QEM "Optimization", WS 2017/18

Part 4. Constrained optimization

(about 4 Lectures)

Supporting Literature: ***Angel de la Fuente*, "Mathematical Methods and Models for Economists",** Chapter 7

Contents

4 Constrained Optimization

4.1 Equality Constrains: The Lagrange Problem

Typical Example from Economics:

A consumer chooses how much of the available income I to spend on:

$$\begin{array}{ccc} \text{goods} & \text{units} & \text{price per unit} \\ 1 & x_1 & p_1 \\ \vdots & \vdots & \vdots \\ n & x_n & p_n \end{array}.$$

The consumer preferences are measured by the **utility function** $u(x_1, \dots, x_n)$. The consumer faces the problem of choosing $(x_1, \dots, x_n)$ in order to maximize $u(x_1, \dots, x_n)$ subject to the budget constraint $p_1 x_1 + \dots + p_n x_n = I$.

Mathematical formalization:

$$\textbf{maximize} \;\; u(x_1, \dots, x_n),$$

$$\textbf{subject to} \;\; p_1 x_1 + \dots + p_n x_n = I.$$

We ignore for a moment that $x_1, \dots, x_n \geq 0$ and that possibly not the whole income I may be spent. To solve this and similar problems economists make use of the **Lagrange multiplier method**.

4.1.1 Lagrange Problem: Mathematical Formulation

Let U be a subset of $\mathbb{R}^n$ and let $m \leq n$. Let moreover $f : U \to \mathbb{R}$ and $g : U \to \mathbb{R}^m$ be functions (usually C^1- or even C^2-class).

Equality Constraint (EC) Problem: maximize the **objective function** $f(x)$ subject to $g(x) = 0$, i.e. find

$$\max_{x \in U,\; g(x)=0} f(x).$$

The components of $g = (g_1, \dots, g_m)$ are called the **constraint functions** and the set $D := \{x \in U \mid g(x) = 0\}$ is called the **constraint set**.

The method is named after the Italian/French mathematician J. L. Lagrange (1736–1813). In economics, the method was first implemented (≈1876) by the Danish economist H. Westergard.

We are looking for points $x^* \in D$ that are (local) maxima of f. We could have a unique x^* or not. x^* could exist or not exist at all.

Definition 4.1.1. *A point $x^* \in D$ is called a* ***local max*** *(resp.* ***min****) for the EC problem if there exists $\varepsilon > 0$ such that for all $x \in B_\varepsilon(x^*) \cap D$, one has*

$$f(x^*) \geq f(x) \quad (\textit{resp. } f(x^*) \leq f(x)).$$

Moreover, this point is a ***global max*** *(resp.* ***min****) if $f(x^*) \geq f(x)$ (resp. $f(x^*) \leq f(x)$) for all $x \in D$.*

4.1.2 The Simplest Case of EC: n=2, m=1

Let $U \subset \mathbb{R}^2$ be open and let $f, g : U \to \mathbb{R}$ be continuously differentiable. We want to compute

$$\max\{f(x_1, x_2) \mid (x_1, x_2) \in U,\ g(x_1, x_2) = 0\}.$$

Let (x_1^*, x_2^*) be *some local maximiser* for the EC problem (provided it exists). How do we find all such (x_1^*, x_2^*)? The Theorem of *Lagrange* (which will be precisely formulated later) gives the *necessary* conditions which should be satisfied by *any local optima* in this problem. Based on the Lagrange Theorem**,** we should proceed as follows to find all possible candidates for (x_1^*, x_2^*).

A Formal Scheme of the Lagrange Method

(1) Write down the so-called Lagrangian function (or simple 'the Lagrangian')

$$\mathcal{L}(x_1, x_2) := f(x_1, x_2) - \lambda g(x_1, x_2)$$

with a constant $\lambda \in \mathbb{R}$ – **Lagrange multiplier.**

(2) Take the partial derivatives of $\mathcal{L}(x_1, x_2)$ **w.r.t.** x_1 and x_2

$$\begin{aligned} \frac{\partial}{\partial x_1}\mathcal{L}(x_1, x_2) &:= \frac{\partial f}{\partial x_1}(x_1, x_2) - \lambda \frac{\partial g}{\partial x_1} g(x_1, x_2), \\ \frac{\partial}{\partial x_2}\mathcal{L}(x_1, x_2) &:= \frac{\partial f}{\partial x_2}(x_1, x_2) - \lambda \frac{\partial g}{\partial x_2} g(x_1, x_2). \end{aligned}$$

As will be explained below, a solution (x_1^*, x_2^*) to the EC problem can only be a point for which

$$\frac{\partial}{\partial x_1}\mathcal{L}(x_1, x_2) = \frac{\partial}{\partial x_2}\mathcal{L}(x_1, x_2) = 0$$

for a suitable $\lambda = \lambda(x_1^*, x_2^*)$. This leads to the next step:

(3) Solve the system of three equations and find all possible solutions $(x_1^*, x_2^*; \lambda^*) \in U \times \mathbb{R}$

$$\begin{cases} \frac{\partial}{\partial x_1}\mathcal{L}(x_1, x_2) = \frac{\partial f}{\partial x_1}(x_1, x_2) - \lambda \frac{\partial g}{\partial x_1} g(x_1, x_2) = 0, \\ \frac{\partial}{\partial x_2}\mathcal{L}(x_1, x_2) = \frac{\partial f}{\partial x_2}(x_1, x_2) - \lambda \frac{\partial g}{\partial x_2} g(x_1, x_2) = 0, \\ \frac{\partial}{\partial \lambda}\mathcal{L}(x_1, x_2) = -g(x_1, x_2) = 0. \end{cases}$$

So, any candidate for local extrema (x_1^*, x_2^*) is a solution, with its own $\lambda^* \in \mathbb{R}$, to the system

$$\frac{\partial \mathcal{L}}{\partial x_1} = 0, \quad \frac{\partial \mathcal{L}}{\partial x_2} = 0, \quad \frac{\partial \mathcal{L}}{\partial \lambda} = 0.$$

These 3 conditions are called the **first order** conditions for the EC problem.

Remark 4.1.2. *This procedure would not have worked if both* $\frac{\partial g}{\partial x_1}$ *and* $\frac{\partial g}{\partial x_2}$ *were zero at* (x_1^*, x_2^*), *i.e.,* (x_1^*, x_2^*) *is a critical point of* g. *The restriction that* U *does not contain critical points of* g *is called a* ***constraint qualification*** *in the domain* U. *The restriction that* $\nabla g(x_1^*, x_2^*) \neq 0$ *implies the constrain qualification in some neighborhood of the point* (x_1^*, x_2^*).

Remark 4.1.3.

i. A magic process!!! To solve the ***constraint*** *problem for* ***two*** *variables* (x_1, x_2) *we transform it into the* ***unconstrained*** *problem in three variables by adding an artificial variable* λ*).*

ii. The same scheme works whether we are minimising $f(x_1, x_2)$. *To distinguish max from min, one needs second order conditions.*

Example 4.1.4. *Maximize* $f(x_1, x_2) = x_1 x_2$ *subject to* $2x_1 + x_2 = 100$.

Solution 4.1.5. *Define* $g(x_1, x_2) = 2x_1 + x_2 - 100$ *and the Lagrangian*

$$\mathcal{L}(x_1, x_2) = x_1 x_2 - \lambda \left(2x_1 + x_2 - 100\right).$$

The 1st order conditions for the solutions of EC problem:

$$\frac{\partial \mathcal{L}}{\partial x_1} = x_2 - 2\lambda = 0, \frac{\partial \mathcal{L}}{\partial x_2} = x_1 - \lambda = 0, g(x_1, x_2) = 2x_1 + x_2 - 100 = 0.$$

Herefrom,

$$x_2 = 2\lambda, x_1 = \lambda, 2\lambda + 2\lambda = 100 \Longleftrightarrow \lambda = 25.$$

The only candidate for the solution

$$x_1 = 25, \quad x_2 = 50, \quad \lambda = 25.$$

The constrain qualification holds at all points $(x, y) \in \mathbb{R}^2$*:*

$$\frac{\partial g}{\partial x_1} = 2, \quad \frac{\partial g}{\partial x_2} = 1.$$

The solution obtained can be confirmed by the substitution method:

$$\begin{aligned} x_2 = 100 - 2x_1 &\Longrightarrow h(x_1) = x_1(100 - 2x_1) = 2x_1(50 - x_1),\ h'(x_1) = -4x_1 + 100 \\ &\Longrightarrow x_1 = 25, h''(x_1) = -4 < 0. \end{aligned}$$

Therefore, $x_1 = 25$ *is a max point for* $h \implies x_1 = 25,\ x_2 = 50$ *is a max point for* f.

Justification of the EC scheme: An analytic argument

How to find a local max / min of $f(x_1, x_2)$ subject to $g(x_1, x_2) = 0$.

Let $(x_1^*, x_2^*) \in U$ be a local extrema for the EC problem and let $\nabla g(x_1^*, x_2^*) \neq 0$. Without loss of generality assume that $\partial g/\partial x_2(x_1^*, x_2^*) \neq 0$. Then by the Implicit Function Theorem (IFT) the equation $g(x_1, x_2) = 0$ defines a differentiable function $x_2 := i(x_1)$ such that

$$g(x_1, i(x_1)) = 0 \quad \text{near} \quad (x_1^*, x_2^*) \in U$$

and

$$i'(x_1^*) = -\frac{\partial g/\partial x_1}{\partial g/\partial x_2}(x_1^*, x_2^*).$$

Then

$$h(x_1) := f(x_1, i(x_1))$$

has a local extremum at the point x_1^*. By the Chain Rule

$$\begin{aligned} 0 &= h'(x_1^*) = \frac{\partial f}{\partial x_1}(x_1^*, x_2^*) + \frac{\partial f}{\partial x_2}(x_1^*, x_2^*) i'(x_1^*) \\ &= \frac{\partial f}{\partial x_1}(x_1^*, x_2^*) - \frac{\partial f}{\partial x_2}(x_1^*, x_2^*) \frac{\partial g/\partial x_1}{\partial g/\partial x_2}(x_1^*, x_2^*). \end{aligned}$$

Hence,

$$\frac{\partial f}{\partial x_1}(x_1^*, x_2^*) = \frac{\partial f/\partial x_2}{\partial g/\partial x_2}(x_1^*, x_2^*) \frac{\partial g}{\partial x_1}(x_1^*, x_2^*).$$

Denoting

$$(!!!) \quad \lambda := \frac{\partial f/\partial x_2}{\partial g/\partial x_2}(x_1^*, x_2^*) \in \mathbb{R},$$

we have

$$\frac{\partial f}{\partial x_1}(x_1^*, x_2^*) - \lambda \frac{\partial f}{\partial x_1}(x_1^*, x_2^*) = 0$$

and

$$\frac{\partial f}{\partial x_2}(x_1^*, x_2^*) - \lambda \frac{\partial f}{\partial x_2}(x_1^*, x_2^*) = 0.$$

4.1.3 More variables: $n \geq 2, m = 1$

Goal: Find min/max of $f(x_1, \ldots, x_n) = 0$ subject to $g(x_1, \ldots, x_n) = 0$.

Define the Lagrangian with the multiplier $\lambda \in \mathbb{R}$ as

$$\mathcal{L}(x_1, \ldots, x_n) := f(x_1, \ldots, x_n) - \lambda g(x_1, \ldots, x_n)$$

where $(x_1, \ldots, x_n) \in U \subset \mathbb{R}^n$ and U is open.

Theorem 4.1.6 (Lagrange Theorem for a single constraint equation)**.** *Let $U \subset \mathbb{R}^n$ be open and let $f, g : U \to \mathbb{R}$ be continuously differentiable. Let $x^* = (x_1^*, \dots, x_n^*) \in U$ be a local extremum for $f(x_1, \dots, x_n)$ under the equality constraint $g(x_1, \dots, x_n) = 0$. Suppose further that $\nabla g(x^*) \neq 0$, i.e., at least one of $\partial g/\partial x_j(x^*) \neq 0$, $1 \leq j \leq n$. Then there exists a unique number $\lambda^* \in \mathbb{R}$ such that*

$$\frac{\partial f}{\partial x_j}(x^*) = \lambda^* \frac{\partial g}{\partial x_j}(x^*), \quad \text{for all } 1 \leq j \leq n,$$
$$\text{or} \quad \nabla f(x^*) = \lambda^* \nabla g(x^*).$$

In particular, for any pair (i, j), where $1 \leq i, j \leq n$, one has

$$\frac{\partial f}{\partial x_i}(x^*) / \frac{\partial f}{\partial x_j}(x^*) = \frac{\partial g}{\partial x_i}(x^*) / \frac{\partial g}{\partial x_j}(x^*).$$

Constraint qualification (CQ): We assume that $\nabla g(x^*) \neq 0$. The method in general **fails** if $\nabla g(x^*) = 0$. All such critical points should be treated separately by calculating $f(x^*)$.

The Theorem of Lagrange only provides **necessary** conditions for local optima x^* and, moreover, only for those which meet **CQ**, i.e., $\nabla g(x^*) \neq 0$. These conditions are **not sufficient**!

Counterexample 4.1.7. *Maximize $f(x_1, x_2) = -x_2$ subject to $g(x_1, x_2) = x_2^3 - x_1^2 = 0$, with $(x_1, x_2) \in U = \mathbb{R}^2$.*

Solution 4.1.8. *Since $x_2^3 = x_1^2 \implies x_2 \geq 0$. Moreover, $x_2 = 0 \Leftrightarrow x_1 = 0$. So, $(x_1^*, x_2^*) = (0, 0)$ is the global max of f under the constraint $g = 0$. But $\nabla g(x_1^*, x_2^*) = 0$, i.e., the constaint qualification does not hold. Furthermore, $\nabla f(x_1, x_2) = (0, -1)$ for all (x_1, x_2), and there cannot exist any $\lambda \in \mathbb{R}$ such that*

$$\nabla f(x^*) - \lambda \nabla g(x^*) = 0 \quad (\text{since} \quad -1 \neq \lambda \cdot 0).$$

*The Lagrange Theorem is **not** applicable.*

Remark 4.1.9.

i. *On the technical side: we need $\nabla g(x^*) \neq 0$ to apply IFT.*

ii. *If $\nabla g(x^*) = 0$, it still can happen that $\nabla f(x^*) = \lambda \nabla g(x^*) = 0$ (Suppose e.g. that $f : U \to \mathbb{R}$ has a strict global min/max in x^* and hence $\nabla f(x^*) = 0$).*

iii. *It is also possible that the constraint qualifications holds, but the EC problem has no solutions, see the example below.*

$$f(x_1, x_2) = x_1^2 - x_2^2 \quad \text{subject to} \quad g(x_1, x_2) = 1 - x_1 - x_2.$$

Then

$$\nabla g(x_1, x_2) = (-1, -1) \neq 0 \quad \textit{everywhere.}$$

Define the Lagrangian

$$\mathcal{L}(x_1, x_2) = f(x_1, x_2) - \lambda g(x_1, x_2).$$

Then

$$\begin{cases} 2x_1 + \lambda = 0 \\ -2x_2 + \lambda = 0 \\ 1 - x_1 - x_2 = 0 \end{cases} \iff \begin{array}{c} \lambda \neq 0,\ x_1 = -x_2,\ x_1 + x_2 = 1,\ \textit{or} \\ \lambda = 0,\ x_1 = x_2 = 0,\ x_1 + x_2 = 1 \end{array}$$

There are thus ***NO solutions to EC****!! Indeed, put*

$$x_2 = 1 - x_1 \quad \textit{and} \quad h(x_1) := x_1^2 - (1 - x_1)^2 = -1 + 2x_1.$$

Then there are ***NO local extrema****!!*

4.1.4 More Variables and More Constraints: $n \geq m$

Theorem 4.1.10 (General Form of the Lagrange Theorem)**.** *Let $U \subset \mathbb{R}^n$ be open and let*

$$f : U \to \mathbb{R}, \quad g : U \to \mathbb{R}^m \quad (m \leq n)$$

be continuously differentiable. Suppose that $x^* = (x_1^*, \ldots, x_n^*) \in U$ is a local extremum for $f(x_1, \ldots, x_n)$ under the equality constraints

$$\begin{cases} g_1(x_1, \ldots, x_n) = 0, \\ \quad\vdots \\ g_m(x_1, \ldots, x_n) = 0. \end{cases}$$

Suppose further that the matrix $Dg(x^*)$ has ***rank*** *m. Then there* ***exists a unique vector*** *$\lambda^* = (\lambda_1^*, \ldots, \lambda_m^*) \in \mathbb{R}^m$ such that*

$$\frac{\partial f}{\partial x_j}(x^*) = \sum_{i=1}^{m} \lambda_i^* \frac{\partial g_i}{\partial x_j}(x^*), \quad \textit{for all } 1 \leq j \leq n.$$

In other words,

$$\underbrace{\nabla f(x^*)}_{1\times n} = \underbrace{\lambda}_{1\times m} \times \underbrace{DG(x^*)}_{m\times n} \quad \textit{(product of } 1\times m \textit{ and } m \times n \textit{ matrices)},$$

$$\left(\frac{\partial f}{\partial x_1}(x^*), \ldots, \frac{\partial f}{\partial x_n}(x^*)\right) = (\lambda_1^*, \ldots, \lambda_m^*) \times \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(x^*) & \cdots & \frac{\partial g_1}{\partial x_n}(x^*) \\ \vdots & & \vdots \\ \frac{\partial g_m}{\partial x_1}(x^*) & \cdots & \frac{\partial g_m}{\partial x_1}(x^*) \end{pmatrix}.$$

Constraint Qualification: The **rank** of the Jacobian matrix

$$Dg(x^*) = \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(x^*) & \cdots & \frac{\partial g_1}{\partial x_n}(x^*) \\ \vdots & & \vdots \\ \frac{\partial g_m}{\partial x_1}(x^*) & \cdots & \frac{\partial g_m}{\partial x_1}(x^*) \end{pmatrix}$$

is equal to the **number of the constraints**, i.e.,

$$\text{rank } Dg(x^*) = m.$$

This ensures that $Dg(x^*)$ contains an **invertible** $m \times m$ submatrix, which will be used to determine $\lambda^* \in \mathbb{R}^m$.

Proof of Theorem 4.1.10. The **main ingredients** of the proof are:

i. Implicit Function Theorem, and

ii. Chain Rule for Derivatives.

By assumption, there exists an $m \times m$ submatrix of $Dg(x^*)$ with **full rank**, i.e., its **determinant is non-zero**. Without loss of generality, such submatrix can be chosen as

$$D_{\leq m}g(x^*) := \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(x^*) & \cdots & \frac{\partial g_1}{\partial x_m}(x^*) \\ \vdots & & \vdots \\ \frac{\partial g_m}{\partial x_1}(x^*) & \cdots & \frac{\partial g_m}{\partial x_m}(x^*) \end{pmatrix}$$

(otherwise we can change the numbering of variables $x_1, \ldots, x_n$). So, we have

$$\det D_{\leq m}g(x^*) \neq 0,$$

and hence there exists the inverse matrix $[D_{\leq m}g(x^*)]^{-1}$. By the IFT there exist C^1-functions

$$i_1(x_{m+1}, \ldots, x_n), \ldots, \; i_m(x_{m+1}, \ldots, x_n)$$

such that

$$g\left(i_1(x_{m+1}, \ldots, x_n), \ldots, i_m(x_{m+1}, \ldots, x_n), x_{m+1}, \ldots, x_n\right) = 0 \text{ near } (x_1^*, \ldots, x_n^*),$$

and moreover

$$\underbrace{Di(x_{m+1}^*, \ldots, x_n^*)}_{m\times(n-m)} = -\left[D_{\leq m}\underbrace{g(x_1^*, \ldots, x_n^*)}_{m\times n}\right]^{-1} \times \underbrace{D_{>m}g(x_1^*, \ldots, x_n^*)}_{n\times(n-m)},$$

where

$$D_{>m}g(x^*) := \begin{pmatrix} \frac{\partial g_1}{\partial x_{m+1}}(x^*) & \cdots & \frac{\partial g_1}{\partial x_n}(x^*) \\ \vdots & & \vdots \\ \frac{\partial g_m}{\partial x_{m+1}}(x^*) & \cdots & \frac{\partial g_m}{\partial x_n}(x^*) \end{pmatrix}.$$

Then $(x^*_{m+1}, \ldots, x^*_n)$ is a local extrema of the C^1-function

$$h(x_{m+1}, \ldots, x_n) := f\left(i_1(x_{m+1}, \ldots, x_n), \ldots, i_m(x_{m+1}, \ldots, x_n), x_{m+1}, \ldots, x_n\right).$$

Hence, by the Chain Rule

$$\begin{aligned} \underbrace{0}_{\in\, \mathbb{R}^{n-m}} &= \nabla h(x^*_{m+1}, \ldots, x^*_n) \\ &= \underbrace{\nabla_{\leq m} f(x^*_1, \ldots, x^*_n)}_{1\times m} \times \underbrace{Di(x^*_{m+1}, \ldots, x^*_n)}_{m\times(n-m)} + \underbrace{\nabla_{>m} f(x^*_1, \ldots, x^*_n)}_{1\times(n-m)} \\ &= -\underbrace{\nabla_{\leq m} f(x^*)}_{1\times m} \times \underbrace{[D_{\leq m} g(x^*)]^{-1}}_{m\times m} \times \underbrace{D_{>m} g(x^*)}_{m\times(n-m)} + \underbrace{\nabla_{>m} f(x^*)}_{1\times(n-m)} \end{aligned}$$

or

$$\begin{aligned} \underbrace{\nabla_{>m} f(x^*)}_{1\times(n-m)} &= \underbrace{\nabla_{\leq m} f(x^*)}_{1\times m} \times \underbrace{[D_{\leq m} g(x^*)]^{-1}}_{m\times m} \times \underbrace{D_{>m} g(x^*)}_{m\times(n-m)} \qquad (*) \\ &= \underbrace{\lambda^*}_{1\times m} \times \underbrace{D_{>m} g(x^*)}_{m\times(n-m)} \end{aligned}$$

where we set

$$\mathbb{R}^m \ni \underbrace{\lambda^*}_{1\times m} := (\lambda^*_1, \ldots, \lambda^*_m) := \underbrace{\nabla_{\leq m} f(x^*)}_{1\times m} \times \underbrace{[D_{\leq m} g(x^*)]^{-1}}_{m\times m}. \qquad (**)$$

So, we have from $(*)$

$$(i) \quad \frac{\partial f}{\partial x_j}(x^*) = \sum_{i=1}^m \lambda^*_i \frac{\partial g_i}{\partial x_j}(x^*), \quad \textit{for all } m+1 \leq j \leq n,$$

and respectively from $(**)$

$$\begin{aligned} \underbrace{\nabla_{\leq m} f(x^*)}_{1\times m} \underbrace{[D_{\leq m} g(x^*)]^{-1}}_{m\times m} = \underbrace{\lambda^*}_{1\times m} \quad &\Longleftrightarrow \\ \underbrace{\nabla_{\leq m} f(x^*)}_{1\times m} = \underbrace{\lambda^*}_{1\times m} \times D_{\leq m} g(x^*) \quad &\Longleftrightarrow \\ (ii) \quad \frac{\partial f}{\partial x_j}(x^*) = \sum_{i=1}^m \lambda^*_i \frac{\partial g_i}{\partial x_j}(x^*), \quad &\textit{for all } 1 \leq j \leq n, \end{aligned}$$

which proves the theorem. □

4.2 A "Cookbook" Procedure: How to use the Multidimensional Theorem of Lagrange

(1) Set up the Lagrangian function

$$U \ni (x_1, \ldots, x_n) \to \mathcal{L}(x_1, \ldots, x_n) := f(x_1, \ldots, x_n) - \sum\nolimits_{i=1}^{m} \lambda_i g_i(x_1, \ldots, x_n)$$

with a vector of Lagrange multipliers $\lambda = (\lambda_1, \ldots, \lambda_m) \in \mathbb{R}^m$.

(2) Take the partial derivatives of $\mathcal{L}(x_1, \ldots, x_n)$ w.r.t. x_j, $1 \leq j \leq n$,

$$\frac{\partial}{\partial x_j}\mathcal{L}(x_1, \ldots, x_n) := \frac{\partial f}{\partial x_j}(x_1, \ldots, x_n) - \sum\nolimits_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_j}(x_1, \ldots, x_n).$$

(3) Find the set of all critical points $(x_1^*, \ldots, x_n^*) \in U$ for the Lagrangian $\mathcal{L}(x_1, \ldots, x_n)$. To this end, solve the system of $(n+m)$ equations

$$\begin{cases} \frac{\partial}{\partial x_j}\mathcal{L}(x_1, \ldots, x_n) = 0, & 1 \leq j \leq n, \\ \frac{\partial}{\partial \lambda_i}\mathcal{L}(x_1, \ldots, x_n) = -g_i(x_1, \ldots, x_n) = 0, & 1 \leq i \leq m, \end{cases}$$

with $(n+m)$ unknowns $(x_1, \ldots, x_n) \in U, (\lambda_1, \ldots, \lambda_m) \in \mathbb{R}^m$.

Every critical point $(x_1^*, \ldots, x_n^*; \lambda_1^*, \ldots, \lambda_m^*) \in U \times \mathbb{R}^m$ for $\mathcal{L}$ gives us the candidate $(x_1^*, \ldots, x_n^*)$ for the local extrema of the EC problem, provided this $(x_1^*, \ldots, x_n^*)$ satisfies the constraint qualification rank $Dg(x^*) = m$. To check whether x^* is a local (global) max / min, we need to evaluate f at each point x^*.

The points x_* at which the constraint qualification fails (i.e., rank $Dg(x_*) < m$) should be considered separately since the Lagrange Theorem is not applicable to them.

Example 4.2.1 (Economic / Numerical Example to EC)**.** *Maximize the **Cobb-Douglas utility function***

$$u(x_1, x_2, x_3) = x_1^2 x_2^3 x_3, \quad x_1, x_2, x_3 \geq 0 \quad (\in \mathbb{R}_+),$$

under the budjet constraint

$$x_1 + x_2 + x_3 = 12.$$

Solution 4.2.2. *The global maximum exists by the **Weierstrass** theorem, since u is a continuous function defined on a compact domain*

$$D := \left\{ (x_1, x_2, x_3) \in \mathbb{R}_+^3 \mid x_1 + x_2 + x_3 = 12 \right\}.$$

If any of x_1, x_2, x_3 is zero, then $u(x_1, x_2, x_3) = 0$, which is not the max value.

So, it is enough to solve the Lagrange optimization problem in the open domain

$$\mathring{U} := \left\{(x_1, x_2, x_3) \in \mathbb{R}^3_{>0}\right\}.$$

The Lagrangian is

$$\mathcal{L}(x_1, x_2, x_3) = x_1^2 x_2^3 x_3 - \lambda(x_1 + x_2 + x_3 - 12).$$

The 1st order conditions are

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial x_1} = 2x_1 x_2^3 x_3 - \lambda = 0, & (i) \\ \frac{\partial \mathcal{L}}{\partial x_2} = 3x_1^2 x_2^2 x_3 - \lambda = 0, & (ii) \qquad (i) + (ii) \Longrightarrow x_2 = 3x_1/2; \\ \frac{\partial \mathcal{L}}{\partial x_3} = x_1^2 x_2^3 - \lambda = 0, & (iii) \qquad (i) + (iii) \Longrightarrow x_3 = x_1/2. \\ x_1 + x_2 + x_3 = 12, & (iv) \end{cases}.$$

Inserting x_2 and x_3 in (iv) $\Longrightarrow$

$$x_1 + 3x_1/2 + x_1/2 = 12 \quad \Longrightarrow$$
$$x_1 = 4,\ x_2 = 6,\ x_3 = 2.$$

The Constraint Qualification in this (as well as in any other) point holds:

$$\frac{\partial g}{\partial x_1} = \frac{\partial g}{\partial x_2} = \frac{\partial g}{\partial x_3} = 1.$$

Answer: *The only possible* ***solution*** *is $(4, 6, 2)$, which is the global* max *point.*

Example 4.2.3 (Harder Example)**.** *Determine the max/min of the objective function*

$$f(x, y) = x^2 + y^2 \quad (\textit{square of distance from } (0,0) \textit{ in } \mathbb{R}^2)$$

subject to $g(x, y) = x^2 + xy + y^2 - 3 = 0$.

Solution 4.2.4. *The constraint $g(x, y) = 0$ defines an ellipse in $\mathbb{R}^2$, so we should find points of the ellipse which have the minimal distance from $(0, 0)$. The Lagrangian is*

$$\mathcal{L}(x, y) = x^2 + y^2 - \lambda(x^2 + xy + y^2 - 3), \ \textit{for } (x, y) \in \mathbb{R}^2.$$

The 1st order conditions are

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial x} = 2x - \lambda(2x + y), & (i) \\ \frac{\partial \mathcal{L}}{\partial y} = 2y - \lambda(x + 2y), & (ii) \\ x^2 + xy + y^2 - 3 = 0. & (iii) \end{cases}$$

We then have $(i) \Longrightarrow \lambda = \frac{2x}{2x+y}$ ***if*** *$y \neq -2x$. Inserting λ in (ii), we get*

$$2y = \frac{2x}{2x + y}(x + 2y) \quad \Longrightarrow \quad y^2 = x^2 \quad \Longleftrightarrow \quad x = \pm y.$$

(a) ***Suppose*** $y = x$. *Then* (*iii*) $\Longrightarrow x^2 = 1$, *so* $x = 1$ *or* $x = -1$. *We have 2 solution* ***candidates:*** $(x, y) = (1, 1)$ *and* $(x, y) = (-1, 1)$ *for* $\lambda = 2/3$.

(b) ***Suppose*** $y = -x$. *Then* (*iii*) $\Longrightarrow x^2 = 3$, *so* $x = \sqrt{3}$ *or* $x = -\sqrt{3}$. *We have 2 solution* ***candidates:*** $(x, y) = (\sqrt{3}, -\sqrt{3})$ *and* $(x, y) = (-\sqrt{3}, \sqrt{3})$ *for* $\lambda = 2$.

(c) *It remains to consider* $y = -2x$. *Then* (*i*) $\Longrightarrow x = y = 0$, *which contradicts* (*iii*). *So, we have 4 candidates for the max/min problem:*

$$f_{\min} = f(1, 1) = f(-1, -1) = 2 \quad \text{and} \quad f_{\max} = f(\sqrt{3}, -\sqrt{3}) = f(-\sqrt{3}, \sqrt{3}) = 6.$$

Next, we check the constraint qualification at these points: $\nabla g(x, y) = (2x + y, 2y + x) \neq 0$. *The only point where* $\nabla g(x, y) = 0$ *is* $x = y = 0$, *but it does not satisfy the constraint* $g(x, y) = 0$.
Answer: $(1, 1)$ *and* $(-1, 1)$ *solve the* min *problem;* $(\sqrt{3}, -\sqrt{3})$ *and* $(-\sqrt{3}, \sqrt{3})$ *solve the* max *problem.*

Example 4.2.5 (Economic Example)**.** *Suppose we have* n *resources with units* $x_1, \ldots, x_n \geq 0$ *and* m *consumers with their* ***utility*** *functions*

$$u_1(x), \ldots, u_m(x), \quad x = (x_1, \ldots, x_n) \in \mathbb{R}^n_+.$$

The vector $x_i := (x_{i1}, \ldots, x_{in}) \in \mathbb{R}^n_+$ *describes the* ***allocation*** *received by the* i*th consumer,* $1 \leq i \leq m$.
Problem: *Find*

$$\max_{x_1, \ldots, x_n \in \mathbb{R}^n_+} \sum\nolimits_{i=1}^m u_i(x_i)$$

under the ***resourse constraint***

$$\begin{aligned} \sum\nolimits_{i=1}^m x_i &= \omega \in \mathbb{R}^n_+ \quad \text{(a given endowment vector), i.e.,} \\ \sum\nolimits_{i=1}^m x_{ij} &= \omega_j \geq 0, \quad 1 \leq j \leq n. \end{aligned}$$

Solution 4.2.6. *The Weierstrass theorem says that the global maximum exists if each of the functions* $u_1(x), \ldots, u_m(x)$ *are continuous. The* ***Lagrangian*** *with the multiplier vector* $\lambda \in \mathbb{R}^n$

$$\begin{aligned} \mathcal{L}(x_1, \ldots, x_n) &= \sum\nolimits_{i=1}^m u_i(x_i) - \left\langle \lambda, \sum\nolimits_{i=1}^m x_i - \omega \right\rangle \\ &= \sum\nolimits_{i=1}^m u_i(x) - \sum\nolimits_{j=1}^n \lambda_j \left(\sum\nolimits_{i=1}^m x_{ij} - \omega_j \right). \end{aligned}$$

1st order conditions

$$\begin{aligned} \frac{\partial u_i}{\partial x_{ij}}(x_i) &= \lambda_j \quad \text{(independent of } i\text{)} \quad \Longrightarrow \\ \frac{\frac{\partial u_i}{\partial x_{ij}}(x_i)}{\frac{\partial u_i}{\partial x_{ik}}(x_i)} &= \frac{\lambda_j}{\lambda_k}, \quad \text{for any pair of resourses } k, j \text{ and any consumer } i. \end{aligned}$$

The left-hand side is the so-called ***marginal rate of substitution (MRS)*** *of resourse k for resourse j. This relation is the same for all consumers, $1 \le i \le m$.*

4.3 Sufficient Conditions

4.3.1 Global Sufficient Conditions

The Lagrange multiplier method gives the **necessary** conditions. They also will be **sufficient** in the following special case.

Concave / Convex Lagrangian

Let everything be as in Theorem 4.2. Namely, let $U \subset \mathbb{R}^n$ be open and let

$$f : U \to \mathbb{R}, \quad g : U \to \mathbb{R}^m \quad (m \le n)$$

be continuously differentiable. Consider the Lagrangian

$$\mathcal{L}(x; \lambda) := f(x) - \sum\nolimits_{i=1}^{m} \lambda_i g_i(x).$$

Let $(x^*, \lambda^*) \in U \times \mathbb{R}$ be a **critical** point of $\mathcal{L}(x; \lambda)$, i.e., it satisfies the 1st order conditions.

Theorem 4.3.1. *The following hold.*

i. If $\mathcal{L}(x; \lambda^*)$ is a ***concave*** *function of $x \in U$, then x^* is the global maximum.*

ii. If $\mathcal{L}(x; \lambda^*)$ is a ***convex*** *function of $x \in U$, then x^* is the global minimum.*

Proof. By assumption, x obeys the constraint $g(x) = 0$. Let $\mathcal{L}(x; \lambda^*)$ be concave on U. Then by Theorem 3.6, for any $x \in U$, one has

$$\begin{aligned} h(x) = \mathcal{L}(x; \lambda^*) &\le \mathcal{L}(x^*; \lambda^*) + \langle \nabla_x \mathcal{L}(x^*; \lambda^*), x - x^* \rangle_{\mathbb{R}^n} \\ &= \mathcal{L}(x^*; \lambda^*) + \left\langle \nabla f(x^*) - \sum\nolimits_{i=1}^{m} \lambda_i^* \nabla g_i(x^*), x - x^* \right\rangle_{\mathbb{R}^n} \\ &= \mathcal{L}(x^*; \lambda^*) + 0 = f(x^*) - \langle \lambda^*, g(x^*) \rangle_{\mathbb{R}^n} = f(x^*) - 0 \\ &= f(x^*). \end{aligned}$$

□

Remark 4.3.2. *In particular, Theorem 4.3.1 holds if f is concave, g is convex and $\lambda^* \ge 0$. Furthermore, all this applies to linear f, g which are both convex and concave.*

Example 4.3.3 (Economic Example)**.** *A firm uses inputs $K > 0$ of capital and $L > 0$ of labour, respectively, to produce a single output Q according to the Cobb-Douglas production function*

$$Q = K^a L^b,$$

where

$$a, b > 0 \qquad \text{and } a + b \leq 1.$$

*The prices of capital and labour are $r > 0$ and $w > 0$, respectively. Solve the **cost minimising problem***

$$\min \{rK + wL\} \quad \text{subject to} \quad K^a L^b = Q.$$

Solution 4.3.4. *The Lagrangian is*

$$\mathcal{L}(K, L) = rK + wL - \lambda \left(K^a L^b - Q \right).$$

Note that

$$f(K, L) := rK + wL \text{ is linear and } g(k, L) := K^a L^b - Q \text{ is concave.}$$

The 1st order conditions are necessary and sufficient:

$$\begin{cases} r = \lambda a K^{a-1} L^b, \\ w = \lambda b K^a L^{b-1}, \\ K^a L^b = Q, \end{cases} \quad \Longrightarrow \quad \begin{cases} \lambda \geq 0, \\ \frac{r}{w} = \frac{aL}{bK} \Rightarrow L = K \frac{br}{aw}, \\ K^{a+b} = Q \left(\frac{aw}{br} \right)^b. \end{cases}$$

Answer:

$$K = Q^{\frac{1}{a+b}} \left(\frac{aw}{br} \right)^{\frac{b}{a+b}}, \; L = K \frac{br}{aw} = Q^{\frac{1}{a+b}} \left(\frac{br}{aw} \right)^{\frac{a}{a+b}}$$

is the global solution of the Lagrange min problem.

4.3.2 Local Sufficient Conditions of 2nd Order

Theorem 4.3.5. *Let $U \subset \mathbb{R}^n$ be open and let*

$$f : U \to \mathbb{R}, \quad g : U \to \mathbb{R}^m \quad (m \leq n)$$

*be **twice continuously differentiable**. Define the Lagrangian*

$$\mathcal{L}(x; \lambda) := f(x) - \langle \lambda, g(x) \rangle_{\mathbb{R}^m}.$$

Let $x^* \in U$ be such that $g(x^*) = 0$ and

$$D_x \mathcal{L}(x; \lambda^*) = \underbrace{\nabla f(x^*)}_{1 \times n} - \underbrace{\lambda^*}_{1 \times m} \times \underbrace{Dg(x^*)}_{m \times n} = 0$$

*for some Lagrange multiplier $\lambda^* \in \mathbb{R}^m$, i.e., (x^*, λ^*) is a critical point of $\mathcal{L}(x; \lambda)$. Consider the **matrix of 2nd partial derivatives** of $\mathcal{L}(x; \lambda^*)$ w.r.t. x*

$$D_x^2 \mathcal{L}(x; \lambda^*) := \underbrace{D^2 f(x)}_{n \times n} - \underbrace{\lambda^*}_{1 \times m} \times \underbrace{D^2 g(x^*)}_{m \times (n \times n)}.$$

*Suppose that $D_x^2 \mathcal{L}(x; \lambda^*)$ is **negative definite subject to the constraint***

$$\underbrace{Dg(x^*)}_{m \times n} \times \underbrace{h}_{n \times 1} = 0,$$

i.e., for all $x \in U$ and for all $0 \neq h \in \mathbb{R}^n$, one has

$$\left\langle D_x^2 \mathcal{L}(x; \lambda^*) h, h \right\rangle_{\mathbb{R}^n} < 0$$

*from the **linear constraint subspace** $\mathcal{Z}(x^*) := \{h \in \mathbb{R}^n \mid Dg(x^*)h = 0\}$. Then x^* is a **strict local maximum** of $f(x)$ subject to $g(x) = 0$ (i.e., there exists a ball $B_\varepsilon(x^*) \subset U$ such that $f(x^*) > f(x)$ for all $x \in B_\varepsilon(x^*)$ satisfying the constraint $g(x) = 0$).*

Proof. (Idea.) By Taylor's formula and the IFT. See e.g. *Simon, Blume*, Sect. 19.3, or *Sundarem*, Sect. 5.3 . □

Example 4.3.6 (Illustrative Example with $n = 2$, $m = 1$ (see Section 4.2))**.** *Find local max / min of*

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

subject to

$$g(x, y) = x^2 + xy + y^2 - 3 = 0.$$

Solution 4.3.7. *We have seen that the 1st order conditions give 4 candidates*

$$\begin{aligned} (1,1),\ (-1,-1) \quad \textit{with}\ \lambda &= 2/3, \\ (\sqrt{3}, -\sqrt{3}),\ \ (-\sqrt{3}, \sqrt{3}) \quad \textit{with}\ \lambda &= 2. \end{aligned}$$

Calculate

$$\nabla g(x, y) = (2x + y, 2y + x),$$
$$\mathcal{L}(x, y) = x^2 + y^2 - \lambda(x^2 + xy + y^2 - 3),$$
$$D^2 \mathcal{L}(x, y) = \begin{pmatrix} 2 - 2\lambda & -\lambda \\ -\lambda & 2 - 2\lambda \end{pmatrix}.$$

i. Let $x^* = y^* = 1$, $\lambda^* = 2/3$, *and* $h = (h_1, h_2) \neq 0$.

$$\begin{aligned} \nabla g(x^*, y^*) &= (3,3), \\ \langle \nabla g(x^*, y^*), h \rangle &= 0 \iff 3h_1 + 3h_2 = 0 \iff h_1 = -h_2. \\ \left\langle D_x^2 \mathcal{L}(x;\lambda^*)h, h \right\rangle_{\mathbb{R}^n} &= (2-2\lambda^*)h_1^2 - 2\lambda^* h_1 h_2 + (2-2\lambda^*)h_2^2 \\ &= 8h_1^2/3 > 0 \quad (\text{for } h \neq 0). \end{aligned}$$

By Theorem 4.3.5, $x^* = y^* = 1$ *is a local min. The same holds for* $x^* = y^* = -1$.

ii. Let $x^* = -y^* = \sqrt{3}$, $\lambda^* = 2$, *and* $h = (h_1, h_2) \neq 0$.

$$\begin{aligned} \nabla g(x^*, y^*) &= (\sqrt{3}, -\sqrt{3}), \\ \langle \nabla g(x^*, y^*), h \rangle &= 0 \iff h_1 = h_2. \\ \left\langle D_x^2 \mathcal{L}(x;\lambda^*)h, h \right\rangle_{\mathbb{R}^n} &= -8h_1^2 < 0 \quad (\text{for } h \neq 0). \end{aligned}$$

By Theorem 4.3.5, $x^* = \sqrt{3}$, $y^* = -\sqrt{3}$ *is a local maximum. The same holds for* $x^* = -\sqrt{3}$, $y^* = \sqrt{3}$.

4.4 Non-linear Programming and the (Karush-) Kuhn-Tucker Theorem: Optimization under Inequality Constraints

In economics one usually meets inequality than equality constraints (certain variables should be nonnegative, budget constraints, etc.).

Formulation of the problem

Let $U \subset \mathbb{R}^n$ be an open set and let $n, m \in \mathbb{N}$ (not necessarily $m \leq n$). Find

$$\max_{x \in U} f(x_1, \ldots, x_n)$$

subject to m **inequality constraints**

$$\begin{cases} g_1(x_1, \ldots, x_n) \leq 0, \\ \qquad \vdots \\ g_m(x_1, \ldots, x_n) \leq 0. \end{cases}$$

The points $x \in U$ which satisfy these constraints are called **admissible** or **feasible**. Respectively,

$$D := \{x \in U \mid g_1(x) \leq 0, \ldots, g_m(x) \leq 0\}$$

is called **admissible** or **feasible set**.

A point $x^* \in U$ is called a **local maximum** (resp. **minimum**) of f under the above inequality constraints, if there exists a ball $B_\varepsilon(x^*) \subset U$ such that $f(x^*) \geq f(x)$ (resp. $f(x^*) \leq f(x)$) for all $x \in D \cap B_\varepsilon(x^*)$.

Remark 4.4.1. *In general, it is possible that $m > n$, since we have some inequality constraints. For the sake of concreteness we consider only the constraints with "$\leq$".*

In principle, the problem can be solved by the Lagrange method. We have to examine the critical points of $\mathcal{L}(x_1, \ldots, x_n)$ in the interior of the domain D and the behaviour of $f(x_1, \ldots, x_n)$ on the boundary of D. However, since the 1950s, the economists generally tacked this such problems by using an extension of the Lagrange multiplier method due to **Karush–Kuhn–Tucker**.

4.4.1 Karush-Kuhn-Tucker (KKT) Theorem

Albert Tucker (1905–1995) was a Canadian-born American mathematician who made important contributions in topology, game theory, and non-linear programming. He chaired the mathematics department of the Princeton University for about 20 years, one of the longest tenures.

Harold Kuhn (born 1925) is an American mathematician who studied game theory. He won the 1980 John von Neumann Theory Prize along with David Gale and Albert Tucker. He is known for his association with *John Nash*, as a fellow graduate student, a lifelong friend and colleague, and a key figure in getting Nash the attention of the Nobel Prize committee that led to Nash's 1994 Nobel Prize in Economics. Kuhn and Nash both had long associations and collaborations with A. Tucker, who was Nash's dissertation advisor. Kuhn is credited as the mathematics consultant in the 2001 movie adaptation of Nash's life, "A Beautiful Mind".

William Karush (1917–1997) was a professor of California State University and is a mathematician best known for his contribution to Karush–Kuhn–Tucker conditions. He was the first to publish the necessary conditions for the inequality constrained problem in his Masters thesis (Univ. of Chicago, 1939), although he became renowned after a seminal conference paper by Kuhn and Tucker (1951).

Definition 4.4.2. *We say that the inequality constraint $g_i(x) \leq 0$ is **effective** (or **active**, **binding**) at a point $x^* \in U$ if $g_i(x^*) = 0$. Respectively, the constraint $g_i(x) \leq 0$ is **passive** (**inactive, not binding**) at a point $x^* \in U$ if $g_i(x^*) < 0$.*

Intuitively, only *active* constraints have effect on the local behaviour of an optimal solution. If we know from the beginning which restrictions would be binding at an optimum, the Karush-Kuhn-Tucker problem would reduce to a Lagrange problem, in which we would take the active constraints as equalities and ignore the rest.

Theorem 4.4.3 (Karush-Kuhn-Tucker Theorem or the 1st Order Necessary Conditions for Optima; without proof here)**.** *Let $U \subset \mathbb{R}^n$ be open and let $f : U \to \mathbb{R}$ and $g : U \to \mathbb{R}^m$, with $m, n \in \mathbb{N}$, be continuously differentiable. Suppose that $x^* = (x_1^*, \ldots, x_n^*) \in U$ is a* ***local maximum*** *for $f(x_1, \ldots, x_n)$ under the inequality constraints*

$$\begin{cases} g_1(x_1, \ldots, x_n) \leq 0, \\ \qquad \vdots \\ g_m(x_1, \ldots, x_n) \leq 0. \end{cases}$$

Without loss of generality, suppose that the first p $(0 \leq p \leq m)$ constraints are ***active*** *at point x^*, while the others are inactive.*

Furthermore, suppose that the ***Constraint Qualification (CQ)*** *holds: the rank of the Jacobian matrix of the binding constraints (which is a $p \times n$ matrix)*

$$Dg_{\leq p}(x^*) = \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(x^*) & \cdots & \frac{\partial g_1}{\partial x_n}(x^*) \\ \vdots & & \vdots \\ \frac{\partial g_p}{\partial x_1}(x^*) & \cdots & \frac{\partial g_p}{\partial x_n}(x^*) \end{pmatrix}$$

is equal to p, i.e.,

$$\text{rank } Dg_{\leq p}(x^*) = p.$$

Then there ***exists a nonnegative vector*** *$\lambda^* = (\lambda_1^*, \ldots, \lambda_m^*) \in \mathbb{R}_+^m$ such that (x^*, λ^*) satisfy the following conditions hold:*

$$\begin{aligned} [\mathbf{KKT-1}] \quad \frac{\partial f}{\partial x_j}(x^*) &= \sum_{i=1}^m \lambda_i^* \frac{\partial g_i}{\partial x_j}(x^*), \quad \textit{for all } 1 \leq j \leq n; \\ [\mathbf{KKT-2}] \quad \lambda_i^* g_i(x^*) &= 0, \quad \textit{for all } 1 \leq i \leq m. \end{aligned}$$

Remark 4.4.4.

i. $[\mathbf{KKT-2}]$ *is called the* ***"Complementary Slackness"*** *condition: if one of the inequalities*

$$\lambda_i^* \geq 0 \quad \textit{or} \quad g_i(x^*) \leq 0$$

is slack (i.e., strict), the other cannot be!

$$\begin{cases} \lambda_i^* > 0 \implies g_i(x^*) = 0, \\ g_i(x^*) < 0 \implies \lambda_i^* = 0. \end{cases}$$

It is also possible that both $\lambda_i^* = g_i(x^*) = 0$.

ii. The ***Constraint Qualification (CQ)*** *claims that the matrix $Dg_{\leq p}(x^*)$ is of full rank p, i.e., there is no redundant binding constraints, both in the sense that there are fewer binding constraints than variables (i.e., $p \leq n$) and in the sense that the constraints which are binding are 'independent' (otherwise, $Dg_{\leq p}(x^*)$ cannot have the full rank p).*

By changing $\min f = \max(-f)$, we get the following:

Corollary 4.4.5. *Suppose f, g are defined as in Theorem 4.4.3 and $x^* \in U$ is a* ***local minimum****. Then the statement of Theorem 4.5 holds true with the only modification*

$$[\mathbf{KKT-1'}] \qquad \frac{\partial f}{\partial x_j}(x^*) = -\sum\nolimits_{i=1}^m \lambda_i^* \frac{\partial g_i}{\partial x_j}(x^*), \quad \textit{for all } 1 \le j \le n.$$

4.5 A "Cookbook" Procedure: How to use the Karush–Kuhn–Tucker Theorem

(1) Set up the Lagrangian function

$$U \ni (x_1, \dots, x_n) \to \mathcal{L}(x_1, \dots, x_n) := f(x_1, \dots, x_n) - \sum\nolimits_{i=1}^m \lambda_i g_i(x_1, \dots, x_n)$$

with a vector of non-negative Lagrange multipliers $\lambda = (\lambda_1, \dots, \lambda_m) \in \mathbb{R}_+^m$ **(i.e., all** $\lambda_i \ge 0$, $1 \le i \le m$).

(2) Equate all 1st order partial derivatives of $\mathcal{L}(x_1, \dots, x_n)$ **w.r.t.** x_j, $1 \le j \le n$, **to zero:**

$$[\mathbf{KKT-1}] \qquad \frac{\partial}{\partial x_j}\mathcal{L}(x_1, \dots, x_n) = \frac{\partial f}{\partial x_j}(x_1, \dots, x_n) - \sum\nolimits_{i=1}^m \lambda_i \frac{\partial g_i}{\partial x_j}(x_1, \dots, x_n) = 0.$$

(3) Require $(x_1, \dots, x_n)$ **to satisfy the constraints**

$$-\frac{\partial}{\partial \lambda_i}\mathcal{L}(x_1, \dots, x_n) = g_i(x_1, \dots, x_n) \le 0, \quad 1 \le i \le m.$$

Impose the Complementary Slackness Condition

$$\begin{gathered}[\mathbf{KKT-2}] \quad \lambda_i g_i(x_1, \dots, x_n) = 0, \quad 1 \le i \le m, \\ \text{whereby} \ \ \lambda_i = 0 \ \text{ if } \ g_i(x_1, \dots, x_n) < 0 \\ \text{and } g_i(x_1, \dots, x_n) = 0 \ \text{ if } \ \lambda_i > 0.\end{gathered}$$

(4) Find all $x^* = (x_1^*, \dots, x_n^*) \in U$ **which together with the corresponding values of** $\lambda_1^*, \dots, \lambda_m^*$ **satisfy Conditions** $[\mathbf{KKT-1}]$, $[\mathbf{KKT-2}]$. **These are the maxima solution candidates, at least one of which solves the problem (if it has a solution at all). For such** x^* **we should check the Constraint Qualification** $\operatorname{rank} Dg_{\le p}(x^*) = p$, **otherwise the method can give a wrong answer.**

Finally, compute all points $x \in U$ **where the Constraint Qualification fails and compare values of** f **at such points**.

4.5.1 Remarks on Applying KKT Method

(1) The **sign** of λ_i is **important**. The multipliers $\lambda_i^* \geq 0$ correspond to the inequality constraints $g_i(x) \leq 0, \quad 1 \leq i \leq m$. Constraints $g_i(x) \geq 0$ formally lead to the multipliers $\lambda_i^* \leq 0$ in $[\mathbf{KKT-1}]$ (by setting $\tilde{g}_i := -g_i$).

(2) $\lambda_i^* \geq 0$ correspond to the **maximum** problem

$$\max_{x \in U;\ g_1(x) \leq 0, \ldots, g_m(x) \leq 0} f(x).$$

In turn, the **minimum** problem

$$\min_{x \in U;\ g_1(x) \leq 0, \ldots, g_m(x) \leq 0} f(x)$$

leads to the following modification of the $[\mathbf{KKT-1}]$ (by setting $\tilde{f} := -f$)

$$[\mathbf{KKT-1'}] \quad \frac{\partial f}{\partial x_j}(x^*) = -\sum_{i=1}^{m} \lambda_i^* \frac{\partial g_i}{\partial x_j}(x^*), \quad \textit{for all } 1 \leq j \leq n.$$

(3) Intuitively, the λ_i means the **sensitivity** of the objective function $f(x)$ w.r.t. a "small" increase of the parameter c_i in tne constraint $g_i(x) \leq c_i$.

(4) Possible reasons leading to **failure** of the Karush-Kuhn-Tucker method:

i. The **Constraint Qualification fails**. Even if an optimum x^* does exit but does not obey **CQ**, it may happen that x^* does not satisfy $[\mathbf{KKT-1}]$, $[\mathbf{KKT-2}]$.

ii. There exists **no global optimum** for the constrained problem at all. Then there may exist solutions to $[\mathbf{KKT-1}]$, $[\mathbf{KKT-2}]$, which are however not global, or maybe even local, optima.

Example 4.5.1 (Worked Examples (with $n = 2$, $m = 1$))**.** *Solve the problem:*

$$\begin{aligned} \max f(x,y) \quad \textit{for} \quad f(x,y) &= x^2 + y^2 + y + 1 \\ \textit{subject to} \quad g(x,y) &= x^2 + y^2 - 1 \leq 0. \end{aligned}$$

Solution 4.5.2. *By the Weierstrass Theorem there exists a global maximum $(x^*, y^*) \in D$ of $f(x,y)$ in the closed bounded domain (unit ball)*

$$D := \left\{ (x,y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1 \right\}.$$

The Lagrangian is defined for all $(x,y) \in \mathbb{R}^2 := U$ by

$$\mathcal{L}(x,y) := x^2 + y^2 + y + 1 - \lambda(x^2 + y^2 - 1).$$

$$[\mathbf{KKT-1}] \quad \begin{cases} \frac{\partial \mathcal{L}(x,y)}{\partial x} = 2x - 2\lambda x = 0, & (i) \\ \frac{\partial \mathcal{L}(x,y)}{\partial y} = 2y + 1 - 2\lambda y = 0. & (ii) \end{cases},$$

$$[\mathbf{KKT-2}] \quad \begin{cases} \lambda \geq 0, \quad x^2+y^2 \leq 1, \\ \lambda = 0 \quad \text{if } x^2+y^2<1, \quad x^2+y^2=1 \quad \text{if} \quad \lambda>0. \end{cases} \qquad (iii).$$

We should find all $(x^*, y^*) \in D$ which satisfy $(i) - (iii)$ for some $\lambda \geq 0$.

$$(i) \iff 2x(1-\lambda) = 0 \iff \lambda = 1 \;\; or \;\; x = 0.$$

But $\lambda = 1 \underbrace{\Longrightarrow}_{(ii)} 2y + 1 - 2y = 0$, *contradiction.* ***Hence***, $x = 0$.

(a) Suppose $x^2 + y^2 = 1 \iff y = \pm 1$.

If $y = 1 \underbrace{\Longrightarrow}_{(ii)} \lambda = 3/2$, *which solves (iii).*

If $y = -1 \underbrace{\Longrightarrow}_{(ii)} \lambda = 1/2$, *which solves (iii).*

(b) Suppose $x^2 + y^2 < 1$; $x = 0 \Longrightarrow -1 < y < 1$, $\lambda = 0$.

Then by (ii) $y = -1/2$.

We get ***3 candidates****:*

1. $(0, 1)$ *with* $\lambda = 3/2$ *and* $f(0, 1) = 3$;
2. $(0, -1)$ *with* $\lambda = 1/2$ *and* $f(0, -1) = 1$;
3. $(0, -1/2)$ *with* $\lambda = 0$ *and* $f(0, -1/2) = 3/4$.

The point $(0, -1/2)$ is inside D, i.e., the constraint is not active.

At the points $(0, 1)$ and $(0, -1)$ the constraint is active, but $\nabla g(x, y) = (2x, 2y) \neq 0$ and rank$Dg(x, y) = 1$, i.e., ***(CQ)*** *holds.*

The only point, where ***(CQ)*** *could fail, i.e., $\nabla g(x, y) = 0$, is $x = y = 0$ with $f(0, 0) = 1$. But this point is inside D, i.e. $g(0, 0) < 0$, and hence the constraint is passive.*

Answer: $x = 0$, $y = 1$ *is the solution (global maximum).*

Example 4.5.3 (Counterexample (KKT method fails))**.** *Find* $\max f(x, y)$ *for*

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

subject to

$$g(x, y) = y^2 - (x - 1)^3 \leq 0.$$

Solution 4.5.4. *Elementary analysis:* $y^2 \le (x-1)^3 \implies x \ge 1$. *In particular, the smallest possible value of* x *is 1, which corresponds to* $y = 0$. *So,*

$$\max_{g(x,y)\le 0} f(x,y) = -\min_{g(x,y)\le 0}(x^2+y^2) = -1$$

is achieved at $x^* = 1$, $y^* = 0$.

Now, we try to apply the Karush-Kuhn-Tucker method. First we note that $g(x^*,y^*) = 0$ *and*

$$\nabla g(x^*,y^*) = (\partial_x g(x^*,y^*), \partial_y g(x^*,y^*)) = (0,0),$$

i.e., the Constrained Qualification ***fails****. Formally, we should find* $\lambda^* \ge 0$ *such that*

$$\begin{cases} \partial_x f(x^*,y^*) = \lambda^* \partial_x g(x^*,y^*) = 0, \\ \partial_y f(x^*,y^*) = \lambda^* \partial_y g(x^*,y^*) = 0, \end{cases}$$

but we see that $\nabla f(x^*,y^*) = (-2x^*, -2y^*) = (-2,0) \ne 0$. *The Kuhn-Tucker method gives no solutions / critical points, hence it is not applicable. On the other hand, elementary analysis gives us the global maximum at the above point* $x^* = 1$, $y^* = 0$.

4.5.2 The Simplest Case of KKT: $n = 2, m = 1$

Problem: Maximize $f(x,y)$ subject to $g(x,y) \le 0$.

Corollary 4.5.5 (Karush-Kuhn-Tucker Theorem with one inequality constraint)**.** *Let* U *be an open subset of* $\mathbb{R}^2$ *and let* $f : U \to \mathbb{R}$ *and* $g : U \to \mathbb{R}$ *be continuously differentiable. Suppose that* $(x^*,y^*) \in U$ *is a* ***local maximum*** *for* $f(x,y)$ *under the inequality constraint* $g(x,y) \le 0$.

If $g(x^*,y^*) = 0$ *(i.e., the constraint* g *is* ***active*** *at point* (x^*,y^*)*), suppose* ***additionally*** *that rank* $Dg(x^*) = \nabla g(x^*) = 1$, *i.e.,*

$$\frac{\partial g}{\partial x}(x^*,y^*) \ne 0 \quad \text{or} \quad \frac{\partial g}{\partial y}(x^*,y^*) \ne 0,$$

i.e., the ***Constraint Qualification (CQ)*** *holds.*

In any case, form the Lagrangian function

$$\mathcal{L}(x,y) := f(x,y) - \lambda g(x,y).$$

Then, there exists a multiplier $\lambda^* \ge 0$ *such that*

$$\begin{aligned} [\mathbf{KKT-1}] \quad \frac{\partial \mathcal{L}}{\partial x}(x^*,y^*) &= \frac{\partial f}{\partial x}(x^*,y^*) - \lambda^* \frac{\partial g}{\partial x}(x^*,y^*) = 0, \\ \frac{\partial \mathcal{L}}{\partial y}(x^*,y^*) &= \frac{\partial f}{\partial y}(x^*,y^*) - \lambda^* \frac{\partial g}{\partial y}(x^*,y^*) = 0; \\ [\mathbf{KKT-2}] \quad \lambda^* \cdot g(x^*,y^*) &= 0, \quad \lambda^* \ge 0, \quad g(x^*,y^*) \le 0. \end{aligned}$$

Remark 4.5.6 (Why Does the Receipe Work? Geometrical Picture ($n = 2, m = 1$))**.** *Since we* ***do not know a priori*** *whether or not the constraint will be binding at the maximizer, we cannot use the only condition* $[\mathbf{KKT-1}]$*, i.e.,* $\partial_x \mathcal{L}(x,y) = \partial_y \mathcal{L}(x,y) = 0$ *that we used with equality constraints. We should complete the statement by the condition* $[\mathbf{KKT-2}]$*, which says that either the* ***constraint is binding*** *or* ***its multiplier is zero*** *(or sometime, both).*

Proof. (Idea of Prooving Theorem 4.4.3.)

Case 1: Passive Constraint $g(x^*, y^*) < 0$**.** The point $p = (x^*, y^*)$ is **inside** the feasible set

$$D := \{(x,y) \in U \mid g(x,y) \leq 0\}.$$

This means that (x^*, y^*) is an interior maximum of $f(x,y)$ and thus

$$\frac{\partial f}{\partial x}(x^*, y^*) = \frac{\partial f}{\partial y}(x^*, y^*) = 0.$$

In this case we set $\lambda^* = 0$.

Case 2: Binding Constraint $g(x^*, y^*) = 0$**.** The point $p = (x^*, y^*)$ is **on the boundary** of the feasible set. In other words, (x^*, y^*) solves the Lagrange problem, i.e., there exists a Lagrange multiplier $\lambda^* \in \mathbb{R}$ such that

$$\frac{\partial f}{\partial x}(x^*, y^*) = \lambda^* \frac{\partial g}{\partial x}(x^*, y^*), \qquad \frac{\partial f}{\partial y}(x^*, y^*) = \lambda^* \frac{\partial g}{\partial y}(x^*, y^*),$$
$$\text{or} \quad \nabla f(x^*, y^*) = \lambda^* \nabla g(x^*, y^*).$$

This time, however, the sign of λ^* is **important**! Let us show that $\lambda^* \geq 0$. Recall from Sect. 2, that $\nabla f(x^*, y^*) \in \mathbb{R}^2$ points in the direction in which f inreases most rapidly at the point (x^*, y^*). In particular, $\nabla g(x^*, y^*)$ points to the set $g(x,y) \geq 0$ and not to the set $g(x,y) \leq 0$. Since (x^*, y^*) maximizes f on the set $g(x,y) \leq 0$, the gradient of f cannot point to the constraint set. If did, we could increase f and still keep $g(x,y) \leq 0$. So, $\nabla f(x^*, y^*)$ must point to the region where $g(x,y) \geq 0$. This means that $\nabla f(x^*, y^*)$ and $\nabla g(x^*, y^*)$ must point in the **same direction**. Thus, if $\nabla f(x^*, y^*) = \lambda^* \nabla g(x^*, y^*)$, the multiplier λ^* must be ≥ 0.

□

Corollary 4.5.7 (Trivial case ($n = m = 1$))**.** *Let* $U \subset \mathbb{R}$ *be open and let* $f, g \in C^1(U)$*. Suppose that* $x^* \in U$ *is a* ***local maximum*** *for* $f(x)$ *under the inequality constraint*

$$g(x) \leq 0.$$

If $g(x^*) = 0$ *(i.e., the constraint is* ***active*** *at* x^**), suppose additionally that*

$$g'(x^*) \neq 0$$

(i.e., the ***CQ*** *holds). Then there exists a multiplier* $\lambda^* \geq 0$ *such that*

$$\begin{array}{lrcl} [KT-1] & f'(x^*) & = & \lambda^* g'(x^*); \\ [KT-2] & \lambda^* g(x^*) & = & 0, \quad \lambda^* \geq 0, \quad g(x^*) \leq 0. \end{array}$$

4.5.3 The case $n = m = 2$

Corollary 4.5.8. *Let* $U \subset \mathbb{R}^2$ *be open and let*

$$f : U \to \mathbb{R}, \quad g_1 : U \to \mathbb{R}, \quad g_2 : U \to \mathbb{R}$$

be continuously differentiable. Suppose that $(x^*, y^*) \in U$ *is a* ***local maximum*** *for* $f(x, y)$ *under the inequality constraints* $g_1(x, y) \leq 0$, $g_2(x, y) \leq 0$.

i. If $g_1(x^*, y^*) = g_2(x^*, y^*) = 0$ *(i.e.,* ***both constraints are active*** *at point* (x^*, y^*)*), suppose additionally that* $\text{rank} Dg(x^*) = 2$, *i.e.,*

$$\det Dg(x^*, y^*) = \begin{vmatrix} \frac{\partial g_1}{\partial x}(x^*, y^*) & \frac{\partial g_1}{\partial y}(x^*, y^*) \\ & \vdots \\ \frac{\partial g_2}{\partial x}(x^*, y^*) & \frac{\partial g_2}{\partial y}(x^*, y^*) \end{vmatrix} \neq 0$$

(i.e., the ***CQ*** *holds).*

ii. If $g_1(x^*, y^*) = 0$ *and* $g_2(x^*, y^*) < 0$, *suppose additionally that* $\text{rank} Dg_1(x^*, y^*) = 1$, *i.e.,* ***at least one*** *of* $\frac{\partial g_1}{\partial x}(x^*, y^*)$ *and* $\frac{\partial g_1}{\partial y}(x^*, y^*)$ *is not zero.*

iii. If $g_1(x^*, y^*) < 0$ *and* $g_2(x^*, y^*) = 0$, *suppose respectively that* $\text{rank} Dg_2(x^*, y^*) = 1$, *i.e.,* ***at least one*** *of* $\frac{\partial g_2}{\partial x}(x^*, y^*)$ *and* $\frac{\partial g_2}{\partial y}(x^*, y^*)$ *is not zero.*

iv. If both $g_1(x^*, y^*) < 0$ *and* $g_2(x^*, y^*) < 0$, *no additional assumptions are needed (i.e., the* ***CQ holds automatically****).*

In any case, form the ***Lagrangian function***

$$\mathcal{L}(x, y) := f(x, y) - \lambda_1 g_1(x, y) - \lambda_2 g_2(x, y).$$

Then there exists a multiplier $\lambda^* = (\lambda_1^*, \lambda_2^*) \in \mathbb{R}^2_+$ *such that:*

$$[\mathbf{KKT-1}] \quad \begin{array}{l} \frac{\partial \mathcal{L}}{\partial x}(x^*, y^*) = \frac{\partial f}{\partial x}(x^*, y^*) - \lambda_1^* \frac{\partial g_1}{\partial x}(x^*, y^*) - \lambda_2^* \frac{\partial g_2}{\partial x}(x^*, y^*) = 0, \\ \frac{\partial \mathcal{L}}{\partial y}(x^*, y^*) = \frac{\partial f}{\partial y}(x^*, y^*) - \lambda_1^* \frac{\partial g_1}{\partial y}(x^*, y^*) - \lambda_2^* \frac{\partial g_2}{\partial y}(x^*, y^*) = 0; \end{array}$$

$$[\mathbf{KKT-2}] \quad \begin{array}{c} \lambda_1^* g_1(x^*, y^*) = 0, \quad \lambda_2^* g_2(x^*, y^*) = 0, \\ \lambda_1^* \geq 0, \quad \lambda_2^* \geq 0, \quad g_1(x^*, y^*) \leq 0, \quad g_2(x^*, y^*) \leq 0. \end{array}$$

Example 4.5.9 (More difficult: $n = m = 2$). *Find* $\min\,(e^{-x} - y)$ *subject to*

$$\begin{cases} e^x + e^y \le 6 \\ y \ge x \end{cases}$$

Solution 4.5.10. *Rewrite the problem as* $\max f(x,y)$ *with*

$$f(x,y) := y - e^{-x} \quad \textit{where} \quad (x,y) \in \mathbb{R}^2 =: U,$$

subject to

$$\begin{cases} g_1(x,y) := e^x + e^y - 6 \le 0 \\ g_2(x,y) := x - y \le 0 \end{cases}.$$

Define the Lagrangian function with $\lambda_1, \lambda_2 \ge 0$

$$\mathcal{L}(x,y) := y - e^{-x} - \lambda_1(e^x + e^y - 6) - \lambda_2(x - y).$$

The 1st order conditions **[KKT-1]**

$$\begin{cases} e^{-x} - \lambda_1 e^x - \lambda_2 = 0, \quad \textit{(i)} \\ 1 - \lambda_1 e^y + \lambda_2 = 0. \quad \textit{(ii)} \end{cases}$$

The Complementary Slackness **[KKT-2]**

$$\begin{cases} \lambda_1(e^x + e^y - 6) = 0; \quad \lambda_1 \ge 0; \quad \lambda_1 = 0 \;\; \textit{if} \;\; e^x + e^y < 6, \quad \textit{(iii)} \\ \lambda_2(x - y) = 0; \quad \lambda_2 \ge 0; \quad \lambda_2 = 0 \;\; \textit{if} \;\; x < y; \quad \textit{(iv)} \\ x \le y, \quad e^x + e^y \le 6. \end{cases}$$

From (ii)

$$\lambda_2 + 1 = \lambda_1 e^y \quad \Longrightarrow \quad \lambda_1 > 0,$$

and then by (iii)

$$e^x + e^y = 6.$$

Suppose in (iv) that $x = y$*, then* $e^x = e^y = 3$*. From (i) and (ii), we get*

$$\begin{cases} \frac{1}{3} - 3\lambda_1 - \lambda_2 = 0, \\ 1 - 3\lambda_1 + \lambda_2 = 0. \end{cases} \quad \Longrightarrow \begin{cases} \lambda_1 = 2/9, \\ \lambda_2 = -1/3, \end{cases}$$

which contradicts to (iv) (since now $\lambda_2 < 0$*). Hence* $x < y$ *and* $\lambda_2 = 0$*, as well as* $e^x + e^y = 6$ *and* $\lambda_1 > 0$.

$$\left.\begin{matrix} \textit{(i)} \;\; \Longrightarrow \;\; \lambda_1 = e^{-2x}, \\ \textit{(ii)} \;\; \Longrightarrow \;\; \lambda_1 = e^{-y} \end{matrix}\right\} \quad \Longrightarrow$$

$$\left.\begin{matrix} y = 2x, \\ e^{2x} + e^x = 6 \end{matrix}\right\} \quad \Longrightarrow \quad e^x = 2 \;\; \textit{or} \;\; e^x = -3 \; \textit{(impossible!)}.$$

So,

$$x^* = \ln 2, \quad y^* = 2x = \ln 4,$$
$$\lambda_1^* = 1/4, \quad \lambda_2^* = 0.$$

We showed that $(x^*, y^*) = (\ln 2, \ln 4)$ *is the* ***only candidate*** *for a solution. At this point the constraint* $g_1(x, y) \leq 0$ *is binding whereas the constraint* $g_2(x, y) \leq 0$ *is passive. The* ***(CQ)*** *now reads as*

$$\frac{\partial g_1}{\partial x}(x^*, y^*) = e^{x^*} \neq 0 \quad \textit{or} \quad \frac{\partial g_1}{\partial y}(x^*, y^*) = e^{y^*} \neq 0$$

*and is satisfied. Actually, (**CQ**) holds at all points* $(x, y) \in \mathbb{R}^2$*. Namely,*

$$Dg(x, y) = \begin{pmatrix} e^x & e^y \\ 1 & -1 \end{pmatrix},$$

with $\det Dg(x, y) = -(e^x + e^y) < 0$ *and* $\nabla g_1(x, y) \neq 0$, $\nabla g_2(x, y) \neq 0$ *for all* $(x, y) \in \mathbb{R}^2$. *The point* $(\ln 2, \ln 4)$ *is the* ***global minimum*** *point we need to find.*