{1}{2} & 0 & 0 & \frac{1}{4} & 2 \\ 0 & 1 & 0 & 0 & -1 & -8 \end{array}\right)$$

STEP 4 If all the number $a_{m+1}, \ldots, a_n$ are less than or equal to 0, then the problem is solved. and the algorithm stops. The solution is $\vec{x}_0$, and the maximum is z_{offset}. - In this case, the last row has one strictly positive entry, and we have to continue.

STEP 5 If for each strictly positive a_k, all entries in the k^{th} row of T_0 are 0 or less, i.e. if $a_{1,k1}, \ldots, a_{m,k} \leq 0$, then the problem is unbounded and hence the problem has no solution. The algorithm stops. - In this case, we have to continue.

STEP 6 Let k be an index so that $a_k > 0$ and $a_{i,k} > 0$ for at least one index i. Let

$$r = \min\left\{\frac{y_i}{a_{i,k}} : a_{i,k} > 0\right\}$$

If $r = 0$ go to step 7. Otherwise, let j be the coordinate for which the minimum in $r = \min\left\{\frac{y_i}{a_{i,k}} : a_{i,k} > 0\right\}$ is obtained. The new basic submatrix B is obtained from removing column j and adding column k.Continue with STEP 1. with T_0 replacing the augmented matrix $\left(A, \vec{b}\right)$:

$$\begin{array}{rcl} \left(A, \vec{b}\right) & \leftarrow & T_0 \\ B & \leftarrow & B \text{ (remove column } j, \text{ add column } k) \end{array}$$

- In this case, we can pick $k = 2$ and

$$\begin{array}{rcl} r & = & \min\left\{\frac{6}{\frac{5}{2}}, \frac{4}{\frac{3}{2}}, \frac{2}{\frac{1}{2}}\right\} \\ & = & \min\left\{\frac{12}{5}, \frac{8}{3}, 4\right\} = \frac{12}{5} \end{array}$$

So the minimum occurs in the first row. That means that the column having a 1 in this row (the third column) leaves the basic submatrix, and the second column is added. The basic submatrix now contains columns $1, 2$ and 4.- Go back to step 2:

STEP 2 / STEP 3 The basic submatrix consists of columns $1, 2$ and 4. Gauss Jordan elimination leads to

$$\left(A, \vec{b}\right) = \left(\begin{array}{ccccc|c} 0 & \frac{5}{2} & 1 & 0 & -\frac{3}{4} & 6 \\ 0 & \frac{3}{2} & 0 & 1 & -\frac{3}{4} & 4 \\ 1 & \frac{1}{2} & 0 & 0 & \frac{1}{4} & 2 \\ 0 & 1 & 0 & 0 & -1 & -8 \end{array}\right) \rightarrow \left(\begin{array}{ccccc|c} 0 & 1 & \frac{2}{5} & 0 & -\frac{3}{10} & \frac{12}{5} \\ 0 & 0 & -\frac{3}{5} & 1 & -\frac{3}{10} & \frac{2}{5} \\ 1 & 0 & -\frac{1}{5} & 0 & \frac{2}{5} & \frac{4}{5} \\ 0 & 0 & -\frac{2}{5} & 0 & -\frac{7}{10} & -\frac{52}{5} \end{array}\right)$$

In this case,

$$T_0 = \left(\begin{array}{ccccc|c} 0 & 1 & \frac{2}{5} & 0 & -\frac{3}{10} & \frac{12}{5} \\ 0 & 0 & -\frac{3}{5} & 1 & -\frac{3}{10} & \frac{2}{5} \\ 1 & 0 & -\frac{1}{5} & 0 & \frac{2}{5} & \frac{4}{5} \\ 0 & 0 & -\frac{2}{5} & 0 & -\frac{7}{10} & -\frac{52}{5} \end{array}\right)$$

STEP 4 If all the number $a_{m+1}, \ldots, a_n$ are less than or equal to 0, then the problem is solved. and the algorithm stops. The solution is $\vec{x}_0$, and the maximum is z_{offset}. - In this case, we stop, and the maximum is $z_{offset} = \frac{52}{5}$, achieved at $x_1 = \frac{4}{5}$, $x_2 = \frac{12}{5}, x_3 = 0, x_4 = \frac{2}{5}$ and $x_5 = 0$.

EXAMPLE 53. *Maximize*

$$f(w, x, y, z) = w + x + y + 8z$$

subject to

$$\begin{aligned} w + z &= 9 \\ x + 2z &= 4 \\ y + 4z &= 1 \end{aligned}$$

We write the matrix for the problem:

$$\begin{array}{cccc|c} 1 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & 2 & 4 \\ 0 & 0 & 1 & 4 & 1 \\ 1 & 1 & 1 & 8 & 0 \end{array}$$

We add a first row with indicators showing where the basic columns are:

$$\begin{array}{cccc|c} \downarrow & \downarrow & \downarrow & & \\ 1 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & 2 & 4 \\ 0 & 0 & 1 & 4 & 1 \\ 1 & 1 & 1 & 8 & 0 \end{array} \quad \begin{array}{c} \\ \\ \\ \\ \\ \rightarrow \end{array} \quad \begin{array}{cccc|c} \downarrow & \downarrow & \downarrow & & \\ 1 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & 2 & 4 \\ 0 & 0 & 1 & 4 & 1 \\ 0 & 0 & 0 & 1 & -8 \end{array}$$

There is only one strictly positive entry in the last column, and this column has to enter the list of basic columns. In order to decide which column has to leave this list, we add a new column containing the fractions "right side / coefficient in the new column":

$$\begin{array}{cccc|cc} \downarrow & \downarrow & \downarrow & & & \\ 1 & 0 & 0 & 1 & 3 & 3/1 = 3 \\ 0 & 1 & 0 & 2 & 4 & 4/1 = 4 \\ 0 & 0 & 1 & 4 & 1 & 1/4 = 0.25 \\ 0 & 0 & 0 & 1 & -8 & \end{array}$$

The row that contains the minimum of all those quotients will be the row containing the new pivot element. Only those entries compete for the minimum that are positive. In our case, this will be row No., 3:

$$\begin{array}{cccc|c} \downarrow & \downarrow & \downarrow & & \\ 1 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & 2 & 4 \\ 0 & 0 & 1 & 4 & 1 \\ 0 & 0 & 0 & 1 & -8 \end{array} \quad \begin{array}{c} \\ \\ \\ \\ \\ \rightarrow \end{array} \quad \begin{array}{cccc|c} \downarrow & \downarrow & & \downarrow & \\ 1 & 0 & -\frac{1}{4} & 0 & \frac{11}{4} \\ 0 & 1 & -\frac{1}{2} & 0 & \frac{7}{2} \\ 0 & 0 & \frac{1}{4} & 1 & \frac{1}{4} \\ 0 & 0 & -\frac{1}{4} & 0 & -\frac{33}{4} \end{array}$$

We now find only negative entries. The algorithm stops, and the maximum is $\frac{33}{4}$, obtained at $w = \frac{11}{4}, x = \frac{7}{2}, y = 0$ and $z = \frac{1}{4}$.

EXAMPLE 54. *Maximize*

$$f(w, x, y, z) = y$$

subject to

$$\begin{aligned} w + y - z &= 0 \\ x - y + 2z &= 1 \end{aligned}$$

The tableaux of the problem has the following entries:

$$\begin{array}{cccc|c} 1 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 2 & 1 \\ 0 & 0 & 1 & 0 & \end{array}$$

The first two columns are used as basic columns:

$$\begin{array}{cccc|c} \downarrow & \downarrow & & & \\ 1 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 2 & 1 \\ 0 & 0 & 1 & 0 & \end{array}$$

The third column is the only column that has a positive entry. Unfortunately, the right side is 0 for the first row, so we have to use a recursive call to the simplex algorithm. The only row having 0 on the right side is the first row. For this row, we introduce an artificial right side different from 0 and work only with the first constraint (with the new right side) and the same objective function:

$$\begin{array}{cccc|cc} \downarrow & \downarrow & & & & \\ 1 & 0 & 1 & -1 & 0 & 1 \\ 0 & 1 & 0 & 2 & 1 & \\ 0 & 0 & 1 & 0 & & \end{array} \rightarrow \begin{array}{cccc|cc} 0 & \downarrow & \downarrow & & & \\ 1 & 0 & 1 & -1 & 0 & 1 \\ 0 & 1 & 0 & 2 & 1 & \\ -1 & 0 & 0 & 1 & & -1 \end{array}$$

The simplex algorithm stops here, because the only positive entry in the last row is in the forth column, and all other entries in the forth column (not counting the second row) are negative. We now delete the new artificial right side and continue with the original problem:

$$\begin{array}{cccc|c} 0 & \downarrow & \downarrow & & \\ 1 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 2 & 1 \\ -1 & 0 & 0 & 1 & \end{array} \rightarrow \begin{array}{cccc|c} 0 & & \downarrow & \downarrow & \\ 1 & \frac{1}{2} & 1 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 & 1 & \frac{1}{2} \\ -1 & -\frac{1}{2} & 0 & 0 & -\frac{1}{2} \end{array}$$

The algorithm stops here. The maximum is $\frac{1}{2}$, obtained at $(w, x, y, z) = \left(0, 0, \frac{1}{2}, \frac{1}{2}\right)$.

EXAMPLE 55. *Maximize*

$$f(v, w, x, y, z,) = v + 3w + 3z$$

subject to

$$\begin{aligned} v + x + y + z &= 1 \\ v + w + y &= 0 \\ x - y + 2z &= 1 \end{aligned}$$

Again, we start with the tableaux:

$$\begin{array}{ccccc|c} 1 & & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 2 & 1 \\ 1 & 3 & 0 & 0 & 3 & \end{array}$$

We decide to make the first three columns the basic column, and check whether this works:

$$\begin{array}{ccccc|c} \downarrow & \downarrow & \downarrow & & & \\ 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 2 & 1 \\ 1 & 3 & 0 & 0 & 3 & \end{array} \quad \rightarrow \quad \begin{array}{ccccc|c} \downarrow & \downarrow & \downarrow & & & \\ 1 & 0 & 0 & 2 & -1 & 0 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 2 & 1 \\ 0 & 0 & 0 & 1 & 1 & \end{array}$$

The algorithm would now either try to make the forth or fifth column a basic column. Because of the $0's$ in the right side, neither of those would work. So we add an artificial right side of 1 and continue working only with the first two constraints:

$$\begin{array}{ccccc|cc} \downarrow & \downarrow & \downarrow & & & & \\ 1 & 0 & 0 & 2 & -1 & 0 & 1 \\ 0 & 1 & 0 & -1 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 & 2 & 1 & \\ 0 & 0 & 0 & 1 & 1 & & \end{array} \quad \rightarrow \quad \begin{array}{ccccc|cc} & \downarrow & \downarrow & \downarrow & & & \\ \frac{1}{2} & 0 & 0 & 1 & -\frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 1 & 0 & 0 & \frac{1}{2} & 0 & \frac{3}{2} \\ 0 & 0 & 1 & -1 & 2 & 1 & \\ -\frac{1}{2} & 0 & 0 & 0 & \frac{3}{2} & & -\frac{1}{2} \end{array}$$

$$\rightarrow \quad \begin{array}{ccccc|cc} & & \downarrow & \downarrow & \downarrow & & \\ 1 & 1 & 0 & 1 & 0 & 0 & 2 \\ 1 & 2 & 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -1 & 2 & 1 & \\ -2 & -3 & 0 & 0 & 0 & & -5 \end{array}$$

This is where the Simplex Algorithm, performed only with the first two constraints, would stop, because now all the entries in the last row are 0 or less. The maximum is found. Actually, since the offset for the original problem will not changed, we also had found the maximum of the original problem before we entered the recursive call: The maximum is 0, obtained at $v = w = 0$, $x = 1$.

8. The Two-Phase Method

In order to find a feasible basic solution that can be used as a starting point for the simplex algorithm, we use the following procedure.

Suppose that we are given the following problem:

Maximize

$$f(\vec{x}) = \vec{a} \cdot \vec{x}$$

subject to

$$\begin{aligned} A\vec{x} &= \vec{b} \\ \vec{x} &\geq 0 \end{aligned}$$

where A has m linearly independent rows and where $\vec{b}$ is positive. Assume that

$$\vec{x} = (x_1, \ldots, x_n)$$

Let I be the identity matrix with m rows and m columns. We add a I to the columns of A, and obtain

$$A_0 = (A, I)$$

We also add m new variables $\vec{x}_1 = (x_{n+1}, \ldots, x_{n+m})$. Then we consider the problem:

Maximize

$$g(\vec{x}, \vec{x}_1) = -x_{n+1} - \cdots - x_{n+m}$$

subject to

$$\begin{aligned} (A, I) \begin{pmatrix} \vec{x} \\ \vec{x}_1 \end{pmatrix} &= \vec{b} \\ \begin{pmatrix} \vec{x} \\ \vec{x}_1 \end{pmatrix} &\geq \begin{pmatrix} \vec{0} \\ \vec{0} \end{pmatrix} \end{aligned}$$

Clearly, the function is bounded by 0 on the feasible set. Any basic feasible solution that gives any of the variables $x_{n+1}, \ldots, x_{n+m}$ a strictly positive value will give the function $g(\vec{x}, \vec{x}_1)$ a strictly negative value. Hence there is a basic feasible solution of $A\vec{x} = \vec{b}$ if and only the maximum of g on its feasible set is equal to 0.

Starting with the basic feasible solution $(\vec{x}, \vec{x}_1) = \left(\vec{0}, \vec{b}\right)$, we use the simplex algorithm to find the maximum of $g(\vec{x}, \vec{x}_1)$ subject to $(A, I) \begin{pmatrix} \vec{x} \\ \vec{x}_1 \end{pmatrix} = \vec{b}$, $\vec{x} \geq \vec{0}$, $\vec{x}_1 \geq 0$. If the maximum is strictly negative, then $A\vec{x} = \vec{b}$, $\vec{x} \geq 0$ has no solution. If the maximum is equal to 0, then there is a basic feasible solution $(\vec{x}, \vec{x}_1)$ so that $g(\vec{x}, \vec{x}_1) = 0$. Some of the variables of $\vec{x}_1$ might still occur in this basic feasible solution. However, they will have the value 0. Remove the corresponding columns from the basic submatrix. The remaining columns are still independent and will constitute a submatrix of A. Complete those columns to a maximal linearly independent set of columns of A. We will have found a basic submatrix that leads to a basic feasible solution.

EXAMPLE 56. *Use the above method to find a basic feasible solution of*

$$\begin{pmatrix} 1 & 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & -1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & -1 \end{pmatrix} \vec{x} = \begin{pmatrix} 2 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

We add artificial variables $\vec{y} = [y_1, y_2, y_4, y_5]^T$ and try solve the system

$$\text{Maximize } -y_1 - y_2 - y_3 - y_4$$

$$\text{subject to } \begin{pmatrix} 1 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{bmatrix} \vec{x} \\ \vec{y} \end{bmatrix} = \begin{pmatrix} 2 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

The simplex algorithm leads to the following steps:

$$\left(\begin{array}{ccccccccccc|c} 1 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 2 \\ 1 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 2 \\ 0 & 1 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 3 \\ 1 & 1 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 4 \\ \hline \overline{0} & \overline{0} & \overline{0} & \overline{0} & \overline{0} & \overline{0} & \overline{0} & \overline{-1} & \overline{-1} & \overline{-1} & \overline{-1} & \overline{\alpha} \end{array}\right) \rightarrow$$

$$\rightarrow \left(\begin{array}{ccccccccccc|c} 1 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 2 \\ 1 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 2 \\ 0 & 1 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 3 \\ 1 & 1 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 4 \\ \hline \overline{3} & \overline{3} & \overline{3} & \overline{-1} & \overline{-1} & \overline{-1} & \overline{-1} & \overline{0} & \overline{0} & \overline{0} & \overline{0} & \overline{\alpha + 11} \end{array}\right)$$

$$\rightarrow \left(\begin{array}{ccccccccccc|c} 1 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 2 \\ 0 & -1 & 1 & 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 1 & 0 & 0 & -1 & -1 & 0 & 0 & 1 & 2 \\ \hline \overline{0} & \overline{0} & \overline{3} & \overline{2} & \overline{-1} & \overline{-1} & \overline{-1} & \overline{-3} & \overline{0} & \overline{0} & \overline{0} & \overline{\alpha + 5} \end{array}\right)$$

$$\rightarrow \left(\begin{array}{ccccccccccc|c} 1 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 2 \\ 0 & -1 & 1 & 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & -1 & 1 & -1 & 0 & 1 & -1 & 1 & 0 & 3 \\ 0 & 1 & 0 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 2 \\ \hline \overline{0} & \overline{3} & \overline{0} & \overline{-1} & \overline{2} & \overline{-1} & \overline{-1} & \overline{0} & \overline{-3} & \overline{0} & \overline{0} & \overline{\alpha + 5} \end{array}\right)$$

$$\rightarrow \left(\begin{array}{ccccccccccc|c} 1 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 1 & 1 & 0 & 0 & -1 & -1 & 0 & 0 & 1 & 2 \\ 0 & 1 & 0 & -1 & 0 & -1 & 1 & 1 & 0 & 1 & -1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 2 \\ \hline \overline{0} & \overline{1} & \overline{0} & \overline{-1} & \overline{0} & \overline{-1} & \overline{1} & \overline{0} & \overline{-1} & \overline{0} & \overline{-1} & \overline{\alpha + 1} \end{array}\right)$$

$$\rightarrow \left(\begin{array}{ccccccccccc|c} 1 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & -1 & 2 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & -1 & -1 & 0 & 0 & 1 & 2 \\ 0 & 1 & 0 & -1 & 0 & -1 & 1 & 1 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & -2 & -1 & -1 & -1 & -2 & 1 \\ \hline \overline{0} & \overline{0} & \overline{0} & \overline{0} & \overline{0} & \overline{0} & \overline{0} & \overline{-1} & \overline{-1} & \overline{-1} & \overline{0} & \overline{\alpha} \end{array}\right)$$

This is where the algorithm stops. The new basic columns are now the columns Nos 1, 2, 3 and 5. We find that

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

We can check that this is indeed a feasible solution.

9. Duality

Consider the linear programming problem

$$\text{Maximize } f(\vec{x}) = \vec{a} \cdot \vec{x}$$

subject to

$$\begin{aligned} A\vec{x} &= \vec{b} \\ \vec{x} &\geq 0 \end{aligned}$$

where $\vec{b} \geq 0$. If we drop the requirement that $\vec{b} \geq 0$, then we can rewrite this problem in each of the following forms

$$\text{Maximize } f(\vec{x}) = \vec{a} \cdot \vec{x}$$

subject to

$$\begin{aligned} A\vec{x} &\leq \vec{b} \\ \vec{x} &\geq 0 \end{aligned}$$

or

$$\text{Minimize } f(\vec{x}) = \vec{a} \cdot \vec{x}$$

subject to

$$\begin{aligned} A\vec{x} &\geq \vec{b} \\ \vec{x} &\geq 0 \end{aligned}$$

For example, instead of maximizing $\vec{a} \cdot \vec{x}$ we could minimize $(-\vec{a}) \cdot \vec{x}$. Furthermore, the equation $A\vec{x} = \vec{b}$ can be replaced by the two inequalities $A\vec{x} \leq \vec{b}$ and $(-A)\,\vec{x} \leq \left(-\vec{b}\right)$ or the two inequalities $A\vec{x} \geq \vec{b}$ and $(-A)\,\vec{x} \geq \left(-\vec{b}\right)$.

The dual of a problem can now be defined as follows:

DEFINITION 24. *Let A be a matrix with m rows and n columns, let $\vec{b} \in \Re^m$ and let $\vec{a} \in \Re^n$. Then the dual of the linear programming problem*

$$\textit{Maximize } \vec{a} \cdot \vec{x}$$

subject to

$$\begin{aligned} A\vec{x} &\leq \vec{b} \\ \vec{x} &\geq 0 \end{aligned}$$

is the linear programming problem

$$\textit{Minimize } \vec{b} \cdot \vec{y}$$

subject to

$$\begin{aligned} A^T\vec{y} &\geq \vec{a} \\ \vec{y} &\geq 0 \end{aligned}$$

The original problem is also called the primal problem.

So in order to find the dual of a problem, we first have to rewrite the original (primal) problem in the right form. This fact will be used in the proof of the following theorem:

THEOREM 42. *The dual of the problem*

$$\textit{Minimize } \vec{a} \cdot \vec{x}$$

subject to

$$\begin{aligned} A\vec{x} &\geq \vec{b} \\ \vec{x} &\geq 0 \end{aligned}$$

is the problem

$$\textit{Maximize } \vec{b} \cdot \vec{y}$$

subject to

$$\begin{aligned} A^T\vec{y} &\leq \vec{a} \\ \vec{y} &\geq 0 \end{aligned}$$

PROOF. In order to find the dual problem, we first have to rewrite the original problem in the form

$$\text{Maximize } (-\vec{a}) \cdot \vec{x}$$

subject to

$$\begin{aligned} (-A)\,\vec{x} &\leq \left(-\vec{b}\right) \\ \vec{x} &\geq 0 \end{aligned}$$

The dual problem is now

$$\text{Minimize } \left(-\vec{b}\right) \cdot \vec{y}$$

subject to

$$\begin{aligned} (-A)^T\,\vec{y} &\geq (-\vec{a}) \\ \vec{y} &\geq 0 \end{aligned}$$

$$\text{Maximize } \vec{b} \cdot \vec{y}$$

subject to

$$\begin{aligned} A^T\vec{y} &\leq \vec{a} \\ \vec{y} &\geq 0 \end{aligned}$$

But this is equivalent to the problem □

THEOREM 43. *Consider the problem*

$$\textit{Maximize } \vec{a} \cdot \vec{x}$$

subject to

$$\begin{aligned} A\vec{x} &\leq \vec{b} \\ \vec{x} &\geq 0 \end{aligned}$$

Then the dual of the dual problem is the primal (original) problem.

PROOF. This follows easily from the previous theorem:.... □

EXAMPLE 57. *Given the problem*

$$\textit{Minimize } x+y+z$$

subject to

$$\begin{aligned} x+y &\geq 2 \\ y+z &\geq 3 \end{aligned}$$

find the dual problem and solve both problems.

Solution. In matrix form, the problem has the form

$$\text{Minimize } [1,1,1]\cdot[x,y,z]$$

subject to

$$\begin{aligned} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} &\geq \begin{bmatrix} 2 \\ 3 \end{bmatrix} \\ x,y,z &\geq 0 \end{aligned}$$

So the dual problem is

$$\text{Maximize } [2,3]\cdot[u,v]$$

subject to

$$\begin{aligned} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} &\leq \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\ u,v &\geq 0 \end{aligned}$$

We have to write the problems in standard form using slack variables. The standard form of the primal problem is

$$\text{Maximize } -x-y-z$$

subject to

$$\begin{aligned} \begin{bmatrix} 1 & 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ a \\ b \end{bmatrix} &= \begin{bmatrix} 2 \\ 3 \end{bmatrix} \\ x,y,z,a,b &\geq 0 \end{aligned}$$

The augmented matrix is

$$\left[\begin{array}{ccccc|c} 1 & 1 & 0 & -1 & 0 & 2 \\ 0 & 1 & 1 & 0 & -1 & 3 \\ -1 & -1 & -1 & 0 & 0 & \alpha \end{array}\right]$$

We use the first columns as initial basic variables, and then go through the simplex method:

$$\left[\begin{array}{ccccc|c} \downarrow & & \downarrow & & & \\ 1 & 1 & 0 & -1 & 0 & 2 \\ 0 & 1 & 1 & 0 & -1 & 3 \\ -1 & -1 & -1 & 0 & 0 & \alpha \end{array}\right] \rightarrow \left[\begin{array}{ccccc|c} \downarrow & & \downarrow & & & \\ 1 & 1 & 0 & -1 & 0 & 2 \\ 0 & 1 & 1 & 0 & -1 & 3 \\ 0 & 1 & 0 & -1 & -1 & \alpha+5 \end{array}\right]$$

$$\rightarrow \left[\begin{array}{ccccc|c} & \downarrow & \downarrow & & & \\ 1 & 1 & 0 & -1 & 0 & 2 \\ -1 & 0 & 1 & 1 & -1 & 1 \\ -1 & 0 & 0 & 0 & -1 & \alpha+3 \end{array}\right]$$

So the maximum is -3, obtained at $x = 0$, $y = 2$, $z = 1$ Hence the minimum of $x + y + z$ subject to the constraints is 3, obtained at $(0, 2, 1)$.
The standard form of the dual problem is

$$\text{Maximize } [2, 3] \cdot [u, v]$$

subject to

$$\left[\begin{array}{ccccc} 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{array}\right] \left[\begin{array}{c} u \\ v \\ c \\ d \\ e \end{array}\right] = \left[\begin{array}{c} 1 \\ 1 \\ 1 \end{array}\right]$$

$$a, b, c, d, e \geq 0$$

This time, we start the simplex algorithm with the last three columns:

$$\left[\begin{array}{ccccc|c} & & \downarrow & \downarrow & \downarrow & \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 2 & 3 & 0 & 0 & 0 & \alpha \end{array}\right] \rightarrow \left[\begin{array}{ccccc|c} \downarrow & & & \downarrow & \downarrow & \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 3 & -2 & 0 & 0 & \alpha-2 \end{array}\right]$$

Here we are stuck, because one entry on the right side of the equations is 0. We have to use a recursive call:

$$\left[\begin{array}{ccccc|c} \downarrow & & & \downarrow & \downarrow & \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 3 & -2 & 0 & 0 & \alpha-2 \end{array}\right] \rightarrow \left[\begin{array}{ccccc|c} \downarrow & \downarrow & & & \downarrow & \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -3 & 0 & \alpha-2 \end{array}\right]$$

The only positive entry in the last row is 1, and in the subproblem, all others in the column no. 3 are negative (rows number 1 and 3 are irrelevant for the subproblem!

So we now can continue:

$$\left[\begin{array}{ccccc|c} \downarrow & \downarrow & & & \downarrow & \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -3 & 0 & \alpha-2 \end{array}\right] \rightarrow \left[\begin{array}{ccccc|c} \downarrow & \downarrow & & & \downarrow & \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1 & 1 \\ 0 & 0 & 1 & -3 & 0 & \alpha-2 \end{array}\right]$$

$$\rightarrow \left[\begin{array}{ccccc|c} & \downarrow & \downarrow & & \downarrow & \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ -1 & 0 & 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & -3 & 0 & \alpha-3 \end{array}\right]$$

Here the algorithm stops. The maximum is 3 obtained at $[u, v, c, d, e] = [0, 1, 1, 0, 0]$. So the maximum of the dual problem is 3, obtained at $[u, v] = [0, 1]$

We observe that both the problem and its dual have the same optimal value, namely 3. Also, if we look at the last row in the tableaux of the simplex algorithm, then we obtain

$$\begin{array}{ccccc|c} -1 & 0 & 0 & 0 & -1 & 3 \end{array} \text{ for the primal problem}$$

$$\begin{array}{ccccc|c} -1 & 0 & 0 & -3 & 0 & -3 \end{array} \text{ for the dual problem}$$

For the primal problem, the entries corresponding to slack variables are $[0, -1] = -[0, 1]$. Note that this is the negative of the solution for the dual problem.

For the dual problem, the entries corresponding to slack variables would by $[0, -3, 0] = -[0, 3, 0]$. Note that $[0, 3, 0]$ would also be a solution for the primal problem.

THEOREM 44. *If $\vec{x}$ and $\vec{y}$ satisfy*

$$\begin{aligned} A\vec{x} &\leq \vec{b} \\ \vec{x} &\geq 0 \end{aligned}$$

and

$$\begin{aligned} A^T\vec{y} &\geq \vec{a} \\ \vec{y} &\geq \vec{0} \end{aligned}$$

then

$$\vec{a} \cdot \vec{x} \leq \vec{b} \cdot \vec{y}$$

PROOF. Since $A\vec{x} \leq \vec{b}$, and since $\vec{y}$ has only positive entries,using the rules of matrix multiplication we find that

$$\begin{aligned} \vec{b} \cdot \vec{y} &\geq (A\vec{x}) \cdot \vec{y} \\ &= (A\vec{x})^T \vec{y} \\ &= \left(\vec{x}^T A^T\right) \vec{y} \\ &= \vec{x}^T \left(A^T \vec{y}\right) \\ &= \vec{x} \cdot \left(A^T \vec{y}\right) \end{aligned}$$

Since $A^T\vec{y} \geq \vec{a}$ and since $\vec{x}$ has only positive entries, we can continue:

$$\vec{x} \cdot \left(A^T \vec{y}\right) \geq \vec{x} \cdot \vec{a}$$

Therefore, $\vec{b} \cdot \vec{y} \geq \vec{a} \cdot \vec{x}$. □

This theorem shows that the minimum of $\vec{b} \cdot \vec{y}$ over all feasible solutions of $A^T \vec{y} \geq \vec{a}, \vec{y} \geq \vec{0}$ is dominated by the maximum of $\vec{a} \cdot \vec{x}$ over all feasible solutions of $A\vec{x} \leq \vec{b}, \vec{x} \geq \vec{0}$. We will now show that this maximum and this minimum are actually the same number. During the proof of this statement, we will also discover how to translate a solution of one problem into a solution of the dual problem. The proof will use the simplex algorithm.

THEOREM 45. *The solution of*

$$\textit{Minimize } f(\vec{x}) = \vec{a} \cdot \vec{x}$$

subject to

$$\begin{aligned} A\vec{x} &\geq \vec{b} \\ \vec{x} &\geq \vec{0} \end{aligned}$$

and the dual problem

$$\textit{Maximize } g(\vec{y}) = \vec{b} \cdot \vec{x}$$

subject to

$$\begin{aligned} A^T \vec{y} &\leq \vec{a} \\ \vec{x} &\geq \vec{0} \end{aligned}$$

have the same value. If we write the primal problem in standard form and use the simplex algorithm, then a feasible solution of the dual problem can be found in the columns representing the slack variables.

PROOF. First, we have to write the problem

$$\text{Maximize } f(\vec{x}) = \vec{a} \cdot \vec{x}$$

subject to

$$\begin{aligned} A\vec{x} &\leq \vec{b} \\ \vec{x} &\geq \vec{0} \end{aligned}$$

in standard form. We need to introduce slack variables. Assume that A has m rows. For each row, we have to add a slack variable ξ_i. So let I be the unit matrix with m rows, and let

$$\vec{\xi} = \left[\vec{\xi}_1, \ldots, \vec{\xi}_m\right]$$

the vector of slack variables. If all the entries of $\vec{b}$ would be positive, then the standard form of the primal problem would be

$$\text{Maximize } f\left(\vec{x}, \vec{\xi}\right) = \vec{a} \cdot \vec{x}$$

subject to

$$\begin{aligned} [A, I] \begin{bmatrix} \vec{x} \\ \vec{\xi} \end{bmatrix} &= \vec{b} \\ \vec{x} &\geq \vec{0}, \vec{\xi} \geq \vec{0} \end{aligned}$$

If not all the entries $\vec{b}$ are positive, we would have to multiply certain equations by -1. This can be achieved by multiplying by the diagonal matrix

$$D = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & d_m \end{bmatrix}$$

where

$$d_i = \begin{cases} 1 & \text{if } b_i \geq 0 \\ -1 & \text{if } b_i < 0 \end{cases}$$

So the standard form is

$$\text{Maximize } f\left(\vec{x}, \vec{\xi}\right) = \vec{a} \cdot \vec{x}$$

subject to

$$\begin{aligned} D\left[A, I\right] \begin{bmatrix} \vec{x} \\ \vec{\xi} \end{bmatrix} &= D\vec{b} \\ \vec{x} &\geq \vec{0}, \vec{\xi} \geq \vec{0} \end{aligned}$$

Since $D\left[A, I\right] = \left[DA, D\right]$, we can rewrite the constraint as

$$\begin{aligned} \left[DA, DI\right] \begin{bmatrix} \vec{x} \\ \vec{\xi} \end{bmatrix} &= D\vec{b} \\ \vec{x} &\geq \vec{0}, \vec{\xi} \geq \vec{0} \end{aligned}$$

So the initial tableaux of for the simplex algorithm has the following form:

$$\left[\begin{array}{cc|c} DA & D & D\vec{b} \\ \vec{a}^T & \vec{0} & \alpha \end{array}\right]$$

We now start the simplex algorithm. Each step is obtained from adding a multiple of one row to another row. So each row of the final tableaux of the simplex algorithm is obtained as linear combination of rows of the initial tableaux. We write the coefficients of those linear combinations into the rows of a $(m+1) \times (m+1)$ matrix T. Then the final tableau of the simplex algorithm is obtained as matrix multiplication by T :

$$T \cdot \begin{bmatrix} DA & D & D\vec{b} \\ \vec{a}^T & \vec{0}^T & \alpha \end{bmatrix} = \begin{bmatrix} A_{final} & D_{final} & \left(D\vec{b}\right)_{final} \\ \vec{a}_{final}^T & \vec{c}_{final}^T & \alpha + r \end{bmatrix}$$

Since during the simplex algorithm, we never add the last row (the row which contains the objective function) to any of the other rows, the matrix T has a special form. We claim that the last row is of the form $\left(\vec{\sigma}^T, 1\right)$, where $\vec{\sigma}$ is a column vector. Indeed, multiplying by the last row of T should be the result of adding a linear combination of the first m rows to the last row, and this is exactly achieved by a row of the form $\left(\sigma^T, 1\right)$.

Also, the last column of T has the form $\begin{bmatrix} \vec{0} \\ 1 \end{bmatrix}$. An entry different from 0 in the last column means that we are adding a multiple of the last row to any of the

other rows, which we never do (except for the last row). Hence T has the form

$$T = \begin{bmatrix} T_0 & \vec{0} \\ \vec{\sigma}^T & 1 \end{bmatrix}$$

So the end product of the simplex algorithm is given by

$$\begin{aligned} \begin{bmatrix} A_{final} & D_{final} & \left(D\vec{b}\right)_{final} \\ \vec{a}_{final}^T & \vec{c}_{final}^T & \alpha + r \end{bmatrix} &= T \cdot \begin{bmatrix} DA & D & D\vec{b} \\ \vec{a}^T & \vec{0}^T & \alpha \end{bmatrix} \\ &= \begin{bmatrix} T_0 & \vec{0} \\ \vec{\sigma}^T & 1 \end{bmatrix} \begin{bmatrix} DA & D & D\vec{b} \\ \vec{a}^T & \vec{0}^T & \alpha \end{bmatrix} \\ &= \begin{bmatrix} T_0 DA & T_0 D & T_0 D\vec{b} \\ \sigma^T DA + \vec{a}^T & \sigma^T D & \alpha + \sigma^T D\vec{b} \end{bmatrix} \end{aligned}$$

The simplex algorithm stops, because the last row (with the possible exception for the entry in the last column) contains only negative entries:

$$\begin{aligned} \sigma^T DA + \vec{a}^T &\leq \vec{0}^T \\ \sigma^T D &\leq \vec{0}^T \end{aligned}$$

Also, during the whole execution of the simplex algorithm, we make sure that the right side (last column) is kept positive:

$$T_0 D\vec{b} \geq \vec{0}$$

Finally, the maximal solution of the standard form of the primal problem is given by $-\sigma^T D\vec{b}$.

We now define

$$\begin{aligned} \vec{y} &= -\left(\sigma^T D\right)^T \\ &= -D^T\sigma = -D\sigma \end{aligned}$$

Then

$$\vec{y} \geq \vec{0}$$

and

$$\begin{aligned} \sigma^T DA + \vec{a}^T &\leq \vec{0}^T \\ \vec{a}^T &\leq -\sigma^T DA \\ \vec{a} &\leq -\left(\sigma^T DA\right)^T \\ \vec{a} &\leq A^T\left(-\sigma^T D\right)^T \\ \vec{a} &\leq A^T y \end{aligned}$$

So $\vec{y}$ is a feasible solution for the dual problem. Moreover, since $-\sigma^T D\vec{b} = \vec{y} \cdot \vec{b}$, and since $-\sigma^T D\vec{b}$ is the maximal solution of the primal problem. $\vec{y} \cdot \vec{b}$ gives the same value as the maximal solution of the primal problem. Since the maximal solution of the primal problem is greater than or equal to the minimal solution of the dual problem, it follows that the maximal solution of the primal problem and the minimal solution of the dual problem yield the same value, namely $-\sigma^T D\vec{b}$. Hence $\vec{y}$ yields to the minimal solution of the dual problem $\square$