QEM "Optimization", WS 2016/17

Part 3. Convexity

(about 3–4 Lectures)

Supporting Literature:

Angel de la Fuente, **"Mathematical Methods and Models for Economists"**, Chapter 6;

Sundaram R.K., **"A First Course in Optimization Theory"**, Chapters 7 and 8.

Contents

For more details see also the (very extended) free E-book *"Convex Optimization"* (about 700 pages) by S. Boyd and L. Vandenberghe from Stanford University:
www.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf

3 Convexity

3.1 Definition of Convex Sets and Convex (resp. Concave) Functions

Definition 3.1.1. *A set $U \subset \mathbb{R}^n$ is* ***convex*** *if for any two points $x, y \in U$, their* ***convex combination*** *also belongs to U, i.e.,*

$$\lambda x + (1-\lambda) y \in U, \ \textit{for all} \ 0 \le \lambda \le 1,$$

or equivalently, the segment $[x,y] \subset U$*.*

Notation 3.1.2. *The segment* $[x,y]$ *connecting* x *and* y *is given by*

$$[x,y] := \{z \in \mathbb{R}^n \mid z = \lambda x + (1-\lambda)y,\ \lambda \in [0,1]\}.$$

Definition 3.1.3. *Let* $U \subset \mathbb{R}^n$ *be a convex set.*

(i) A function $f\colon U \to \mathbb{R}$ *is called* ***convex*** *if for all* $x, y \in U$ *and* $\lambda \in [0,1]$

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

(ii) A function $f\colon U \to \mathbb{R}$ *is called* ***strictly convex*** *if for all pairs of* ***distinct*** *points* $x, y \in U$*,* $x \neq y$*, and all* $\lambda \in (0,1)$

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

(iii) A function $f\colon U \to \mathbb{R}$ *is called* ***(strictly) concave*** *if* $(-f)$ *is (strictly) convex.*

These conditions are not easy to check in multidimensional case ($n \ge 1$). To this end, it is more convenient to use Differential Calculus.

Theorem 3.1.4. *(will be proved later)*

Let $f\colon U \to \mathbb{R}$ *be a twice continuously differentiable function defined on an open convex set* $U \subset \mathbb{R}^n$*, i.e.,* $f \in C^2(U)$*. Then:*

(i) f *is convex if and only if the Hessian matrix* $D^2 f(x)$ *is* ***positive semidefinite*** *for all* $x \in U$*, i.e.,*

$$\langle D^2 f(x)h, h\rangle \ge 0 \text{ for any } h \in \mathbb{R}^n.$$

(ii) If the Hessian is ***positive definite****, i.e., for all* $x \in U$

$$\langle D^2 f(x)h, h\rangle > 0 \text{ for any } h \in \mathbb{R}^n \setminus \{0\},$$

then f *is strictly convex.*

(iii) f *is concave if and only if the Hessian matrix* $D^2 f(x)$ *is* ***negative semidefinite*** *for all* $x \in U$*, i.e.,*

$$\langle D^2 f(x)h, h\rangle \le 0 \text{ for any } h \in \mathbb{R}^n.$$

(iv) If the Hessian is ***negative definite****, i.e., for all* $x \in U$

$$\langle D^2 f(x)h, h\rangle < 0 \text{ for any } h \in \mathbb{R}^n \setminus \{0\},$$

then f *is strictly concave.*

Warning: The positive (resp. negative) definiteness of $D^2 f(x)$ is *sufficient but not necessary* for the strict convexity (resp. concavity) of f.

Counterexample: $f(x) := x^4$, $x \in \mathbb{R}$. Obviously, f is strictly convex on the whole line $\mathbb{R}$, but $f''(0) = 0$.

3.2 Properties of Convex Sets

Lemma 3.2.1. *Any intersection of convex sets is convex.*

In other words, let each U_i, $i \in I$, be convex and let I be an arbitrary (not necessary countable) index set. Then

$$\bigcap_{i \in I} U_i$$

is convex.

Proof. Trivial (Homework). $\square$

Lemma 3.2.2. *Let U and V be convex sets in $\mathbb{R}^n$. Then the following sets are also convex*

$$\begin{aligned} U + V &:= \{z \in \mathbb{R}^n \mid z = x + y \text{ for some } x \in U,\ y \in V\}, \\ \alpha U &:= \{z \in \mathbb{R}^n \mid z = \alpha x \text{ for some } x \in U\},\ \alpha \in \mathbb{R}. \end{aligned}$$

Proof. Trivial (Homework). $\square$

Corollary 3.2.3. *Any linear combination of convex sets*

$$\alpha U + \beta V = \{z \in \mathbb{R}^n \mid z = \alpha x + \beta y \text{ for some } x \in U,\ y \in V\},\ \alpha,\ \beta \in \mathbb{R},$$

is also convex.

Definition 3.2.4. *A point $y \in \mathbb{R}^n$ is a* ***convex combination*** *of given points $x_1, \dots, x_m \in \mathbb{R}^n$ ($m \in \mathbb{N}$) if it can be written as*

$$y = \sum_{i=1}^{m} \lambda_i x_i \text{ with some } \lambda_i \in [0,1],\ 1 \le i \le m,\ \text{ s.t. } \sum_{i=1}^{m} \lambda_i = 1.$$

Lemma 3.2.5. *A set $U \subset \mathbb{R}^n$ is convex if and only if every convex combination of points of U lies in U.*

Proof. ($\Longleftarrow$) Obvious.
($\Longrightarrow$) By the method of mathematical induction. The statement is true for $m = 2$.

Suppose it is true for any $(m-1)$ points from U. It suffices to consider

$$\lambda_i \in (0,1),\ \sum_{i=1}^{m} \lambda_i = 1.$$

Define

$$\begin{aligned} y &:= \sum_{i=1}^{m} \lambda_i x_i = \lambda_m x_m + \sum_{i=1}^{m-1} \lambda_i x_i \\ &= \lambda_m x_m + (1 - \lambda_m) \sum_{i=1}^{m-1} \frac{\lambda_i}{1 - \lambda_m} x_i = \lambda_m x_m + (1 - \lambda_m) z, \end{aligned}$$

where we let z denote

$$z := \sum_{i=1}^{m-1} \frac{\lambda_i}{1-\lambda_m} x_i = \sum_{i=1}^{m-1} \mu_i x_i,$$

$$\text{with } \mu_i := \frac{\lambda_i}{(1-\lambda_m)} \in (0,1), \ \sum_{i=1}^{m-1} \mu_i = \sum_{i=1}^{m-1} \frac{\lambda_i}{(1-\lambda_m)} = 1.$$

By induction $z \in U$ and by convexity $y := \lambda_m x_m + (1-\lambda_m) z \in U$. $\square$

Definition 3.2.6. *Let $A \subset \mathbb{R}^n$ be an arbitrary set. The* ***convex hull*** *of A, which will be denoted by* conv$A$*, is the smallest convex set containing A. That is,*

$$\mathrm{conv} A := \bigcap_{A \subseteq U,\ U \, convex} U.$$

which is convex by lemma 3.2.1

Theorem 3.2.7. *The convex hull of A coincides with the set of all convex combinations of elements from A.*

Proof. Very similar to the proof of Lemma 3.2.5. $\square$

Question: How many points are required in Theorem 3.2.7? The answer is given by

Theorem 3.2.8 (Carathéodory)**.** *If y is a convex combination of some points $x_1, \dots, x_m \in \mathbb{R}^n$ (with arbitrary $m \in \mathbb{N}$, possibly $m > n$), then y can be represented as a convex combination of $(n+1)$ or fewer points from this list.*

Equivalently, y lies in an r-simplex with vertices in $\mathbb{R}^n$, where $r \le n+1$.

Lemma 3.2.9 (metric properties of convex sets)**.** *Let $U \subset \mathbb{R}^n$ be convex. Then the interior* $\mathrm{int}\, U$ *and the closure $\bar{U}$ are convex too.*

Proof. The proof is left as an easy exercise. $\square$

Remark 3.2.10. *The definitions and properties of convex sets extend from $\mathbb{R}^n$ to any Banach or normed space X.*

3.3 Properties of Concave Functions

Lemma 3.3.1. *Let $U \subset \mathbb{R}^n$ be a convex set, $f \colon U \to \mathbb{R}$ be a concave (resp. convex) function and $g \colon \mathbb{R} \to \mathbb{R}$ be an increasing concave (resp. convex) function. Then the composition $h = g \circ f$ is concave (resp. convex).*

Lemma 3.3.2. *Let $U \subset \mathbb{R}^n$ be a convex set and $f, g \colon U \to \mathbb{R}$ be concave (resp. convex) functions. Then, for all $\alpha, \beta \ge 0$, the linear combination $h := \alpha f + \beta g$ is concave (resp. convex).*

Exercise 3.3.3. *Prove the above lemmas by directly using the definition of convex/concave functions. Complete solutions can be found in the book of de la Fuente.*

Theorem 3.3.4. *Let $f: U \to \mathbb{R}$ be a concave (resp.* ***convex****) function defined on an open convex set $U \subset \mathbb{R}^n$. Then f is continuous on U.*

Proof. We'll skip this proof. For details see de la Fuente, p. 252. □

Indeed, it can be shown that any such f is continuously differentiable almost everywhere on U, except possibly at a set of points having the 'Lebesgue measure' (a generalisation of area) zero.

Lemma 3.3.5. *Let $\{f_i\}_{i \in I}$ be a (possibly infinite) family of concave functions on U, all of which are bounded below,*

$$\exists C \in \mathbb{R}: \quad f_i(x) \geq C \text{ for all } x \in U,\ i \in I.$$

(i) Then the function $f: U \to \mathbb{R}$ given by

$$f(x) = \inf\{f_i(x) \mid i \in I\}$$

is concave.

(ii) (Exercise) Reformulate this statement for convex functions.

The idea for the proof is based on the notion of *hypographs* which we do not introduce here.

3.4 Concavity of Smooth Functions

Theorem 3.4.1 (Characterization of concavity by means of the tangent hyperplane)**.** *Let $U \subset \mathbb{R}^n$ be an open convex set and $f: U \to \mathbb{R}$ be continuously differentiable ($f \in C^1(U)$). Then:*

(i) f is concave if and only if

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle, \text{ for all } x, y \in U; \qquad (*)$$

(ii) f is strictly concave if and only if

$$f(y) < f(x) + \langle \nabla f(x), y - x \rangle, \text{ for all } x, y \in U,\ x \neq y. \qquad (**)$$

Proof. (only necessary conditions):

(i) f concave $\Longrightarrow f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle$, $\forall x, y \in U$?

Consider for $\lambda \in (0,1)$

$$z^{\lambda} := (1-\lambda)x + \lambda y = x + \lambda(y - x) \in U.$$

By concavity of f

$$\begin{array}{rcl} (1-\lambda)f(x) + \lambda f(y) & \leq & f\left[(1-\lambda)x + \lambda y\right], \text{ or} \\ f(y) - f(x) & \leq & \frac{f[x+\lambda(y-x)]-f(x)}{\lambda}. \end{array}$$

Taking $\lambda \to 0$, in the RHS we get the directional derivative along the vector $(y - x)$, i.e.,

$$f(y) - f(x) \leq \langle \nabla f(x), y - x \rangle.$$

Comment: the graph of f lies below the tangent hyperplane drawn at every point $x \in U$.

(ii) f strictly concave $\Longrightarrow f(y) < f(x) + \langle \nabla f(x), y - x \rangle$, $x \neq y$?

Indeed, analogously to the proof of (i), we have by strict concavity of f that

$$f(y) - f(x) < \frac{f(z^{\lambda}) - f(x)}{\lambda}, \quad \lambda \in (0,1).$$

Now we use the following trick: For $z^{\lambda} := (1-\lambda)x + \lambda y$, we have by the claim **(i)**

$$\begin{array}{rcl} f(z^{\lambda}) - f(x) & \leq & \langle \nabla f(x), z^{\lambda} - x \rangle, \text{ and thus} \\ f(y) - f(x) & < & \frac{1}{\lambda}\langle \nabla f(x), z^{\lambda} - x \rangle, \ \lambda \in (0,1). \end{array}$$

Calculate

$$\frac{1}{\lambda}(z^{\lambda} - x) = \frac{1}{\lambda}\left[(1-\lambda)x + \lambda y - x\right] = y - x.$$

Hence,

$$f(y) - f(x) < \langle \nabla f(x), y - x \rangle_{\mathbb{R}^n}.$$

□

Theorem 3.4.2. *Let $f\colon U \to \mathbb{R}$ be a twice continuously differentiable function (i.e., $f \in C^2(U)$) defined on an open convex set $U \subset \mathbb{R}^n$. Then f is concave on U if and only if the Hessian matrix $D^2 f(x)$ is negative semidefinite for all $x \in U$, i.e.,*

$$\langle D^2 f(x)h, h \rangle \leq 0 \quad \text{for all } h \in \mathbb{R}^n.$$

Moreover, if the Hessian is negative definite, i.e.,

$$\langle D^2 f(x)h, h \rangle < 0 \quad \text{for all } h \in \mathbb{R}^n \setminus \{0\},$$

then f is strictly concave.

Remark 3.4.3. *(i) Negative definiteness of $D^2 f(x)$ is sufficient but* ***not necessary*** *for the strict concavity of f.*

(ii) The positive / negative definiteness or semidefiniteness of the matrix $D^2 f(x)$ can be checked using the principal-minor test.

Proof. **(a)** (f concave $\Longrightarrow D^2 f(x) \leq 0$)

Pick some direction $h \in \mathbb{R}^n$ and define a function

$$\alpha \mapsto g(\alpha) := f(x + \alpha h) - f(x) - \alpha \langle \nabla f(x), h \rangle,$$

for small enough $|\alpha| < \delta$ (such that $B_\delta(x) \subset U$). By concavity of f and Theorem 3.4.1,

$$f(x + \alpha h) - f(x) \leq \langle \nabla f(x), \alpha h \rangle,$$

which implies

$$g(\alpha) \leq 0 \text{ for all } |\alpha| < \delta.$$

On the other hand, $g(0) = 0$, $g \in C^1((-\delta, \delta))$, and g has a local maximum at $\alpha = 0$. By the necessary condition (well-known from Calculus) for local max of scalar functions we have

$$g'(0) = 0, \; g''(0) \leq 0$$

(it is obvious that $g''(0) \leq 0$ since supposing the opposite $g''(0) > 0$ would give $\alpha = 0$ is a strict local minimum for g). But

$$g''(0) = \langle D^2 f(x) h, h \rangle,$$

which shows that $\langle D^2 f(x) h, h \rangle \leq 0$ for all $h \in \mathbb{R}^n$.

(b) ($D^2 f(x) \leq 0 \Longrightarrow f$ concave)

By Theorem 3.4.1, it is enough to show that for all $h \in \mathbb{R}^n$ such that $x, x + h \in U$

$$f(x + h) \leq f(x) + \langle \nabla f(x), h \rangle.$$

Pick any $x, x + h \in U$. Since $f \in C^2(U)$, by Taylor's formula $\exists \theta \in (0, 1)$ such that

$$\begin{aligned} f(x+h) &= f(x) + \langle \nabla f(x), h \rangle + \tfrac{1}{2} \underbrace{\langle D^2 f(x + \theta h) h, h \rangle}_{\leq 0} \\ &\Longrightarrow f(x+h) \leq f(x) + \langle \nabla f(x), h \rangle. \end{aligned}$$

By Theorem 3.4.1 this implies that f is concave. $\square$

3.5 Examples

Example 3.5.1 (Quadratic function with parameters)**.** *Let $a, b > 0$:*

$$f(x,y) := ax^2 + by^2, \quad x, y \in \mathbb{R}.$$

The Hessian

$$D^2 f(x,y) = \mathrm{Hess} f(x,y) = \begin{pmatrix} 2a & 0 \\ 0 & 2b \end{pmatrix}$$

is positive definite (since $2a > 0$, $2a \cdot 2b > 0$). So $f(x,y)$ is strictly convex on $\mathbb{R}^2$.

Example 3.5.2 (Euclidean norm function in $\mathbb{R}^n$)**.** *Let $f \colon \mathbb{R}^n \to \mathbb{R}$ be given by*

$$f(x) := \|x\| = \sqrt{\sum_{i=1}^{n} x_i^2}, \quad x = (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

We claim that f is convex. Indeed by the properties of $\|\cdot\|$, we have for any $x, y \in \mathbb{R}^n$ and $\lambda \in [0,1]$

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= \|\lambda x + (1-\lambda)y\| \\ &\leq \lambda\|x\| + (1-\lambda)\|y\| = \lambda f(x) + (1-\lambda) f(y). \end{aligned}$$

Alternatively, by Theorem 3.4.2:

$$\nabla f(x) = \frac{x}{\|x\|}, \quad \text{for all } x \neq 0.$$

Then by the Cauchy-Schwartz inequality,

$$\langle \nabla f(x), y - x \rangle = \frac{\langle x, y - x \rangle}{\|x\|} = \frac{\langle x, y \rangle}{\|x\|} - \|x\| \leq \|y\| - \|x\| = f(y) - f(x)$$
$$\text{or} \quad f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle.$$

Example 3.5.3 (CES production functions)**.** *Let $a, b > 0$:*

$$f(x,y) := (ax^p + by^p)^{1/p},$$
$$U := \{(x,y) \mid x > 0,\ y > 0\} \subset \mathbb{R}^2.$$

Exercise: *For which $p > 0$ is this function convex / concave?*

Example 3.5.4 (Additive utility)**.** *Set the coefficients $c_i > 0$.*

$$F(x) := \sum_{i=1}^{n} c_i f_i(x_i) \quad \text{on } \mathbb{R}^n.$$

F is concave (convex) on $\mathbb{R}^n \iff$ all f_i are concave (convex) on $\mathbb{R}$.

Example 3.5.5 (Leontieff function)**.** *The following function is concave.*

$$f(x,y) = \min\{x,y\}, \quad (x,y) \in \mathbb{R}^2.$$

Example 3.5.6. *The simplest utility or production function*

$$f(x,y) := xy, \ \ U = \{(x,y) \mid x, y > 0\},$$

$$D^2 f(x,y) = \mathrm{Hess} f(x,y) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \textit{for all } x, y.$$

Since $\det D^2 f(x,y) = -1$*, the matrix* $D^2 f(x,y)$ *is indefinite and hence is neither concave nor convex.*

Example 3.5.7 (Cobb-Douglas function)**.** *Set the parameters* $a, b > 0$

$$f(x,y) = x^a y^b, \quad U = \{(x,y) \mid x, y > 0\}.$$

$$D^2 f(x,y) = \begin{pmatrix} a(a-1)x^{a-2}y^b & abx^{a-1}y^{b-1} \\ abx^{a-1}y^{b-1} & b(b-1)x^a y^{b-2} \end{pmatrix}.$$

$$\det D^2 f(x,y) = ab(1-a-b)x^{2a-2}y^{2b-2}.$$

Then, f *is concave on* U *if and only if* $D^2 f(x,y) \leq 0$*, i.e.,*

$$\begin{cases} a(a-1) \leq 0, \\ b(b-1) \leq 0, \\ ab(1-a-b) \geq 0, \end{cases} \iff \begin{cases} 0 \leq a \leq 1, \\ 0 \leq b \leq 1, \\ a+b \leq 1. \end{cases}$$

For strict concavity it suffices to have $D^2 f(x,y) < 0$*, i.e.,*

$$\begin{cases} a(a-1) < 0, \\ ab(1-a-b) > 0, \end{cases} \implies \begin{cases} 0 < a < 1, \\ 0 < b < 1, \\ a+b < 1. \end{cases}$$

And f *is neither concave nor convex if* $a + b > 1$*.*

3.6 Extrema of Concave Functions

For concave / convex functions, the 1st order conditions are both necessary and sufficient to identify their global maxima (or minima). We need not to analyse the 2nd derivatives to separate maxima, minima and saddle points.

A critical point of a concave (resp. convex) function is automatically its global maximum (resp. minimum).

Theorem 3.6.1. *Let $U \subset \mathbb{R}^n$ be an open convex set and $f: U \to \mathbb{R}$ be a continuously differentiable (i.e., $f \in C^1(U)$), concave function. Then x^* is a critical point of f, i.e.,*

$$\nabla f(x^*) = 0,$$

if and only if x^* is a global maximum of f on U.

Proof. Necessity of the condition $\nabla f(x^*) = 0$ for any local (and hence for any global) maximum x^* is given as a theorem in Part II.

So, we need to prove sufficiency only. Suppose that $\nabla f(x^*) = 0$.

Since $f \in C^1(U) \Longrightarrow$ by Theorem 3.4.1

$$f \text{ concave } \iff f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle, \quad \forall x, y \in U.$$

Since $\nabla f(x^*) = 0$, we get $f(y) \leq f(x^*)$ for all $y \in U$, which means that x^* is a global maximum. □

In the next theorem we do **not** need to assume that $f \in C^1(U)$.

Theorem 3.6.2. *Let $U \subset \mathbb{R}^n$ be an open convex set and $f: U \to \mathbb{R}$ be concave. Then:*

(i) Any local maximum of f is its global maximum on U.

(ii) If $f: U \to \mathbb{R}$ is strictly concave, then it has at most one maximum.

Proof. **(i)** Let x^* be a local maximum, but not a global one. Hence,

$$\begin{aligned} \exists \epsilon &> 0 \text{ s.t. } f(x^*) \geq f(y), \forall y \in B_\epsilon(x^*) \subseteq U, \quad \text{but} \\ \exists z &\in U: \; f(z) > f(x^*). \end{aligned}$$

Since U is convex, we have $y_\lambda := \lambda x^* + (1-\lambda)z \in U$ for all $\lambda \in [0,1]$. Choose λ sufficiently close to 1 such that $y_\lambda \in B_\epsilon(x^*)$.

By concavity of f, for such $\lambda \in (0,1)$

$$f(y_\lambda) \geq \lambda f(x^*) + (1-\lambda) f(z) > f(x^*),$$

since $f(z) > f(x^*)$. On the other hand, since x^* is a local maximum,

$$f(\lambda x^* + (1-\lambda)z) = f(y_\lambda) \leq f(x^*),$$

which leads to a contradiction.

(ii) Suppose x^* and x_* are two different maxima, i.e., $x^* \neq x_*$ and

$$f(x^*) = f(x_*) \geq f(y), \quad \forall y \in U.$$

For $z := (x^* + x_*)/2$ we get by strict concavity

$$f(z) > \frac{1}{2}\left[f(x^*) + f(x_*)\right] = f(x_*),$$

which contradicts the assumption that x^* and x_* are maxima. □

3.7 Summary: Unconstrained Optimization for Differentiable functions

Here, we summarise some of the most important results from this section and the last on finding and classifying interior extrema for unconstrained optimisation problems involving differentiable functions.

Theorem 3.7.1 (2.10.1 in Part II - 1st Order Necessary Condition for Local Extrema)**.** *Let $U \subset \mathbb{R}^n$ be an open set. A differential function $f : U \to \mathbb{R}$ can only have a local maximum or minimum at a point $x^* \in U$ if this point is critical, i.e.,*

$$\mathrm{grad} f(x^*) := \nabla f(x^*) = \left(\frac{\partial f}{\partial x_1}(x^*), \ldots, \frac{\partial f}{\partial x_n}(x^*) \right) = 0.$$

Theorem 3.7.2 (2.11.2 in Part II - 2nd Order Sufficient Conditions for Local Extrema)**.** *Let $U \subset \mathbb{R}^n$ be open, $f : U \to \mathbb{R}$ be twice continuously differentiable on U (i.e., $f \in C^2(U)$), and let $x^* \in U$ be a critical point of f. Then:*

(i) The Hessian $\mathrm{Hess} f(x^*)$ (also denoted by $D^2 f(x^*)$) is positive definite (i.e., $\mathrm{Hess} f(x^*) > 0$) implies that x^* is a strict local min.

The positive definiteness of $\mathrm{Hess} f(x^*)$ is equivalent to the positivity of all n leading principal minors of $D^2 f(x^*)$:

$$\partial^2_{1,1} f(x^*) > 0, \quad \begin{vmatrix} \partial^2_{1,1} f(x^*) & \partial^2_{1,2} f(x^*) \\ \partial^2_{2,1} f(x^*) & \partial^2_{2,2} f(x^*) \end{vmatrix} > 0,$$

$$\begin{vmatrix} \partial^2_{1,1} f(x^*) & \partial^2_{1,2} f(x^*) & \partial^2_{1,3} f(x^*) \\ \partial^2_{2,1} f(x^*) & \partial^2_{2,2} f(x^*) & \partial^2_{2,3} f(x^*) \\ \partial^2_{3,1} f(x^*) & \partial^2_{3,2} f(x^*) & \partial^2_{3,3} f(x^*) \end{vmatrix} > 0, \ldots, \; |D^2 f(x^*)| > 0.$$

(ii) $\mathrm{Hess} f(x^*)$ is negative definite (i.e., $\mathrm{Hess} f(x^*) < 0$) implies that x^* is a strict local max.

The matrix $\mathrm{Hess} f(x^*)$ is negative definite if and only if its n leading principal minors alternate in sign:

$$\partial^2_{1,1} f(x^*) < 0, \quad \begin{vmatrix} \partial^2_{1,1} f(x^*) & \partial^2_{1,2} f(x^*) \\ \partial^2_{2,1} f(x^*) & \partial^2_{2,2} f(x^*) \end{vmatrix} > 0,$$

$$\begin{vmatrix} \partial^2_{1,1} f(x^*) & \partial^2_{1,2} f(x^*) & \partial^2_{1,3} f(x^*) \\ \partial^2_{2,1} f(x^*) & \partial^2_{2,2} f(x^*) & \partial^2_{2,3} f(x^*) \\ \partial^2_{3,1} f(x^*) & \partial^2_{3,2} f(x^*) & \partial^2_{3,3} f(x^*) \end{vmatrix} < 0, \ldots, \; (-1)^n |D^2 f(x^*)| > 0.$$

(iii) $\mathrm{Hess} f(x^*)$ is indefinite (i.e, is neither positive semidefinite nor negative semidefinite) $\Longrightarrow$ x^* is not a local extremum (i.e., x^* is a saddle point).

(iv) If x^* is a local min then $\mathrm{Hess} f(x^*)$ *is positive semi-definite. The matrix* $\mathrm{Hess} f(x^*)$ *is positive semi-definite if and only if each of its $(2^n - 1)$ principal minors is non-negative. The principal minors of order $k = 1, \ldots, n$ are the determinants of $k \times k$ submatrices obtained by deleting any $n - k$ columns, say $i_1, \ldots, i_{n-k}$, and the same $n - k$ rows, $i_1, \ldots, i_{n-k}$, from $D^2 f(x^*)$.*

(v) If x^* is a local max then $\mathrm{Hess} f(x^*)$ *is negative semidefinite. This is the case if and only if the principal minors alternate in sign, so that the odd ordered ones are ≤ 0 and the even ordered ones are ≥ 0.*

Remark 3.7.3. *Be careful:*

- *The claims (i) and (ii) of Theorem 2.11.2 are not necessary conditions.*
- $\mathrm{Hess} f(x^*) \geq 0$ *(or* $\mathrm{Hess} f(x^*) \leq 0$*) does not imply in general that x^* is a local (non-strict) minimum (or maximum). Counterexample: $f(x) = \pm x^4$ at $x = 0$.*

Example 3.7.4. *For a function f of two variables, the Hessian is*

$$D^2 f(x^*) := \begin{pmatrix} \partial^2_{1,1} f(x^*) & \partial^2_{1,2} f(x^*) \\ \partial^2_{2,1} f(x^*) & \partial^2_{2,2} f(x^*) \end{pmatrix}, \quad \text{with } \partial^2_{1,2} f(x^*) = \partial^2_{2,1} f(x^*).$$

This matrix is negative definite if $\partial^2_{1,1} f(x^*) < 0$ and $\det D^2 f(x^*) = \partial^2_{1,1} f(x^*) \cdot \partial^2_{2,2} f(x^*) - (\partial^2_{1,2} f(x^*))^2 > 0$. (These two inequalities imply that $\partial^2_{2,2} f(x^*) < 0$). This is a sufficient condition for a critical point x^* of a function of two variables to be a local max.

Similarly, a sufficient condition for x^* to be a local min is $D^2 f(x^*) > 0$, that is $\partial^2_{1,1} f(x^*) > 0$ and $\det D^2 f(x^*) > 0$ (which imply $\partial^2_{2,2} f(x^*) > 0$). In particular, if $\det D^2 f(x^*) < 0$, then x^* is neither a local maximizer nor a local minimizer, i.e., x^* is a saddle point. Note that this condition is only sufficient, not necessary.

The matrix $D^2 f(x^*)$ is negative semidefinite if $\partial^2_{1,1} f(x^*) \leq 0$, $\partial^2_{2,2} f(x^*) \leq 0$ and $\det D^2 f(x^*) \geq 0$. Finally, $D^2 f(x^*)$ is positive semidefinite if $\partial^2_{1,1} f(x^*) \geq 0$, $\partial^2_{2,2} f(x^*) \geq 0$ and $\det D^2 f(x^*) \geq 0$.

Theorem 3.7.5 (Sufficient Conditions for Convexity/Concavity)**.** *Let $f \colon U \to \mathbb{R}$ be a twice continuously differentiable function (i.e., $f \in C^2(U)$) defined on an open convex set $U \subset \mathbb{R}^n$. Then:*

(i) f is convex if and only if the Hessian matrix $D^2 f(x)$ is positive semidefinite for all $x \in U$, i.e.,

$$\langle D^2 f(x) h, h \rangle \geq 0 \ \text{ for all } h \in \mathbb{R}^n.$$

Every critical point $x^* \in U$ will surely be a global minimum point for f in U.

(ii) f *is concave if and only if the Hessian matrix* $D^2 f(x)$ *is negative semidefinite for all* $x \in U$*, i.e.,*

$$\langle D^2 f(x)h, h\rangle \leq 0 \text{ for all } h \in \mathbb{R}^n.$$

Every critical point $x^* \in U$ *will surely be a global maximum point for* f *in* U*.*

(iii) *If the Hessian is positive definite, i.e.,*

$$\langle D^2 f(x)h, h\rangle > 0 \text{ for all } h \in \mathbb{R}^n \setminus \{0\},$$

then f *is strictly convex.*

(iv) *If the Hessian is negative definite, i.e.,*

$$\langle D^2 f(x)h, h\rangle < 0 \text{ for all } h \in \mathbb{R}^n \setminus \{0\},$$

then f *is strictly concave.*

Remark 3.7.6. *The positive (resp. negative) definiteness of* $D^2 f(x)$ *is sufficient but not necessary for the strict convexity (resp. concavity) of* f*. Counterexample:* $f(x) = x^4$ *is strictly convex, but* $f''(0) = 0$*.*

So, we can state briefly the results of the above theorems for maximizers together:

(i) **Sufficient conditions for local max:** if x^* is a stationary point of f and $\text{Hess} f(x^*)$ is negative definite at this x^*, then x^* is a local maximizer of f.

(ii) **Sufficient conditions for global max:** if x^* is a stationary point of f and $\text{Hess} f(x)$ is negative semidefinite for all values of $x \in U$, then x^* is a global maximizer of f.

3.8 Separation Theorems

This section and the next on Farkas' Lemma are not examinable - see *A. de la Fuente, Section 6.1 d* for further details.

Definition 3.8.1. *A* ***hyperplane*** *in* $\mathbb{R}^n$ *is given by*

$$H_{c,v} := \{x \in \mathbb{R}^n \mid \langle x, v\rangle = c\},$$

for a vector $v \in \mathbb{R}^n \setminus \{0\}$ *and some* $c \in \mathbb{R}$*.*

The vectors $x \in H_{c,v}$ *satisfy a linear equation with* n *unknowns* $x_1, \ldots, x_n \in \mathbb{R}$*:*

$$\sum_{j=1}^{n} x_j v_j = c.$$

The hyperplane $H_{c,v}$ *divides* $\mathbb{R}^n$ *into 2 regions:*

(i) $x \in \mathbb{R}^n$ *satisfying* $\langle x, v\rangle \geq c$;

(ii) $x \in \mathbb{R}^n$ *satisfying* $\langle x, v\rangle \leq c$.

We say that two sets $A, B \subset \mathbb{R}^n$ *are separated by* $H_{c,v}$ *if they lie on different sides of the hyperplane, i.e.,*

$$\begin{aligned} \langle x,v\rangle &\geq c \ \textit{ for } x\in A, \quad \langle x,v\rangle \leq c \ \textit{ for } x \in B, \\ \textit{or } \langle x,v\rangle &\leq c \ \textit{ for } x\in A, \quad \langle x,v\rangle \geq c \ \textit{ for } x \in B. \end{aligned}$$

We say that A *and* B *are* ***strictly separated****, if these are strict inequalities.*

Theorem 3.8.2 (Separation Theorem)**.** *Let* $A \neq \varnothing$ *be closed and convex, and let* $z \notin A$. *Then there exists a hyperplane* $H_{c,v}$ *with some* $v \neq 0$ *and* $c \in \mathbb{R}$ *such that*

$$\langle z, v\rangle < c < \langle x, v\rangle \quad \textit{ for all } x \in A.$$

Proof. (Idea) By the Weierstrass theorem, $\min_{x\in A}\|z-x\| := \operatorname{dist}(z, A)$ is attained at some $x_* \in A$. Furthermore, $\operatorname{dist}(z, A) =: \delta > 0$ since $z \notin A$ and $A = \bar{A}$. Set

$$0 \neq v := x_* - z \quad \text{and} \quad c := \langle x_*, v\rangle.$$

Then $H_{c,v}$ separates z and A:

$$\begin{aligned} \langle z,v\rangle &= \langle x_* - v, v\rangle = c - \|v\|^2 < c, \\ \langle x,v\rangle &> c \text{ (this is checked by convexity of } A \text{).} \end{aligned}$$

□

Further improvements of this theorem

Theorem 3.8.3 (Supporting Theorem)**.** *Let* $A \neq \varnothing$ *be closed and convex, and let* $z \notin A$. *Then there exists a point* $x_0 \in \partial A$ *and a* ***supporting hyperplane*** $H_{c,v}$ *through* x_0 *such that*

$$\langle z, v\rangle < c := \langle x_0, v\rangle = \inf_{x\in A}\{\langle x, v\rangle\}.$$

Theorem 3.8.4 (Minkowski Theorem)**.** *Let* A *and* B *be disjoint and nonempty convex sets in* $\mathbb{R}^n$. *Then there exists a hyperplane* $H_{c,v}$ *separating* A *and* B, *i.e.,*

$$\langle x, v\rangle \leq c \leq \langle y, v\rangle \ \textit{ for all } x \in A \textit{ and } y \in B.$$

3.9 Farkas' Lemma

Every student has learned how to solve a system of linear equations, but solving systems of linear inequalities is less well-known. How can one solve $Ax \leq b$ for $x \geq 0$ (or show that there is no solution)?

The answer is given by **Farkas' lemma**. It was originally proved by Julius Farkas ("*Uber die Theorie der linearen Ungleichungen*", 1902). This lemma is at the core of *linear programming* and *game theory*. Farkas' lemma is also used to prove the *Kuhn-Tucker theorem* in nonlinear programming. Another important application is *arbitrage problems* in theoretical finance.

Farkas' lemma (roughly) states that a vector is either in a given convex cone or that there exists a (hyper-) plane separating this vector from the cone, but not both. So, Farkas' lemma is an example of a **theorem of alternatives** (or **dichotomy**): a theorem stating that of two systems, one or the other has a solution, but not both or none.

One of equivalent statements of Farkas' lemma is as follows:

Lemma 3.9.1 (Farkas' Lemma)**.** *Let A be an $m \times n$ matrix and $b \in \mathbb{R}^m$. Then either:*

(i) There is an $x \in \mathbb{R}^n$ such that $Ax = b$ and $x \geq 0$; or

(ii) There is a $y \in \mathbb{R}^m$ such that

$$\underbrace{y^{\mathrm{t}}}_{1\times m} \times \underbrace{A}_{m\times n} \geq 0 \quad \textit{and} \quad \langle y, b\rangle = \underbrace{y}_{1\times m} \times \underbrace{b}_{m\times 1} < 0.$$

Proof. The trivial claim is that (i) and (ii) contradict each other:

$$0 = \langle y, Ax - b\rangle = \underbrace{\langle yA, x\rangle}_{\geq 0} - \langle y, b\rangle \geq -(y, b) > 0.$$

Contradiction to (ii) implies (i) - This part is highly non-trivial! □

Most of the known proofs are based on *Separation Hyperplane Theorem* (Theorem 3.8.2).

Farkas' lemma is still considered a pedagogical annoyance: it has an obvious formulation but no elementary proof.

3.9.1 Geometric Interpretation

Let $a_1, \ldots, a_n \in \mathbb{R}^m$ denote the columns of a matrix A. In terms of these vectors, Farkas' lemma states that exactly one of the following two statements is true:

(i) There exist non-negative coefficients $x_1, \ldots, x_n \in \mathbb{R}$ such that $b = x_1a_1 + \cdots + x_na_n$.

(ii) There exists a vector $y \in \mathbb{R}^m$ such that $\langle a_j, y \rangle \geq 0$ for all $1 \leq j \leq n$ and $\langle y, b \rangle < 0$.

The vectors $x_1 a_1 + \cdots + x_n a_n \in \mathbb{R}^m$ with nonnegative coefficients $x_1, \ldots, x_n$ constitute the **convex cone** $K(a_1, \ldots, a_n)$ of the set $\{a_1, \ldots, a_n\}$, so the first statement says that b is in this cone.

The second statement says that there exists a vector y such that the angle of y with the vectors a_j is at most $90°$, while the angle of y with the vector b is more than $90°$. The hyperplane normal to this vector has the vectors a_j on one side and the vector b on the other side. Hence, this hyperplane separates the vectors in the cone $K(a_1, \ldots, a_n)$ and the vector b.

(One more equivalent) **Algebraic formulation:**

The system of inequalities $Ax \leq b$ has a solution $x \geq 0$ **if and only if** $\langle y, b \rangle \geq 0$ for any vector $y \geq 0$ such that $y^{\mathrm{t}} A \geq 0$.