MATHEMATICAL ECONOMICS: OPTIMIZATION

JOÃO LOPES DIAS

Contents

These notes address the topics on "Optimization", the second part of the course "Mathematical Economics". To follow them it is assumed good knowledge of vector differential calculus and matrices.

Date: October 16, 2015.

1. Introduction

1.1. **Preliminaries.** Consider a subset D of $\mathbb{R}^n$ and a scalar function $f\colon D \to \mathbb{R}$. We say that the *image* of D by f is the set

$$f(D) = \{f(x) \in \mathbb{R} \colon x \in D\}$$

and, given $A \subset \mathbb{R}$, the *pre-image* of A by f is

$$f^{-1}(A) = \{x \in D \colon f(x) \in A\}.$$

This last set corresponds to the points in D that are mapped by f into A.

1.2. **Optimal points and values.** A point $x^* \in D$ is a *maximizer* of f on D if for any $x \in D$ it satisfies

$$f(x) \leq f(x^*).$$

The value $f(x^*)$ is the *maximum* of f on D and it is usually denoted by any of the following notations:

$$\max_D f = \max_{x \in D} f(x) = \max\{f(x) \colon x \in D\} = \max f(D).$$

Whenever there is no ambiguity regarding the set D, we simply write $\max f$. Notice that the maximum of f can also be seen as the maximum of the image set $f(D)$. Moreover, it is clear that maximizers and the maximum might not exist.

Similarly, we define a *minimizer* and the *minimum* of f on D writing

$$\min_D f = \min_{x \in D} f(x) = \min\{f(x) \colon x \in D\} = \min f(D).$$

Example 1.1. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ given by $f(x,y) = x^2 + y^2$ and $D = \mathbb{R}^2$. Then $\min_D f = 0$ but there is no maximum. However, if $D = [0,1] \times [0,1]$ we have $\max_D f = 2$.

When restricting to local neighbourhoods, we obtain a more general definition of a maximizer. A point x^* is a *local maximizer* of f on D if there is $\varepsilon > 0$ such that

$$f(x) \leq f(x^*), \quad x \in D \cap B_\varepsilon(x^*),$$

where

$$B_\varepsilon(x^*) = \{x' \in \mathbb{R}^n \colon \|x' - x\| < \varepsilon\}$$

is the open ball centered at x^* with radius ε. If in addition we have the inequality $f(x) < f(x^*)$ when $x \in D \cap B_\varepsilon(x^*) \setminus \{x^*\}$, we say that x^* is a *strict local maximizer*. The image by f of a local maximizer is called a *local maximum*.

Again, we have the corresponding definitions of *local minimizer*, *strict local minimizer* and *local minimum*.

Maximizers and minimizers are called (global) *optimal points* and *local optimal* points in case of being local. The maximum and the minimum are known as (global) *optimal values* and *local optimal values* in the local situation.

1.3. **The optimization problems.** Let D be a subset of $\mathbb{R}^n$ and a scalar function $f\colon D \to \mathbb{R}$. The computation of the (global) optimal points of f on D is called the *optimization problem*. Moreover, the *local optimization problem* is to find the local optimal points of f on D. Observe that the local optimization solutions include the global ones.

In some contexts f is called the *objective function* and D the *constraint set*.

Depending on the choices of f and D there are different names for the optimization problem. Consider f to be a linear function[1] and[2]

$$D = \{x \in \mathbb{R}^n \colon g(x) = a, h(x) \geq b\}$$

where $g\colon \mathbb{R}^n \to \mathbb{R}^m$ and $h\colon \mathbb{R}^n \to \mathbb{R}^\ell$ are also linear functions, $a \in \mathbb{R}^m$ and $b \in \mathbb{R}^\ell$. The corresponding optimization problem is usually called *linear optimization* (or linear programming). Otherwise, it is called *nonlinear optimization* or (nonlinear programming).

Generalizations of linear optimization are the *convex* and the *concave optimizations*. They consist in taking f, g and h as convex functions or concave functions (see section 5 for the definitions).

When restricting to solutions which are integers we are dealing with *integer optimization* (or integer programming).

We present below some examples from Economics.

Example 1.2 (Utility maximization). Consider a consumer buying $x_i \geq 0$ units of the commodity i with price p_i, $i = 1, \dots, n$, and with income $I > 0$. The utility is given by $u(x_1, \dots, x_n)$. We want to know how much should the consumer buy of each commodity in order to maximize the utility.

Since the spending is given by

$$p \cdot x = \sum_{i=1}^{n} p_i x_i,$$

and it should be less or equal than the income I, we choose

$$D = \{x = (x_1, \dots, x_n) \in (\mathbb{R}_0^+)^n \colon p \cdot x \leq I\}.$$

[1]A function $f\colon \mathbb{R}^n \to \mathbb{R}^k$ is linear if $f(\alpha x + \beta x') = \alpha f(x) + \beta f(x')$ for any $x, x' \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$. It thus have to be in the form $f(x) = Mx$ where M is a $k \times n$ matrix.

[2]We write $h(x) \geq b$ to mean that $h_i(x) \geq b_i$ for every $i = 1, \dots, \ell$.

We want to find the maximizers of u on D.

Example 1.3 (Spending minimization). Given an utility value $\bar{u}$, we want to minimize the spending that gives us an utility at least equal to $\bar{u}$.

Writing

$$D = \{x \in (\mathbb{R}_0^+)^n \colon u(x) \geq \bar{u}\}$$

and $f(x) = p \cdot x$, we need to find the minimizers of f on D.

Example 1.4 (Profit maximization). A firm produces y units of one output product which sells at price $p(y)$ using n inputs. Let x_i be the amount of units of the i-th input and w_i its price. So, the cost of the inputs is $w \cdot x = (w_1, \dots, w_n) \cdot (x_1, \dots, x_n)$. We want to find the input quantities x that maximize the profit.

Since the profit function is given by $f(x) = p(g(x)) \cdot g(x) - w \cdot x$ and we write $D = (\mathbb{R}^+)^n$, the goal is to find the maximizers of f on D.

Example 1.5 (Cost minimization). For the previous example, we now want to minimize the cost of producing at least $\bar{y}$ units of output.

So, for $D = \{x \in (\mathbb{R}^+)^n \colon g(x) \geq \bar{y}\}$ and the cost function $f(x) = w \cdot x$, we need to find the minimizers of f on D.

1.4. **Existence of optimal points.** The following theorem is a well-known criterium to determine if there are solutions to the optimization problem.

Theorem 1.6 (Weierstrass). *If $D \subset \mathbb{R}^n$ is compact and $f \colon D \to \mathbb{R}$ is continuous, then there exists the maximum and the minimum of f on D.*

Proof. This follows from the fact that the image of a compact set by a scalar continuous function is a compact set in $\mathbb{R}$. Compact sets in $\mathbb{R}$ have always a maximum and a minimum. We can therefore apply this to $f(D)$ whose extreme points are the maximum and minimum of f on D. □

Remark 1.7. The Weierstrass theorem only states the existence of optimal points, it does not compute them. For that we need to introduce other tools like differential calculus as is done in the next sections. This restricts our study to differentiable functions. In the last sections we will be dealing with non-differentiable functions which are convex or concave.

The following simple examples show that the Weierstrass theorem can not be improved.

Example 1.8. Let $D = \mathbb{R}$ (not bounded) and $f(x) = x$. Then, $f(D) = \mathbb{R}$ and f does not have optimal values on D.

Example 1.9. Let $D =]0,1[$ (not closed) and $f(x) = x$. Then, $f(D) =]0,1[$ and f does not have optimal values on D.

Example 1.10. Let $D = [0,1]$ and

$$f(x) = \begin{cases} x, & x \in]0,1[\\ \frac{1}{2}, & x \in \{0,1\} \end{cases}$$

(not continuous). Then, $f(D) =]0,1[$ and f does not have optimal values on D.

2. Unconstrained local optimization

We want here to find local optimal points inside the interior of D. We first show that optimal points are critical points. By first finding all critical points, we will try to decide which ones are optimal points.

Let $f\colon D \to \mathbb{R}$ be differentiable in the interior of $D \subset \mathbb{R}^n$ (denoted by $\operatorname{Int} D$). Recall that the derivative of f at $x^* \in \operatorname{Int} D$ is a linear function given by the Jacobian matrix:

$$Df(x^*) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x^*) & \cdots & \frac{\partial f}{\partial x_n}(x^*) \end{bmatrix}.$$

Moreover, the second derivative of f at x^* is a quadratic form given by the $n \times n$ Hessian matrix:

$$D^2 f(x^*) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x^*) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x^*) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x^*) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x^*) \end{bmatrix}.$$

We say that f is C^1 if its first partial derivatives are continuous, and C^2 if the second partial derivatives are continuous. Notice that if f is C^2 then the second derivative is a symmetric matrix (Schwarz theorem).

2.1. **Critical points.**

We say that $x^* \in \operatorname{Int} D$ is a *critical point* of f (or stationary point) if $Df(x^*) = 0$.

Theorem 2.1. *If $x^* \in \operatorname{Int} D$ is a local optimal point, then it is a critical point of f.*

Proof. For each $i = 1, \dots, n$ and a sufficiently small $\delta > 0$, let $\psi_i \colon]-\delta, \delta[\to \mathbb{R}$, $\psi_i(t) = f(x^* + te_i)$, where e_i is the i-th vector of the canonical basis of $\mathbb{R}^n$. Since $\psi_i(0) = f(x^*)$ is a local optimal value, then $\psi_i'(0) = 0$. Therefore, $Df(x^*)\, e_i = 0$ for any i. This means that all the columns of $Df(x^*)$ are zero. □

Remark 2.2.

(1) The converse of the above theorem is not true. There are examples of critical points which are not local optimal points. For the function $f(x) = x^3$ the point $x^* = 0$ is critical but not optimal.
(2) The above theorem is restricted to interior points of D where the function is differentiable. For example, $f(x, y) = \sqrt{x^2 + y^2}$ on $D = \{(x, y) \in \mathbb{R}^s : x^2 + y^2 \leq 1\}$ has a minimizer at the origin (which is in the interior of D), but f is not differentiable there.

A critical point which is not a local optimal point is called *saddle point*. Hence, the set of critical points is the union of the sets of local optimal and saddle points:

$$\{\text{critical points}\} = \{\text{local optimal points}\} \cup \{\text{saddle points}\}.$$

Exercise 2.3. Consider the function $f(x, y) = xye^{-(x^2+y^2)/2}$ defined in $\mathbb{R}^2$. Show that its critical points are $(0, 0)$, $(1, 1)$, $(1, -1)$, $(-1, 1)$, $(-1, -1)$. Try to determine if they are optimal points.

Exercise 2.4. The function $g(x, y) = (x^2 - y^2)^2 + x^2y^2$ is always larger or equal to 0. Prove that there is only one critical point $(0, 0)$, and as $g(0, 0) = 0$ we have that the origin is a minimizer.

2.2. **Classification of critical points.** Recall now the following classification of a symmetric $n \times n$ matrix A.

(1) A is **positive definite** ($A > 0$) if $v^T Av > 0$ for any $v \in \mathbb{R}^n \setminus \{0\}$.
(2) A is **negative definite** ($A < 0$) if $v^T Av < 0$ for any $v \in \mathbb{R}^n \setminus \{0\}$.
(3) A is **positive semi-definite** ($A \geq 0$) if $v^T Av \geq 0$ for any $v \in \mathbb{R}^n \setminus \{0\}$.
(4) A is **negative semi-definite** ($A \leq 0$) if $v^T Av \leq 0$ for any $v \in \mathbb{R}^n \setminus \{0\}$.
(5) A is **indefinite** if it is not any of the above.

Notice that the above definitions are equivalent to statements about the signs of the eigenvalues of A. In fact,

(1) $A > 0$ iff all eigenvalues are positive.
(2) $A < 0$ iff all eigenvalues are negative.
(3) $A \geq 0$ iff all eigenvalues are non-negative.
(4) $A \leq 0$ iff all eigenvalues are non-positive.
(5) A is indefinite if it has positive and negative eigenvalues.

We are going to check the above condition for the second derivative of f at critical points.

Theorem 2.5. *If f is C^2 and $x^* \in \operatorname{Int} D$ is a critical point and*

(1) $D^2 f(x^*) > 0$, *then x^* is a strict local minimizer.*

(2) *$D^2 f(x^*) < 0$, then x^* is a strict local maximizer.*
(3) *x^* is a local minimizer, then $D^2 f(x^*) \geq 0$.*
(4) *x^* is a local maximizer, then $D^2 f(x^*) \leq 0$.*
(5) *$D^2 f(x^*)$ is indefinite, then x^* is a saddle point.*

Proof. Take the Taylor expansion of f on a neighborhood $U \subset \operatorname{Int} D$ of x^* given by

$$f(x) - f(x^*) = \frac{1}{2}(x - x^*)^T D^2 f(c)\,(x - x^*)$$

where $x \in U$ and some t is sufficiently small such that $c = x^* + t(x - x^*) \in U$. Notice that we have used the assumption $Df(x^*) = 0$.

First consider the case $D^2 f(x^*) > 0$. Because $D^2 f$ is continuous, we can take U small enough so that $D^2 f(c) > 0$ (notice that the same could have not be done if we were considering a positive semi-definite matrix). Hence, $f(x) - f(x^*) > 0$ when $x \neq x^*$, which means that x^* is a strict local minimizer. The same idea could be used to prove the claims (2-4). (5) follows from (3) and (4) directly. □

Exercise 2.6. Find the optimal points of

(1) $f(x, y, z) = z \log(x^2 + y^2 + z^2)$ on $\mathbb{R}^3 \setminus \{(0, 0, 0)\}$.
(2) $f(x, y, z) = x^4 + y^4 + z^4 - 4xyz$ on $\mathbb{R}^3$.

3. Equality constraints

We now want to find local optimal points of a given function f when restricting to a non open set defined as the zeros of another function g. This will be achieved by the computation and classification of the critical points of a function combining f and g.

3.1. **Lagrange theorem.**

Given an open subset $U \subset \mathbb{R}^n$ and C^1 functions $f\colon U \to \mathbb{R}$ and $g\colon U \to \mathbb{R}^m$, we want to find the optimal points of f on the constraint set given by the zeros of g:

$$D = \{x \in U \colon g(x) = 0\}.$$

Recall that the rank of a matrix A (denoted by $\operatorname{rank} A$) is the number of linear independent rows or columns.

Theorem 3.1 (Lagrange)**.** *If $x^* \in D$ is a local optimal point of f on D and* $\operatorname{rank} Dg(x^*) = m$*, then there is $\lambda^* \in \mathbb{R}^m$ such that*

$$Df(x^*) + \lambda^{*T} Dg(x^*) = 0.$$

Proof. First, reorder the coordinates such that the first m rows of $Dg(x^*)$ are linearly independent. Write the coordinates $x = (w, z) \in \mathbb{R}^m \times \mathbb{R}^{n-m}$ and suppose that $x^* = (w^*, z^*) \in D$ is a local maximizer (we case use the same ideas for a minimizer).

Since rank $\frac{\partial g}{\partial w}(x^*) = m$ we have that det $\frac{\partial g}{\partial w}(x^*) \neq 0$. By the implicit function theorem, there is a C^1 function $h\colon V \to \mathbb{R}^m$ defined on a neighborhood V of z^* such that

- $h(z^*) = w^*$,
- $g(h(z), z) = 0$, $z \in V$,
- $Dh(z^*) = -\frac{\partial g}{\partial w}(x^*)^{-1}\frac{\partial g}{\partial z}(x^*)$.

Choose $\lambda^{*T} = -\frac{\partial f}{\partial w}(x^*)\frac{\partial g}{\partial w}(x^*)^{-1}$, so that

$$\frac{\partial f}{\partial w}(x^*) + \lambda^{*T}\frac{\partial g}{\partial w}(x^*) = 0.$$

Finally, let $F\colon V \to \mathbb{R}$, $F(z) = f(h(z), z)$, which has a local maximum at $z^* \in V$. This is an unconstrained problem since V is open, thus $DF(z^*) = 0$. That is,

$$Df(h(z^*), z^*)\begin{bmatrix} Dh(z^*) \\ I \end{bmatrix} = 0.$$

This yields, after simplification,

$$\frac{\partial f}{\partial z}(x^*) + \lambda^{*T}\frac{\partial g}{\partial z}(x^*) = 0.$$

□

Exercise 3.2. Let $f(x, y) = x^3 + y^3$ and $g(x, y) = x - y$. Solve with respect to (x, y) and λ the equations $f(x, y) + \lambda^T g(x, y) = 0$ and $g(x, y) = 0$.

3.2. **Classification of critical points.** Given a symmetric $n \times n$ matrix A, an $m \times n$ matrix B and

$$\mathcal{Z} = \{v \in \mathbb{R}^n \colon B\, v = 0\}.$$

We have the following classification of A when restricting to $\mathcal{Z}$.

(1) A is **positive definite on** $\mathcal{Z}$ ($A > 0$ on $\mathcal{Z}$) if $v^T Av > 0$ for any $v \in \mathcal{Z} \setminus \{0\}$.
(2) A is **negative definite on** $\mathcal{Z}$ ($A < 0$ on $\mathcal{Z}$) if $v^T Av < 0$ for any $v \in \mathcal{Z} \setminus \{0\}$.
(3) A is **positive semi-definite on** $\mathcal{Z}$ ($A \geq 0$ on $\mathcal{Z}$) if $v^T Av \geq 0$ for any $v \in \mathcal{Z} \setminus \{0\}$.
(4) A is **negative semi-definite on** $\mathcal{Z}$ ($A \leq 0$ on $\mathcal{Z}$) if $v^T Av \leq 0$ for any $v \in \mathcal{Z} \setminus \{0\}$.
(5) A is **indefinite on** $\mathcal{Z}$ if it is not any of the above.

It is clear that if A is positive definite, then it is also positive definite on $\mathcal{Z}$.

There is a criterion for the positive and negative definite cases. For each $i = m+1, \dots, n$ define the $(m+i) \times (m+i)$ symmetric matrix

$$C_i = \begin{bmatrix} 0 & B_i \\ B_i^T & A_i \end{bmatrix},$$

where A_i is the matrix consisting of the first i rows and i columns of A, and B_i is made of the first i columns of B.

(1) $A > 0$ on $\mathcal{Z}$ iff $(-1)^m \det C_i > 0$ for every $i = m+1, \dots, n$.
(2) $A < 0$ on $\mathcal{Z}$ iff $(-1)^i \det C_i > 0$ for every $i = m+1, \dots, n$.

Example 3.3. Let

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 2 \end{bmatrix}.$$

In this case $m = 1$ and $n = 2$. Thus,

$$C_2 = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix}$$

and $(-1)^1 \det C_2 = 5 > 0$. Therefore, $A > 0$ on $\mathcal{Z}$.

Consider f and g of class C^2. Then,

$$A(x, \lambda) = D^2 f(x) + \sum_{i=1}^{m} \lambda_i D^2 g_i(x)$$

is a symmetric $n \times n$ matrix. Choose also $B(x) = Dg(x)$, thus

$$\mathcal{Z}(x) = \{v \in \mathbb{R}^n \colon Dg(x)\, v = 0\}.$$

Theorem 3.4. *Let $x^* \in D$ and $\lambda^* \in \mathbb{R}^m$ such that* $\operatorname{rank} B(x^*) = m$ *and*

$$Df(x^*) + \lambda^{*T} Dg(x^*) = 0.$$

If

(1) *$A(x^*, \lambda^*) > 0$ on $\mathcal{Z}$, then x^* is a strict local minimizer of f on D.*
(2) *$A(x^*, \lambda^*) < 0$ on $\mathcal{Z}$, then x^* is a strict local maximizer of f on D.*
(3) *x^* is a local minimizer of f on D, then $A(x^*, \lambda^*) \geq 0$ on $\mathcal{Z}$.*
(4) *x^* is a local maximizer of f on D, then $A(x^*, \lambda^*) \leq 0$ on $\mathcal{Z}$.*
(5) *$A(x^*, \lambda^*)$ is indefinite on $\mathcal{Z}$, then x^* is a saddle point of f on D.*

Proof. We are going to show that if x^* is not a strict local minimizer of f on D, then there is $y \in \mathcal{Z}(x^*) \setminus \{0\}$ such that $y^T A(x^*, \lambda^*) y \leq 0$. The remaining assertions are left as an exercise.

For $\varepsilon > 0$ there is $x(\varepsilon) \in D \cap B_\varepsilon(x^*) \setminus \{x^*\}$ such that

$$f(x(\varepsilon)) \leq f(x^*).$$

Consider the sequences $\varepsilon_n > 0$ such that $\varepsilon_n \to 0$, $x_n = x(\varepsilon_n) \to x^*$ and

$$y_n = \frac{x_n - x^*}{\|x_n - x^*\|}.$$

As y_n lies inside the compact $C = \{z \in \mathbb{R}^n \colon \|z\| = 1\}$ for any $n \geq 1$, there is a convergent subsequence $y_{k_n} \to y \in C$ with $y \neq 0$.

The Taylor expansion of g on a neighborhood of x^* is given by

$$g(x_{k_n}) = g(x^*) + Dg(c_n)\,(x_{k_n} - x^*),$$

where $c_n = x^* + t(x_{k_n} - x^*)$ and t is sufficiently small. Since $g(x_{k_n}) = g(x^*) = 0$ because they are points in D, we have

$$Dg(c_n)\, y_{k_n} = 0.$$

Since g is C^1 and $c_n \to x^*$ because $x_{k_n} \to x^*$, we have $Dg(x^*)y = 0$ and $y \in \mathcal{Z}(x^*)$.

Consider the C^2 function $L(x) = f(x) + \lambda^{*T} g(x)$ on D by fixing $\lambda = \lambda^*$. Notice that $DL(x^*) = 0$ and Taylor's formula around x^* yields

$$L(x_{k_n}) = L(x^*) + \frac{1}{2}(x_{k_n} - x^*)^T D^2 L(c_n)(x_{k_n} - x^*)$$

for some c_n as before. Hence,

$$\frac{1}{2} y_{k_n}^T D^2 L(c_n) y_{k_n} = \frac{f(x_{k_n}) - f(x^*)}{\|x_{k_n} - x^*\|^2} \leq 0.$$

For $n \to +\infty$ we get $y^T D^2 L(x^*) y \leq 0$. □

Exercise 3.5. Determine the local optimal points of f on D for:

(1) $f(x, y) = xy$,
$$D = \{(x, y) \in \mathbb{R}^2 \colon x^2 + y^2 = 2a^2\}$$
(2) $f(x, y) = 1/x + 1/y$,
$$D = \{(x, y) \in \mathbb{R}^2 \colon (1/x)^2 + (1/y)^2 = (1/a)^2\}$$
(3) $f(x, y, z) = x + y + z$,
$$D = \{(x, y, z) \in \mathbb{R}^3 \colon (1/x) + (1/y) + (1/z) = 1\}$$
(4) $f(x, y, z) = x^2 + 2y - z^2$,
$$D = \{(x, y, z) \in \mathbb{R}^3 \colon 2x - y = 0, x + z = 6\}$$
(5) $f(x, y) = x + y$,
$$D = \{(x, y) \in \mathbb{R}^2 \colon xy = 16\}$$
(6) $f(x, y, z) = xyz$,
$$D = \{(x, y, z) \in \mathbb{R}^3 \colon x + y + z = 5, xy + xz + yz = 8\}$$

4. Inequality constraints

Consider now the problem of finding the optimal points of a function under restrictions given by inequalities.

4.1. **Kuhn-Tucker conditions.**

Given an open set $U \subset \mathbb{R}^n$ and a C^1 function $h\colon U \to \mathbb{R}^\ell$, consider the restriction set

$$D = \{x \in U \colon h(x) \geq 0\}.$$

Notice that we are using the notation $v = (v_1, \dots, v_\ell) \geq 0$ to stand for $v_i \geq 0$ for every $i = 1, \dots, \ell$. We use similarly $v \leq 0$, $v > 0$ and $v < 0$.

Theorem 4.1 (Kuhn-Tucker). *If x^* is a local optimal point of f on D,*

$$h_1(x^*) = \cdots = h_m(x^*) = 0, \quad h_{m+1}(x^*), \dots, h_\ell(x^*) > 0,$$

and $\operatorname{rank} D(h_1, \dots, h_m)(x^*) = m$, *then there is $\lambda^* \in \mathbb{R}^\ell$ such that (Kuhn-Tucker conditions):*

(1) $Df(x^*) + \lambda^{*T} Dh(x^*) = 0$,
(2) $\lambda_i^* h_i(x^*) = 0$, $i = 1, \dots, \ell$.

Moreover,

- *if x^* is a local minimizer, then $\lambda^* \leq 0$.*
- *if x^* is a local maximizer, then $\lambda^* \geq 0$.*

Proof. A simple generalization of Theorem 3.1 implies that under our assumptions there is $\lambda^* \in \mathbb{R}^\ell$ such that $\lambda^*_{m+1} = \cdots = \lambda^*_\ell = 0$ and $Df(x^*) + \lambda^{*T} Dh(x^*) = 0$. Thus we have shown (1) and (2).

We now want to prove that if in addition x^* is a local maximizer, then $\lambda_1^*, \dots, \lambda_m^* \geq 0$. The case of a minimizer follows simply by considering a maximizer for $-f$.

Let $j \in \{1, \dots, m\}$, $e_j = (0, \dots, 0, 1, 0, \dots, 0) \in \mathbb{R}^m$ where the only non-zero component is the j-th and write $H = (h_1, \dots, h_m)$. It is enough to show that there is a C^1 map $\phi\colon]-\delta, \delta[\to \mathbb{R}^n$ for some $\delta > 0$ such that $\phi(0) = x^*$ and $H \circ \phi(t) = t e_j$. Thus, $H \circ \phi(t) \geq 0$ for $t \geq 0$. Since $(h_{m+1}, \dots, h_\ell)(x^*) > 0$, the continuity of h implies for a sufficiently small δ that

$$h(\phi(t)) = (H, h_{m+1}, \dots, h_\ell) \circ \phi(t) \geq 0, \quad t \geq 0.$$

Hence, $\phi(t) \in D$ for $t \geq 0$.

Furthermore, since x^* is a local maximizer of f on D, $f(\phi(t)) - f(\phi(0)) \leq 0$ for $t \geq 0$, and

$$\begin{aligned}
\lambda_j^* &= (\lambda_1^*, \dots, \lambda_m^*) \cdot e_j \\
&= (\lambda_1^*, \dots, \lambda_m^*) \cdot (H \circ \phi)'(0) \\
&= (\lambda_1^*, \dots, \lambda_m^*, 0, \dots, 0) \cdot (h \circ \phi)'(0) \\
&= \lambda^{*T} Dh(x^*)\, D\phi(0) \\
&= -Df(x^*)\, D\phi(0) = (f \circ \phi)'(0) \\
&= -\lim_{t \to 0^+} \frac{f(\phi(t)) - f(\phi(0))}{t} \geq 0.
\end{aligned}$$

Hence $\lambda^* \geq 0$.

Finally, we show the existence of the above map ϕ. Recall the condition $\operatorname{rank} DH(x^*) = m$. By reordering the variables x_i, $i = 1, \dots, n$, we make the first m columns of $DH(x^*)$ linearly independent. Write $\pi(x) = (x_{m+1}, \dots, x_n)$. We want to use the implicit function theorem applied to $F\colon U \times \mathbb{R} \to \mathbb{R}^n$ given by

$$F(x,t) = (H(x) - te_j, \pi(x - x^*)).$$

Notice that $F(x^*, 0) = 0$ and $\frac{\partial F}{\partial x}(x^*, 0)$ has n linearly independent rows. Then, there is a C^1 map ϕ on a neighbourhood of zero with values around x^*, such that $\phi(0) = x^*$ and $F(\phi(t), t) = 0$. So, $H \circ \phi(t) = te_j$. $\square$

Example 4.2. Consider the case $f(x,y) = x^2 - y$ and $h(x,y) = 1 - x^2 - y^2$. Hence, $D = \{(x,y) \in \mathbb{R}^2 \colon x^2 + y^2 \leq 1\}$ is a compact set (the disk of radius 1 centered at the origin). As f is continuous, the Weierstrass theorem guarantees the existence of minimizer and maximizer points of f on D. The Kuhn-Tucker conditions along with the condition that $(x,y) \in D$ are

$$\begin{cases}
\frac{\partial f}{\partial x} + \lambda \frac{\partial h}{\partial x} = 0 \\
\frac{\partial f}{\partial y} + \lambda \frac{\partial h}{\partial y} = 0 \\
\lambda h(x,y) = 0 \\
h(x,y) \geq 0.
\end{cases}$$

This implies that

$$\begin{cases}
2x - 2\lambda x = 0 \\
-1 - 2\lambda y = 0 \\
\lambda(1 - x^2 - y^2) = 0 \\
x^2 + y^2 \leq 1
\end{cases}$$

with solutions (x^*, y^*, λ^*) given by

$$(0, -1, 1/2), (0, 1, -1/2), (\sqrt{3}/2, -1/2, 1), (-\sqrt{3}/2, -1/2, 1).$$

Since $h(x^*, y^*) = 0$ for all cases and

$$\operatorname{rank} Dh(x^*, y^*) = 1,$$

the points $(0, -1)$, $(0, 1)$, $(\sqrt{3}/2, -1/2)$ and $(-\sqrt{3}/2, -1/2)$ are thus candidates to be local optimal points according to Theorem 4.1. In particular, $(0, 1)$ has to be the minimizer because it is the only one with $\lambda^* < 0$. Comparing all the other values we can check that $f(\sqrt{3}/2, -1/2) = f(-\sqrt{3}/2, -1/2) = 5/4$ is the maximum of f on D.

Exercise 4.3. Find the optimal points of f on D for:

(1) $f(x, y) = 2x^2 + 3y^2$,
$$D = \{(x, y) \in \mathbb{R}^2 \colon 11 - x - 2y \geq 0, x \geq 0, y \geq 0\}.$$
(2) $f(x, y) = (2x + y)^2$,
$$D = \{(x, y) \in \mathbb{R}^2 \colon 16 - x^2 - y \geq 0, x \geq 0, y \geq 0\}.$$

4.2. **Mixed constrains.** Using the previous results we can now mix equality with inequality constraints in order to have a more general optimization theorem.

Let $U \subset \mathbb{R}^n$ be open, $g \colon U \to \mathbb{R}^k$ and $h \colon U \to \mathbb{R}^\ell$ be C^1 functions, and

$$D = \{x \in U \colon g(x) = 0, h(x) \geq 0\}.$$

Theorem 4.4. *If x^* is a local optimal point of f on D,*

$$h_1(x^*) = \cdots = h_m(x^*) = 0, \quad h_{m+1}(x^*), \ldots, h_\ell(x^*) > 0$$

and $\operatorname{rank} D(g_1, \ldots, g_k, h_1, \ldots, h_m)(x^*) = k + m$, *then there is $\lambda^* \in \mathbb{R}^{k+\ell}$ such that for $j = k + 1, \ldots, k + \ell$,*

(1) $Df(x^*) + \lambda^{*T} D(g, h) = 0$,
(2) $\lambda^*_{k+i} h_i(x^*) = 0$, $i = 1, \ldots, \ell$.

Moreover,

- *if x^* is a local minimizer, then $(\lambda^*_{k+1}, \ldots, \lambda^*_{k+\ell}) \leq 0$.*
- *if x^* is a local maximizer, then $(\lambda^*_{k+1}, \ldots, \lambda^*_{k+\ell}) \geq 0$.*

This theorem is a direct consequence of the Lagrange and Kuhn-Tucker theorems, hence we omit its proof.

Example 4.5. Consider $f(x, y) = \log(xy)$ and

$$D = \{(x, y) \in \mathbb{R}^2 \colon x^2 + y^2 = 1, x > 0, y > 0\}.$$

So, we write $g(x, y) = x^2 + y^2 - 1$, $h_1(x, y) = x$ and $h_2(x, y) = y$. In addition, $n = 2$, $k = 1$, $\ell = 2$ and $m = 0$. The solutions of the system of equations and inequalities:

$$\begin{cases} \frac{\partial f}{\partial x}+\lambda_1\frac{\partial g}{\partial x}+\lambda_2\frac{\partial h_1}{\partial x}+\lambda_3\frac{\partial h_2}{\partial x}=0 \\ \frac{\partial f}{\partial y}+\lambda_1\frac{\partial g}{\partial y}+\lambda_2\frac{\partial h_1}{\partial y}+\lambda_3\frac{\partial h_2}{\partial y}=0=0 \\ g(x,y)=0 \\ \lambda_2 h_1(x,y)=0 \\ \lambda_3 h_2(x,y)=0 \\ h_1(x,y)>0 \\ h_2(x,y)>0. \end{cases}$$

The last four conditions above impliy easily that $\lambda_2=\lambda_3=0$. Hence,

$$\begin{cases} \frac{1}{x}+2\lambda_1 x=0 \\ \frac{1}{y}+2\lambda_1 y=0 \\ x^2+y^2-1=0 \\ x>0 \\ y>0 \end{cases}$$

and the solutions are $(x,y)=(\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2})$ and $\lambda_1=-1$. Since rank $Dg(x,y)=1$, this is the only candidate to a local optimization point. Now, $f(\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2})=\log\frac{1}{2}$ is larger to the image of any point in D close to $(0,1)$, it must correspond to a maximum.

5. Convex and concave optimizations

A set $D\subset\mathbb{R}^n$ is *convex* if

$$\lambda x+(1-\lambda)y\in D$$

for any $\lambda\in[0,1]$ and $x,y\in D$. That is, the line joining any two points of D is still contained in D.

Given a convex set $D\subset\mathbb{R}^n$, $f\colon D\to\mathbb{R}$ is a *convex* function on D if

$$f(\lambda x+(1-\lambda)y)\le\lambda f(x)+(1-\lambda)f(y)$$

for any $\lambda\in[0,1]$ and $x,y\in D$. If the above inequality is strict, we say that f is *strictly convex* on D.

On the other hand, f is a *concave* function on D if $-f$ is convex In addition, f is *strictly concave* on D if $-f$ is strictly convex.

Notice that functions could be both convex and concave or neither convex nor concave. Moreover, a function both convex and concave can not be neither strictly convex nor strictly concave.

Example 5.1. Let $f(x)=x^3$ on $\mathbb{R}$ and $x=1$, $y=-1$. For $\lambda=1/4$ we have

$$f(\lambda x+(1-\lambda)y)<\lambda f(x)+(1-\lambda)f(y)$$

and for $\lambda = 3/4$,

$$f(\lambda x + (1-\lambda)y) > \lambda f(x) + (1-\lambda)f(y).$$

Thus, f is neither convex nor concave.

Exercise 5.2. Show that $f\colon \mathbb{R}^n \to \mathbb{R}$ is both convex and concave iff f is affine (i.e. $f(x) = a^T x + b$ for some $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$).

Theorem 5.3. *If $f\colon D \to \mathbb{R}$ is differentiable on $D \subset \mathbb{R}^n$ open and convex, then*

(1) *f is concave on D iff*

$$Df(x)\,(y-x) \geq f(y) - f(x), \quad x, y \in D.$$

(2) *f is convex on D iff*

$$Df(x)\,(y-x) \leq f(y) - f(x), \quad x, y \in D.$$

Proof. □

Theorem 5.4. *Let $f\colon D \to \mathbb{R}$ be C^2 on $D \subset \mathbb{R}^n$ open and convex.*

(1) *f is concave on D iff $D^2 f(x) \leq 0$ for any $x \in D$.*
(2) *f is convex on D iff $D^2 f(x) \geq 0$ for any $x \in D$.*
(3) *If $D^2 f(x) < 0$ for any $x \in D$, then f is strictly concave.*
(4) *If $D^2 f(x) > 0$ for any $x \in D$, then f is strictly convex.*

Proof. □

Example 5.5. Consider the function $f(x) = \log(x_1^\alpha \dots x_n^\alpha)$ for $x_i > 0$ and $\alpha > 0$. Hence,

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \begin{cases} -\frac{\alpha}{x_i^2}, & i = j \\ 0, & i \neq 0. \end{cases}$$

So, $D^2 f(x) < 0$ and f is strictly concave.

5.1. **Optimization.**

Theorem 5.6. *Let $D \subset \mathbb{R}^n$ be convex and $f\colon D \to \mathbb{R}$ convex (concave). Then,*

(1) *any local minimizer (maximizer) of f on D is in fact global.*
(2) *the set of minimizers (maximizers) of f on D is either empty or convex.*
(3) *if f is strictly convex (concave), then the set of minimizers (maximizers) of f on D is either empty or it contains a single point.*

Proof. □

5.1.1. *Unconstrained optimization.*

Theorem 5.7. *Let $D \subset \mathbb{R}^n$ open and convex and $f\colon D \to \mathbb{R}$ convex (concave) and differentiable. Then, x^* is a minimizer (maximizer) of f on D iff $Df(x^*) = 0$.*

Proof. We already know that a maximizer is a critical point. It remains to show the reverse.

Consider f convex (use the same idea for f concave). For any $y \in D$ we have

$$f(y) - f(x^*) \geq Df(x^*)\,(y - x^*) = 0.$$

So, $f(y) \geq f(x^*)$ and x^* is a minimizer. □

5.1.2. *Kuhn-Tucker conditions under convexity.* Given an open and convex set $U \subset \mathbb{R}^n$ and a C^1 function $h\colon U \to \mathbb{R}^\ell$, take as before the restriction set

$$D = \{x \in U\colon h(x) \geq 0\}.$$

We say that h is convex (concave) if all its components h_i, $i = 1, \dots, \ell$ are convex (concave)

Theorem 5.8 (Convex Kuhn-Tucker)**.** *If f and h_i are convex C^1 functions and there is $\bar{x} \in U$ satisfying $h(\bar{x}) > 0$, then x^* is a minimizer of f on D iff there is $\lambda^* \in \mathbb{R}^\ell$ such that*

(1) $Df(x^*) + \lambda^{*T} Dh(x^*) = 0$,
(2) $\lambda_i^* h_i(x^*) = 0$, $i = 1, \dots, \ell$.
(3) $\lambda^* \leq 0$.

Remark 5.9. Notice that the above conditions will give either no solutions or a unique solution.

Proof. □

The corresponding version for concave functions follows.

Theorem 5.10 (Concave Kuhn-Tucker)**.** *If f and h_i are concave C^1 functions and there is $\bar{x} \in U$ satisfying $h(\bar{x}) > 0$, then x^* is a maximizer of f on D iff there is $\lambda^* \in \mathbb{R}^\ell$ such that*

(1) $Df(x^*) + \lambda^{*T} Dh(x^*) = 0$,
(2) $\lambda_i^* h_i(x^*) = 0$, $i = 1, \dots, \ell$.
(3) $\lambda^* \geq 0$.

Exercise 5.11. Find the optimal points of f on D:

(1) $f(x, y) = x + y$,
$$D = \{(x, y) \in \mathbb{R}^2\colon 8 - 2x - y \geq 0, x \geq 0, y \geq 0\}.$$

(2) $f(x, y) = (10 - x - y)(x + y) - ax - y - y^2,$
$$D = \{(x, y) \in \mathbb{R}^2 \colon x \geq 0, y \geq 0\},$$
where $a > 0$.

References

[1] R. K. Sundaram. *A First Course in Optimization Theory.* Cambridge University Press, 1996.

Index

Departamento de Matemática, ISEG, Universidade de Lisboa, Rua do Quelhas 6, 1200-781 Lisboa, Portugal

E-mail address: jldias@iseg.utl.pt