

---

# Static optimization

---

**Keywords:** local and global extremal points, unconstrained optimization, constrained optimization, Lagrange function, Lagrange multipliers, Lagrange method, Kuhn-Tucker method

**Compare:** Vorlesungen Mathematik 1 und Mathematik 2

## Contents

<b>1</b>	<b>Overview about (static) optimization problems</b>	<b>2</b>
<b>2</b>	<b>Unconstrained optimization problems</b>	<b>3</b>
2.1	Local minimizer . . . . .	3
2.2	Global minimizer . . . . .	4
<b>3</b>	<b>Constrained optimization problems</b>	<b>5</b>
3.1	General remarks . . . . .	5
3.2	$D = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{g}(\mathbf{x}) = \mathbf{c}\}$ . . . . .	8
3.3	$D = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{c}\}$ . . . . .	11

# 1 Overview about (static) optimization problems

In a general static optimization problem there is

- a real-valued function

$$f(\mathbf{x}) = f(x_1, \dots, x_n)$$

in  $n$  variables, the so-called objective function, whose value is to be optimized (maximized or minimized) and

- a set  $D \subset \mathbb{R}^n$ , the so-called admissible set.

Then the problem is to find (global) maximum or minimum points  $\mathbf{x}^* \in D$  of  $f$ :

$$\max(\min) f(\mathbf{x}) \text{ subject to } \mathbf{x} \in D.$$

**From now on we will always assume that  $f$  is at least 2-times continuously partial differentiable.**

Because  $\max f(\mathbf{x}) = \min -f(\mathbf{x})$  subject to  $\mathbf{x} \in D$  we could focus our attention (without loss of generality) on minimizing problems.

Depending on the set  $D$  and the function  $f$  several different types of optimization problems can arise. At the first level we will distinguish between so-called

1. unconstrained optimization problems:

$D$  contains no boundary points of  $D$ . This means that the set  $D$  is an open subset of  $\mathbb{R}^n$  and a solution of the optimization problem (if it exists) is an interior point of  $D$ .

**Example 1.1** *Solve the following problems or explain why there are no solutions:*

$$\min x^2 \text{ subject to } x \in D = (-1, 1)$$

$$\min -x^2 \text{ subject to } x \in D = (-1, 1)$$

$$\min x^2 \text{ subject to } x \in D = \mathbb{R}$$

$$\min 1/x \text{ subject to } x \in D = (0, 1)$$

$$\min -1/x \text{ subject to } x \in D = (0, 1)$$

$$\min x^2 - x^4 \text{ subject to } x \in D = (-2, 2)$$

$$\min x^2 - x^4 \text{ subject to } x \in D = (-1, 1)$$

$$\min x^2 - x^4 \text{ subject to } x \in D = (-0.1, 0.1)$$

$$\min \sin(1/x)/x \text{ subject to } x \in D = (0, 1)$$

2. constrained optimization problems:

$D$  contains some boundary points of  $D$ . A solution of the optimization problem may be an interior point or a point on the boundary of  $D$ .

## 2 Unconstrained optimization problems

### 2.1 Local minimizer

Consider a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ . Let  $D$  be some **open** subset of  $\mathbb{R}^n$  and  $\mathbf{x}^* \in D$  a local minimizer of  $f$  over  $D$ . This means that there exists an  $\epsilon > 0$  such that for all  $\mathbf{x} \in D$  satisfying  $\|\mathbf{x} - \mathbf{x}^*\| < \epsilon$  we have  $f(\mathbf{x}^*) \leq f(\mathbf{x})$ .

The term „unconstrained” usually refers to the situation where all points  $\mathbf{x}$  sufficiently near  $\mathbf{x}^*$  are in  $D$ . This is automatically true if  $D$  is an open set.

We already know:

#### Theorem 2.1 (First- and second order necessary conditions for optimality)

Suppose that  $\nabla^2 f$  is continuous in an open neighbourhood  $U$  of  $\mathbf{x}^*$  then

$$\mathbf{x}^* \text{ is a local minimizer of } f \implies \nabla f(\mathbf{x}^*) = \mathbf{0} \text{ and } \nabla^2 f(\mathbf{x}^*) \text{ is pos.semidef.}$$

Note that these necessary conditions are not sufficient.

#### Theorem 2.2 (First- and second order sufficient conditions for optimality)

Suppose that  $\nabla^2 f$  is continuous in an open neighbourhood  $U$  of  $\mathbf{x}^*$  then

$$\nabla f(\mathbf{x}^*) = \mathbf{0} \text{ and } \nabla^2 f(\mathbf{x}^*) \text{ is pos.def.} \implies \mathbf{x}^* \text{ is a (strict) local minimizer of } f$$

#### Proof:

Because  $\nabla^2 f$  is continuous and positive definite at  $\mathbf{x}^*$ , we can choose an open ball  $B = \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}^*\| < \epsilon\} \subset D$  where  $\nabla^2 f$  remains positive definite. Taking any nonzero vector  $\mathbf{v}$  with  $\|\mathbf{v}\| < \epsilon$ , we have  $\mathbf{x}^* + \mathbf{v} \in B$  and by Taylor’s theorem:

$$\begin{aligned} f(\mathbf{x}^* + \mathbf{v}) &= f(\mathbf{x}^*) + \mathbf{v}^T \nabla f(\mathbf{x}^*) + \frac{1}{2} \mathbf{v}^T \nabla^2 f(\mathbf{z}) \mathbf{v} \\ &= f(\mathbf{x}^*) + \frac{1}{2} \mathbf{v}^T \nabla^2 f(\mathbf{z}) \mathbf{v} \end{aligned}$$

for some  $\mathbf{z} = \mathbf{x}^* + t \cdot \mathbf{v}$  with  $t \in (0, 1)$ .

Since  $\mathbf{z} = \mathbf{x}^* + t \cdot \mathbf{v} \in B$ , we have  $\mathbf{v}^T \nabla^2 f(\mathbf{z}) \mathbf{v} > 0$  and therefore  $f(\mathbf{x}^* + \mathbf{v}) > f(\mathbf{x}^*)$ .  $\square$

## 2.2 Global minimizer

Of course, all local minimizers of a function  $f$  are candidates for global minimizing, but obviously, an arbitrary function may not realise a global minimum in an open set  $D$ . For instance, look at  $f(x) = -x^2$  subject to  $x \in D = (-1, 1)$ .

There are only general results in the case where  $f$  is a convex function on  $D$ . Because we define convexity of the function  $f$  by the inequality

$$f(t\mathbf{x} + (1-t)\mathbf{y}) \leq tf(\mathbf{x}) + (1-t)f(\mathbf{y})$$

for all  $\mathbf{x}, \mathbf{y} \in D$  and all  $t \in [0, 1]$ , all points  $t\mathbf{x} + (1-t)\mathbf{y}$  (points between  $\mathbf{x}$  and  $\mathbf{y}$ ) should lie in  $D$ . Hence  $D$  must be a convex set.

**Theorem 2.3** *Let  $f$  be a convex (resp. concave) and differentiable function on the convex (and open) set  $D$ . Then*

$$\mathbf{x}^* \text{ is a global minimizer (resp. maximizer) of } f \iff \nabla f(\mathbf{x}^*) = \mathbf{0}$$

**Proof** (for convex  $f$ ):

• „ $\implies$ “

Clear!?

• „ $\impliedby$ “

Let  $\nabla f(\mathbf{x}^*) = \mathbf{0}$  and suppose that  $\mathbf{x}^*$  is **not** a global minimizer of  $f$  on  $D$ . Then we can find a point  $\mathbf{y} \in D$  with  $f(\mathbf{y}) < f(\mathbf{x}^*)$ .

Consider the line segment that joins  $\mathbf{x}^*$  to  $\mathbf{y}$ , that is

$$\mathbf{z} = \mathbf{z}(t) = t\mathbf{y} + (1-t)\mathbf{x}^* = \mathbf{x}^* + t(\mathbf{y} - \mathbf{x}^*)$$

for all  $t \in [0, 1]$ . Of course,  $\mathbf{z} \in D$  because  $D$  is a convex set. Hence

$$\begin{aligned} \nabla f(\mathbf{x}^*)^T (\mathbf{y} - \mathbf{x}^*) &= \left. \frac{d}{dt} f(\mathbf{x}^* + t(\mathbf{y} - \mathbf{x}^*)) \right|_{t=0} \\ &= \lim_{t \rightarrow 0^+} \frac{f(\mathbf{x}^* + t(\mathbf{y} - \mathbf{x}^*)) - f(\mathbf{x}^*)}{t} \\ &\leq \lim_{t \rightarrow 0^+} \frac{tf(\mathbf{y}) + (1-t)f(\mathbf{x}^*) - f(\mathbf{x}^*)}{t} \\ &= \lim_{t \rightarrow 0^+} \frac{t(f(\mathbf{y}) - f(\mathbf{x}^*))}{t} \\ &= f(\mathbf{y}) - f(\mathbf{x}^*) < 0. \end{aligned}$$

Therefore,  $\nabla f(\mathbf{x}^*) \neq \mathbf{0}$ ! Contradiction.

Hence,  $\mathbf{x}^*$  is a global minimizer of  $f$  on  $D$ .

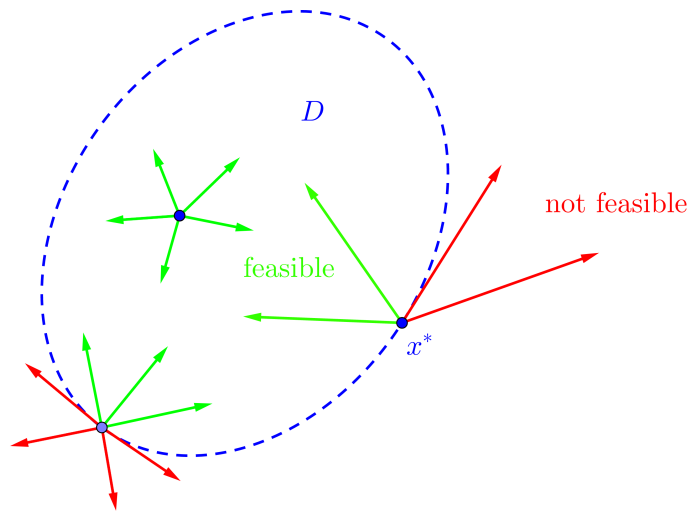
□

### 3 Constrained optimization problems

#### 3.1 General remarks

In the previous case we have used the fact that for every direction  $\mathbf{v}$  points of the form  $\mathbf{x}^* + t\mathbf{v}$  belong to  $D$  (for sufficiently small  $t$ ). This is no longer true if  $D$  has a boundary and  $\mathbf{x}^*$  is a point on this boundary.

**Definition 3.1** Let  $D \subset \mathbb{R}^n$  and  $\mathbf{x}^* \in D$ . A vector  $\mathbf{v} \in \mathbb{R}^n$  is called a feasible direction in  $\mathbf{x}^*$  if  $\mathbf{x}^* + t\mathbf{v} \in D$  for all  $t$  with  $0 \leq t < t_0$ .



If not all directions  $\mathbf{v}$  are feasible in  $\mathbf{x}^*$ , then the condition  $\nabla f(\mathbf{x}^*) = \mathbf{0}$  is no longer necessary for local optimality. But we can prove the following result.

**Theorem 3.1** If  $\mathbf{x}^*$  is a local minimum of the continuous differentiable function  $f$  on  $D$ , then

$$\partial_{\mathbf{v}} f(\mathbf{x}^*) = \nabla f(\mathbf{x}^*)^T \mathbf{v} \geq 0$$

for every feasible direction  $\mathbf{v}$  and

$$\mathbf{v}^T \nabla^2 f(\mathbf{x}^*) \mathbf{v} \geq 0$$

for all feasible directions with  $\partial_{\mathbf{v}} f(\mathbf{x}^*) = 0$ .

There are two cases:

1.  $\partial D \not\subset D$

There are boundary points of  $D$  which are not elements of  $D$ . This case is too difficult and we need a specific method, adapted to the concrete set  $D$ , to solve the optimization problem. We will **not** follow up on this type of problem.

2.  $\partial D \subset D$

The complete boundary  $\partial D$  of  $D$  is in  $D$ ; this means that  $D$  is closed.

From now on let  $D$  be always closed.

We recall the following basic existence result for **closed and bounded** sets  $D$ :

**Theorem 3.2 (Weierstrass-Theorem)** *If  $f$  is a continuous function and  $D$  is a closed and bounded set then there exists a global minimum of  $f$  over  $D$ .*

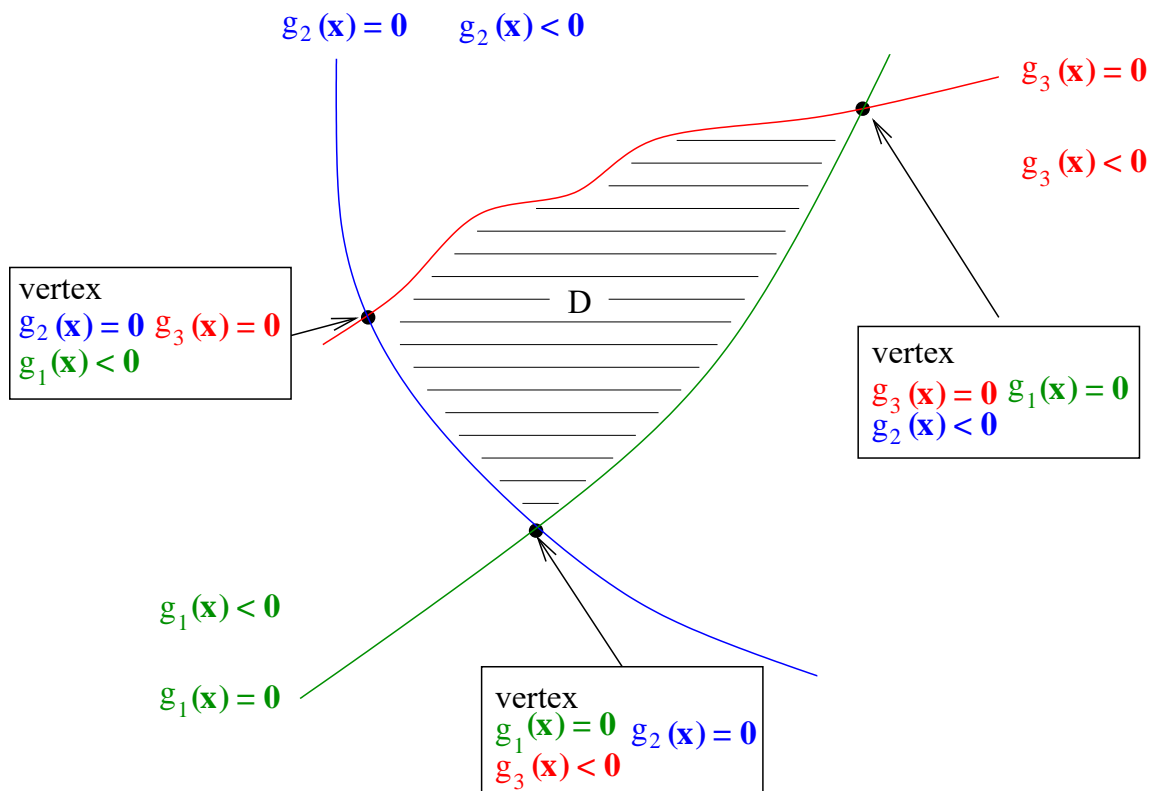
(General) Algorithm for finding a global minimum

1. Find all interior points of  $D$  satisfying  $\nabla f(\mathbf{x}^*) = \mathbf{0}$  (stationary points).
2. Find all points where  $\nabla f$  does not exist (critical points).
3. Find all boundary points satisfying  $\partial_{\mathbf{v}} f(\mathbf{x}^*) \geq 0$  for all feasible directions  $\mathbf{v}$ .
4. Compare all values at all these candidate points and choose one smallest one.

In almost all interesting optimization problems the admissible set  $D$  is given by a set of inequalities (or equations):

$$D = \{\mathbf{x} \in \mathbb{R}^n \mid g_1(\mathbf{x}) \leq c_1, g_2(\mathbf{x}) \leq c_2, \dots, g_m(\mathbf{x}) \leq c_m\} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{c}\}$$

with  $\mathbf{g} = (g_1, \dots, g_m)^T$ ,  $g_1, \dots, g_m : \mathbb{R}^n \rightarrow \mathbb{R}$  and  $\mathbf{c} = (c_1, \dots, c_m)^T$ .



It is easy to see that one equation of the form  $g(\mathbf{x}) = c$  can be expressed by the two inequalities  $g(\mathbf{x}) \leq c$  and  $-g(\mathbf{x}) \leq -c$ . Hence all sets described by a set of equations could be described by a set of inequalities and it would be enough to study sets described by inequalities.

But for practical reasons we will discuss the two cases separately.

**Definition 3.2** For the optimization problem

$$\begin{aligned} \max(\min) \quad & y = f(x_1, x_2, \dots, x_n) = f(\mathbf{x}) \\ \text{subject to} \quad & \begin{cases} g_1(x_1, x_2, \dots, x_n) = g_1(\mathbf{x}) \leq c_1 \\ g_2(x_1, x_2, \dots, x_n) = g_2(\mathbf{x}) \leq c_2 \\ \dots \\ g_m(x_1, x_2, \dots, x_n) = g_m(\mathbf{x}) \leq c_m \end{cases} \end{aligned}$$

the function (in  $n + m$  variables)

$$L(x_1, \dots, x_n, \lambda_1, \dots, \lambda_m) = f(x_1, x_2, \dots, x_n) - \sum_{j=1}^m \lambda_j (g_j(x_1, x_2, \dots, x_n) - c_j)$$

shortly

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) - \sum_{j=1}^m \lambda_j (g_j(\mathbf{x}) - c_j) = f(\mathbf{x}) - \boldsymbol{\lambda}^T (\mathbf{g}(\mathbf{x}) - \mathbf{c})$$

is called Lagrange function of the optimization problem.

### 3.2 $D = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{g}(\mathbf{x}) = \mathbf{c}\}$

Given the following optimization problem:

$$\begin{aligned} \max(\min) \quad & y = f(x_1, x_2, \dots, x_n) = f(\mathbf{x}) \\ \text{subject to} \quad & \begin{cases} g_1(x_1, x_2, \dots, x_n) = g_1(\mathbf{x}) = c_1 \\ g_2(x_1, x_2, \dots, x_n) = g_2(\mathbf{x}) = c_2 \\ \dots \\ g_m(x_1, x_2, \dots, x_n) = g_m(\mathbf{x}) = c_m \end{cases} \end{aligned}$$

**Theorem 3.3** *Suppose that*

- $f, g_1, \dots, g_m$  are defined on a set  $S \subset \mathbb{R}^n$
- $\mathbf{x}^* = (x_1^*, \dots, x_n^*)$  is an interior point of  $S$  that solves the optimization problem
- $f, g_1, \dots, g_m$  are continuously partial differentiable in a ball around  $\mathbf{x}^*$
- the Jacobi-matrix of the constraint functions

$$D\mathbf{g}(\mathbf{x}) = \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(\mathbf{x}) & \frac{\partial g_1}{\partial x_2}(\mathbf{x}) & \dots & \frac{\partial g_1}{\partial x_n}(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(\mathbf{x}) & \frac{\partial g_m}{\partial x_2}(\mathbf{x}) & \dots & \frac{\partial g_m}{\partial x_n}(\mathbf{x}) \end{pmatrix}$$

has rank  $m$  in  $\mathbf{x} = \mathbf{x}^*$ .

#### Necessary condition

Then there exists unique numbers  $\lambda_1^*, \dots, \lambda_m^*$  such that  $(\mathbf{x}^*, \boldsymbol{\lambda}^*) = (x_1^*, x_2^*, \dots, x_n^*, \lambda_1^*, \dots, \lambda_m^*)$  is a stationary point of the Lagrange-function:

$$L_{x_1}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = 0, \dots, L_{x_n}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = 0$$

and

shortly

$$\boxed{\nabla L(\mathbf{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}}$$

$$L_{\lambda_1}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = 0, \dots, L_{\lambda_m}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = 0$$

or expanded

$$\boxed{\nabla f(\mathbf{x}^*) - \sum_{j=1}^m \lambda_j^* \nabla g_j(\mathbf{x}^*) = \mathbf{0} \quad (\star)}$$

#### Sufficient condition

If there exist numbers  $\lambda_1^*, \dots, \lambda_m^*$  and an admissible  $\mathbf{x}^*$  which together satisfy the necessary condition, and if the Lagrange function  $L$  is concave (convex) in  $\mathbf{x}$  and  $S$  is convex, then  $\mathbf{x}^*$  solves the maximization (minimization) problem.



**Proof:**

**Necessary condition** We get a nice argument for condition  $(\star)$  by studying the optimal value function

$$f^*(\mathbf{c}) = \max\{f(\mathbf{x}) \mid \mathbf{g}(\mathbf{x}) = \mathbf{c}\}$$

If  $f$  is a profit function and  $\mathbf{c} = (c_1, \dots, c_m)$  denotes a resource vector, then  $f^*(\mathbf{c})$  is the maximum profit obtainable given the available resource vector  $\mathbf{c}$ .

In the following argument we **assume that  $f^*(\mathbf{c})$  is differentiable**.

Fix a vector  $\mathbf{c}^*$  and let  $\mathbf{x}^*$  be the corresponding optimal solution. Then  $f(\mathbf{x}^*) = f^*(\mathbf{c}^*)$  and obviously for all  $\mathbf{x}$  we have  $f(\mathbf{x}) \leq f^*(\mathbf{g}(\mathbf{x}))$ .

Hence

$$\phi(\mathbf{x}) := f(\mathbf{x}) - f^*(\mathbf{g}(\mathbf{x})) \leq 0$$

has a maximum in  $\mathbf{x} = \mathbf{x}^*$ , so

$$0 = \frac{\partial \phi}{\partial x_i}(\mathbf{x}^*) = \frac{\partial f}{\partial x_i}(\mathbf{x}^*) - \sum_{j=1}^m \left[ \frac{\partial f^*}{\partial c_j}(\mathbf{c}) \right]_{\mathbf{c}=\mathbf{g}(\mathbf{x}^*)} \frac{\partial g_j}{\partial x_i}(\mathbf{x}^*)$$

Define

$$\lambda_j^* := \frac{\partial f^*}{\partial c_j}(\mathbf{c}) \approx f^*(\mathbf{c} + \mathbf{e}_j) - f^*(\mathbf{c})$$

and equation  $(\star)$  follows.

**Sufficient condition** Suppose that  $L = L(\mathbf{x})$  is a concave (resp. convex) function in the variable  $\mathbf{x}$ . The necessary condition means that  $\mathbf{x}^*$  is a stationary point of  $L$ , this means  $\nabla_{\mathbf{x}} L(\mathbf{x}^*) = \mathbf{0}$ . Then by Theorem 2.3 we know that  $\mathbf{x}^*$  is a global maximizer (resp. minimizer) of  $L$  and this means that

$$\begin{aligned} L(\mathbf{x}^*) &= f(\mathbf{x}^*) - \sum_{j=1}^m \lambda_j^* (g_j(\mathbf{x}^*) - c_j) \\ &\geq f(\mathbf{x}) - \sum_{j=1}^m \lambda_j^* (g_j(\mathbf{x}) - c_j) \\ &= L(\mathbf{x}) \end{aligned}$$

for all  $\mathbf{x} \in S$ . But for all admissible  $\mathbf{x}$  we have  $g_j(\mathbf{x}) = c_j$ . Hence  $f(\mathbf{x}^*) \geq f(\mathbf{x})$  for all admissible  $\mathbf{x} \in S$ .  $\square$

**Example 3.1** Given the following optimization problem:

$$\begin{aligned} \max \quad & f(x_1, x_2) = x_1^\alpha x_2^\beta \\ \text{subject to} \quad & g(x_1, x_2) = p_1 x_1 + p_2 x_2 = c \end{aligned}$$

Then step by step we get:

- $L(x_1, x_2, \lambda) = x_1^\alpha x_2^\beta - \lambda(p_1 x_1 + p_2 x_2 - c)$
- $\nabla L(x_1, x_2, \lambda) = \nabla f(x_1, x_2) - \lambda \nabla g(x_1, x_2) = \begin{pmatrix} \alpha x_1^{\alpha-1} x_2^\beta - \lambda p_1 \\ \beta x_1^\alpha x_2^{\beta-1} - \lambda p_2 \\ -(p_1 x_1 + p_2 x_2 - c) \end{pmatrix}$
- $\begin{pmatrix} \alpha x_1^{\alpha-1} x_2^\beta - \lambda p_1 \\ \beta x_1^\alpha x_2^{\beta-1} - \lambda p_2 \\ -(p_1 x_1 + p_2 x_2 - c) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$  or
  - $E1: \quad \alpha x_1^{\alpha-1} x_2^\beta = \lambda p_1$
  - $E2: \quad \beta x_1^\alpha x_2^{\beta-1} = \lambda p_2$
  - $E3: \quad p_1 x_1 + p_2 x_2 = c$

- $E1/E2$

$$\frac{\alpha x_1^{\alpha-1} x_2^\beta}{\beta x_1^\alpha x_2^{\beta-1}} = \frac{\lambda p_1}{\lambda p_2} \Leftrightarrow \frac{\alpha x_2}{\beta x_1} = \frac{p_1}{p_2} \Leftrightarrow x_2 = \frac{p_1 \beta}{p_2 \alpha} x_1$$

- $x_2$  in  $E3$

$$p_1 x_1 + p_2 x_2 = c \Leftrightarrow p_1 x_1 + p_2 \left( \frac{p_1 \beta}{p_2 \alpha} x_1 \right) = c \Leftrightarrow x_1^* = \frac{c \alpha}{p_1 (\alpha + \beta)}$$

- $x_1$  in  $x_2$

$$x_2^* = \frac{p_1 \beta}{p_2 \alpha} x_1 = \frac{p_1 \beta}{p_2 \alpha} \frac{c \alpha}{p_1 (\alpha + \beta)} = \frac{c \beta}{p_2 (\alpha + \beta)}$$

- $x_1^*$  and  $x_2^*$  in  $E1$

$$\lambda^* = \frac{\alpha \left( \frac{c \alpha}{p_1 (\alpha + \beta)} \right)^{\alpha-1} \left( \frac{c \beta}{p_2 (\alpha + \beta)} \right)^\beta}{p_1} = \frac{\alpha^\alpha \beta^\beta c^{\alpha+\beta-1}}{p_1^\alpha p_2^\beta (\alpha + \beta)^{\alpha+\beta-1}}$$

- Hesse matrix of  $L$  with respect to  $\mathbf{x}$

$$\nabla_{\mathbf{x}}^2 L(\mathbf{x}) = \begin{pmatrix} \alpha(\alpha-1)x_1^{\alpha-2}x_2^\beta & \alpha\beta x_1^{\alpha-1}x_2^{\beta-1} \\ \alpha\beta x_1^{\alpha-1}x_2^{\beta-1} & \beta(\beta-1)x_1^\alpha x_2^{\beta-2} \end{pmatrix}$$

- If  $\nabla_{\mathbf{x}}^2 L(\mathbf{x})$  is negative definite (for all  $x_1, x_2 > 0$ ) then  $L$  is concave and  $\mathbf{x}^* = (x_1^*, x_2^*)$  solves the maximization problem.

Is  $\nabla_{\mathbf{x}}^2 L(\mathbf{x})$  negative definite?

### 3.3 $D = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{c}\}$

Given the following optimization problem:

$$\begin{aligned} \max \quad & y = f(x_1, x_2, \dots, x_n) = f(\mathbf{x}) \\ \text{subject to} \quad & \begin{cases} g_1(x_1, x_2, \dots, x_n) = g_1(\mathbf{x}) \leq c_1 \\ g_2(x_1, x_2, \dots, x_n) = g_2(\mathbf{x}) \leq c_2 \\ \dots \\ g_m(x_1, x_2, \dots, x_n) = g_m(\mathbf{x}) \leq c_m \end{cases} \end{aligned}$$

**Definition 3.3** Let  $\mathbf{x}^*$  be the solution of the maximization problem. The constraint  $g_i(\mathbf{x}) \leq c_i$  is called

- binding (or activ) at  $\mathbf{x}^*$ , if  $g_i(\mathbf{x}^*) = c_i$  and
- not binding (or inactiv) at  $\mathbf{x}^*$ , if  $g_i(\mathbf{x}^*) < c_i$ .

**Theorem 3.4** Suppose that

- $f, g_1, \dots, g_m$  are defined on a set  $S \subset \mathbb{R}^n$
- $\mathbf{x}^* = (x_1^*, \dots, x_n^*)$  is an interior point of  $S$  that solves the maximization problem
- $f, g_1, \dots, g_m$  are continuously partial differentiable in a ball around  $\mathbf{x}^*$
- the constraints are ordered in such a way, that the first  $m_0$  constraints are binding at  $\mathbf{x}^*$  and all the remaining  $m - m_0$  constraints are not binding,
- the Jacobi-matrix of the binding constraint functions

$$\begin{pmatrix} \frac{\partial g_1}{\partial x_1}(\mathbf{x}^*) & \dots & \frac{\partial g_1}{\partial x_n}(\mathbf{x}^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_{m_0}}{\partial x_1}(\mathbf{x}^*) & \dots & \frac{\partial g_{m_0}}{\partial x_n}(\mathbf{x}^*) \end{pmatrix}$$

has rank  $m_0$  in  $\mathbf{x} = \mathbf{x}^*$ .

#### Necessary condition

Then there exist unique real numbers  $\boldsymbol{\lambda}^* = (\lambda_1^*, \dots, \lambda_m^*)$  such that

1.  $L_{x_1}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = 0, \dots, L_{x_n}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = 0,$
2.  $\lambda_1^* \geq 0, \dots, \lambda_m^* \geq 0,$
3.  $\lambda_1^* \cdot [g_1(\mathbf{x}^*) - c_1] = 0, \dots, \lambda_m^* \cdot [g_m(\mathbf{x}^*) - c_m] = 0$  and
4.  $g_1(\mathbf{x}^*) \leq c_1, \dots, g_m(\mathbf{x}^*) \leq c_m.$

Conditions 1., 2. and 3. are often called Kuhn-Tucker-conditions.

**Proof:**

**Necessary condition** We study the optimal value function

$$f^*(\mathbf{c}) = \max\{f(\mathbf{x}) \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{c}\}$$

These value function must be nondecreasing in each variable  $c_1, \dots, c_m$ . This is because as  $c_j$  increases with all other variables held fixed, the admissible set becomes larger; hence  $f^*(\mathbf{c})$  can not decrease.

In the following argument we **assume that  $f^*(\mathbf{c})$  is differentiable**.

Fix a vector  $\mathbf{c}^*$  and let  $\mathbf{x}^*$  be the corresponding optimal solution. Then  $f(\mathbf{x}^*) = f^*(\mathbf{c}^*)$ . For any  $\mathbf{x}$  we have  $f(\mathbf{x}) \leq f^*(\mathbf{g}(\mathbf{x}))$  because  $\mathbf{x}$  obviously satisfies the constraints when each  $c_j^*$  is replaced by  $g_j(\mathbf{x})$ .

But then

$$f^*(\mathbf{g}(\mathbf{x})) \leq f^*(\mathbf{g}(\mathbf{x}) + \underbrace{\mathbf{c}^* - \mathbf{g}(\mathbf{x}^*)}_{\geq 0})$$

since  $\mathbf{g}(\mathbf{x}^*) \leq \mathbf{c}^*$  and  $f^*$  is nondecreasing.

Hence

$$\phi(\mathbf{x}) := f(\mathbf{x}) - f^*(\underbrace{\mathbf{g}(\mathbf{x}) + \mathbf{c}^* - \mathbf{g}(\mathbf{x}^*)}_{=: \mathbf{u}(\mathbf{x})}) \leq 0$$

for all  $\mathbf{x}$  and since  $\phi(\mathbf{x}^*) = 0$ ,  $\phi(\mathbf{x})$  has a maximum in  $\mathbf{x} = \mathbf{x}^*$ , so

$$\begin{aligned} 0 = \frac{\partial \phi}{\partial x_i}(\mathbf{x}^*) &= \frac{\partial f}{\partial x_i}(\mathbf{x}^*) - \sum_{j=1}^m \frac{\partial f^*}{\partial u_j}(\mathbf{u}(\mathbf{x}^*)) \frac{\partial u_j}{\partial x_i}(\mathbf{x}^*) \\ &= \frac{\partial f}{\partial x_i}(\mathbf{x}^*) - \sum_{j=1}^m \frac{\partial f^*}{\partial u_j}(\mathbf{c}^*) \frac{\partial g_j}{\partial x_i}(\mathbf{x}^*) \end{aligned}$$

Since  $f^*$  is nondecreasing, we have

$$\lambda_j^* := \frac{\partial f^*}{\partial c_j}(\mathbf{c}) \geq 0.$$

and we should (but will not) prove that if  $g_j(\mathbf{x}^*) < c_j^*$  then  $\lambda_j^* = 0$ . □

How should we solve a maximization problem with Kuhn-Tucker? Lets have a look at two examples.

**Always:**  $\lambda_j \geq 0$  and if  $g_j(\mathbf{x}) < c_j$  then  $\lambda_j = 0$ . **Respect the direction of the implication!**

**Not true:** If  $\lambda_j = 0$  then  $g_j(\mathbf{x}) < c_j$ .

### Example 3.2

$$\begin{aligned} \max \quad & f(x_1, x_2) = x_1^2 + x_2^2 + x_2 - 1 \\ \text{subject to} \quad & g(x_1, x_2) = x_1^2 + x_2^2 \leq 1 \end{aligned}$$

1. We have one constraint and need one Lagrange-multiplicator  $\lambda = \lambda_1$ . The Lagrange-funktion is:

$$L(x_1, x_2) = x_1^2 + x_2^2 + x_2 - 1 - \lambda (x_1^2 + x_2^2 - 1)$$

2. Write down the Kuhn-Tucker-conditions

(I)	$L_{x_1}(x_1, x_2) = 2x_1 - 2\lambda x_1 = 2x_1(1 - \lambda) = 0$
(II)	$L_{x_2}(x_1, x_2) = 2x_2 + 1 - 2\lambda x_2 = 0$
(III)	$\lambda \geq 0$ and $\lambda(x_1^2 + x_2^2 - 1) = 0$

3. Find all points  $(x_1, x_2, \lambda)$  which satisfy all Kuhn-Tucker-conditions and pay attention that for all these points  $x_1^2 + x_2^2 \leq 1$  (constraint).

#### Systematic way

From equation (I) we see, that  $\lambda = 1$  or  $x_1 = 0$ . The case  $\lambda = 1$  with equation (II) gives a contradiction. **Hence:**  $x_1 = 0$ .

All constraints could be binding (=) or not binding (<) and there are 2 possibilities, shortened by = and <.

=	$x_1^2 + x_2^2 = 1 \Rightarrow \lambda \geq 0$ with (III), first part
<	$x_1^2 + x_2^2 < 1 \Rightarrow \lambda = 0$ with (III), second part

- (a) Case = (or  $x_1^2 + x_2^2 = 1$ )

Then with  $x_1 = 0$  we get  $x_2 = \pm 1$ . With (II) we can compute the associated  $\lambda$  and get the two candidats for maximization:  $(0, 1, 3/2)$  and  $(0, -1, 1/2)$

- (b) Case < (or  $x_1^2 + x_2^2 < 1$ )

With  $\lambda = 0$  and  $x_1 = 0$  we get with (II) that  $x_2 = -1/2$ . We have found an third candidat for maximization:  $(0, -1/2, 0)$ .

With

$$f(0, 1) = 1, \quad f(0, -1) = -1 \quad \text{and} \quad f(0, -1/2) = -5/4$$

we see that  $(0, 1)$  (with  $\lambda = 3/2$ ) is the solution of the maximization problem.

**Example 3.3**

$$\begin{aligned} \max \quad & y = f(m, x) = m + \ln x \\ \text{subject to} \quad & \begin{cases} g_1(m, x) = m + x \leq 5 \\ g_2(m, x) = -m \leq 0 \\ g_3(m, x) = -x \leq 0 \end{cases} \end{aligned}$$

1. We have three constraints and need three Lagrange-multiplier  $\lambda_1, \lambda_2, \lambda_3$ . The Lagrange-funktion is:

$$\begin{aligned} L(x_1, x_2) &= m + \ln x - \lambda_1 (m + x - 5) - \lambda_2 (-m) - \lambda_3 (-x) \\ &= m + \ln x - \lambda_1 (m + x - 5) + \lambda_2 m + \lambda_3 x \end{aligned}$$

2. Write down the Kuhn-Tucker-conditions

(I)	$L_{x_1}(x_1, x_2) = 1 - \lambda_1 + \lambda_2$	$= 0$
(II)	$L_{x_2}(x_1, x_2) = \frac{1}{x} - \lambda_1 + \lambda_3$	$= 0$
(III)	$\lambda_1 \geq 0$ and $\lambda_1(m + x - 5) = 0$	
(IV)	$\lambda_2 \geq 0$ and $\lambda_2(-m) = 0$	
(V)	$\lambda_3 \geq 0$ and $\lambda_3(-x) = 0$	

3. Find all points  $(x_1, x_2, \lambda_1, \lambda_2, \lambda_3)$  which satisfy all Kuhn-Tucker-conditions and all constraints.

**Systematic way**

All constraints could be binding (=) or not binding (<) and there are  $2 \cdot 2 \cdot 2 = 8$  possibilities. Of course, some of these combinations are obviously impossible.

(=, =, =)	$m + x = 5$	$-m = 0$	$-x = 0$	$\Rightarrow$	$\lambda_1 \geq 0$	$\lambda_2 \geq 0$	$\lambda_3 \geq 0$	no solution
(<, =, =)	$m + x < 5$	$-m = 0$	$-x = 0$	$\Rightarrow$	$\lambda_1 = 0$	$\lambda_2 \geq 0$	$\lambda_3 \geq 0$	no solution
(=, <, =)	$m + x = 5$	$-m < 0$	$-x = 0$	$\Rightarrow$	$\lambda_1 \geq 0$	$\lambda_2 = 0$	$\lambda_3 \geq 0$	no solution
(=, =, <)	$m + x = 5$	$-m = 0$	$-x < 0$	$\Rightarrow$	$\lambda_1 \geq 0$	$\lambda_2 \geq 0$	$\lambda_3 = 0$	no solution
(<, <, =)	$m + x < 5$	$-m < 0$	$-x = 0$	$\Rightarrow$	$\lambda_1 = 0$	$\lambda_2 = 0$	$\lambda_3 \geq 0$	no solution
(<, =, <)	$m + x < 5$	$-m = 0$	$-x < 0$	$\Rightarrow$	$\lambda_1 = 0$	$\lambda_2 \geq 0$	$\lambda_3 = 0$	no solution
(=, <, <)	$m + x = 5$	$-m < 0$	$-x < 0$	$\Rightarrow$	$\lambda_1 \geq 0$	$\lambda_2 = 0$	$\lambda_3 = 0$	(4, 1, 1, 0, 0)
(<, <, <)	$m + x < 5$	$-m < 0$	$-x < 0$	$\Rightarrow$	$\lambda_1 = 0$	$\lambda_2 = 0$	$\lambda_3 = 0$	no solution

Confirm all these results!

**Elegant way**

If  $m + x < 5$  then  $\lambda_1 = 0$  by (III) and  $1 + \lambda_2 = 0$  or  $\lambda_2 = -1 < 0$  by (I) which contradicts (IV). Hence  $m + x = 5$  and we have to check only 4 possibilities (=, \*, \*).

Because  $\ln(x)$  is not defined in  $x = 0$  (and by equation (II)) we see that  $-x < 0$  (resp.  $x > 0$ ). Hence we have to check the two possibilities (=, \*, <).

- (=, =, <) means  $m + x = 5$ ,  $m = 0$  and  $x > 0$  (and  $\lambda_3 = 0$  by (V)). Then  $m = 5$  and  $\lambda_1 = 1/5$  by (II),  $\lambda_2 = -4/5$  by (I). This contradicts (IV).
- (=, <, <) means  $m + x = 5$ ,  $m > 0$  and  $x > 0$  (and  $\lambda_2 = \lambda_3 = 0$  by (IV) and (V)). Then  $\lambda_1 = 1$  by (I),  $x = 1$  and  $m = 5 - 1 = 4$ . We get the unique solution (4, 1, 1, 0, 0).