EC9A0: Pre-sessional Advanced Mathematics Course

Lecture Notes: Unconstrained Optimisation
By Pablo F. Beker[1]

1 Infimum and Supremum

Definition 1. *Fix a set* $Y \subseteq \mathbb{R}$*. A number* $\alpha \in \mathbb{R}$ *is* an upper bound of Y *if* $y \leq \alpha$ *for all* $y \in Y$*, and is* a lower bound of Y *if the opposite inequality holds.*

Definition 2. *Number* $\alpha \in \mathbb{R}$ *is* the least upper bound of Y*, denoted* $\alpha = \sup Y$*, if:*
(1) α *is an upper bound of* Y*; and*
(2) $\gamma \geq \alpha$ *for any other upper bound* γ *of* Y*.*

Definition 3. *Analogously, number* $\beta \in \mathbb{R}$ *is* the greatest lower bound of Y*, denoted* $\beta = \inf Y$*, if:*
(1) β *is a lower bound of* Y*; and*
(2) if γ *is a lower bound of* Y*, then* $\gamma \leq \beta$*.*

Theorem 1. $\alpha = \sup Y$ *if and only if for all* $\varepsilon > 0$ *it is true that*
(a) $y < \alpha + \varepsilon$ *for all* $y \in Y$*; and*
(b) $\alpha - \varepsilon < y$ *for some* $y \in Y$ *.*

Proof: ($\Rightarrow$) Let $\alpha = \sup Y$ and $\varepsilon > 0$ be arbitrary. Since α is an upper bound of Y, then $y \leq \alpha$ for all $y \in Y$. Thus, $y \leq \alpha + \varepsilon$ for all $y \in Y$, that is, (a) is true. Suppose (b) is not true, that is, $\alpha - \varepsilon \geq y$ for all $y \in Y$. Then $\alpha - \varepsilon$ is an upper bound of Y that is strictly smaller than α, a contradiction. We conclude that $\alpha - \varepsilon < y$ for some $y \in Y$.

($\Leftarrow$) Suppose for all $\varepsilon > 0$, $y < \alpha + \varepsilon$ for all $y \in Y$. This implies that $\alpha \geq y$ for all $y \in Y$ and so it is an upper bound of Y. To get a contradiction, suppose α is not the supremum of Y. Then, there must exist another upper bound α' of Y such that $\alpha' < \alpha$. Let $\varepsilon' = \frac{\alpha - \alpha'}{2}$. Then there exists $y^{'}$ such that $\alpha^{'} = \alpha - (\alpha' - \alpha^{'}) < \alpha - \varepsilon' < y'$ which contradicts α' is an upper bound of Y. *Q.E.D.*

Corollary 1. *Let* $Y \subseteq \mathbb{R}$ *and* $\alpha \equiv \sup Y$*. Then there exists a sequence* $\{y_n\}_{n=1}^{\infty}$ *in* Y *that converges to* α*.*

2 Maximisers

From now on, maintain the assumptions that set $D \subseteq \mathbb{R}^K$, for a finite K, is nonempty.

We need a stronger concept of extremum, in particular one that implies that the extremum lies within the set. Thus,

Definition 4. *A point* $x \in \mathbb{R}$ *is the* the maximum of set $Y \subseteq \mathbb{R}$*, denoted* $x \equiv \max A$*, if* $x \in Y$ *and* $y \leq x$ *for all* $y \in Y$*.*

[1]Based on notes by Andrés Carvajal

The proofs of the following two results are left as exercises

THEOREM 2. *If* $\max Y$ *exists, then*

(a) it is unique.

(b) $\sup Y$ *exists and* $\sup Y = \max Y$*.*

Proof: Suppose $\max Y$ exists.
(a) Suppose $y_1 \neq y_2$ are maxima of the set Y. Then, $y_1 \in Y$ and $y_2 \in Y$. But then, by definition of maximum, $y_1 \leq y_2$ and $y_2 \leq y_1$ and we conclude that $y_1 = y_2$, a contradiction.
(b) Let γ be an arbitrary upper bound of Y. Since $\max Y \in Y$, then $\gamma \geq \max Y$ and $\max Y$ is an upper bound of Y, it follows that $\sup Y$ exists and $\max Y = \sup Y$. *Q.E.D.*

THEOREM 3. *If* $\sup Y$ *exists and* $\sup Y \in Y$*, then* $\max Y$ *exists and* $\max Y = \sup Y$*.*

Proof: Suppose $\sup Y$ exists and $\sup Y \in Y$. Then $\sup Y \geq y$ for all $y \in Y$ and $\sup Y \in Y$. Thus, $\max Y$ exists and $\max Y = \sup Y$. *Q.E.D.*

Now, it typically is of more interest in economics to find extrema of functions, rather than extrema of sets. To a large extent, the distinction is only apparent: what we will be looking for are the extrema of the image of the domain under the function.

DEFINITION 5. *A point* $\bar{x} \in D \subset \mathbb{R}^K$ *is a* a global maximizer of $f : D \to \mathbb{R}$ *if* $f(x) \leq f(\bar{x})$ *for all* $x \in D$*.*

DEFINITION 6. *A point* $\bar{x} \in D \subset \mathbb{R}^K$ *is the* a local maximizer of $f : D \to \mathbb{R}$ *if there exists some* $\varepsilon > 0$ *such that* $f(x) \leq f(\bar{x})$ *for all* $x \in B_\varepsilon(\bar{x}) \cap D$*.*

When $\bar{x} \in D$ is a local (global) maximizer of $f : D \to \mathbb{R}$, the number $f(\bar{x})$ is said to be *a local (the global) maximum of* f. Notice that, in the latter case, $f(\bar{x}) = \max f[D]$, although more standard notation for $\max f[D]$ is $\max_D f$ or $\max_{x \in D} f(x)$.[2] Notice that there is a conceptual difference between maximum and maximizer! Also, notice that a function can have only one global maximum even if it has multiple global maximizers, but the same is not true for the local concept. The set of maximizers of a function is usually denoted by $\text{argmax}_D f$.

By analogy, $b \in \mathbb{R}$ is said to be *the supremum of* $f : D \to \mathbb{R}$, denoted $b = \sup_D f$ or $b = \sup_{x \in D} f(x)$, if $b = \sup f[D]$. Importantly, note that there is no reason why $\exists x \in D$ such that $f(x) = \sup_D f$ even if the supremum is defined.

3 Existence

THEOREM 4 (Weierstrass). *Let* $C \subseteq D \subset \mathbb{R}^K$ *be nonempty and compact. If the function* $f : D \to \mathbb{R}$ *is continuous, then there are* $\bar{x}, \underline{x} \in C$ *such that* $f(\underline{x}) \leq f(x) \leq f(\bar{x})$ *for all* $x \in C$*.*

[2] A point $\bar{x} \in D$ is said to be *a local minimizer of* $f : D \to \mathbb{R}$ if there is an $\varepsilon > 0$ such that for all $x \in B_\varepsilon(\bar{x}) \cap D$ it is true that $f(x) \geq f(\bar{x})$. Point $\bar{x} \in D$ is said to be *a global minimizer of* $f : D \to \mathbb{R}$ if for every $x \in D$ it is true that $f(x) \geq f(\bar{x})$. From now on, we only deal with maxima, although the minimization problem is obviously covered by analogy.

Proof: Since C is compact and f is continuous, then $f[C]$ is compact. By Corollary 1, there is $\{y_n\}_{n=1}^{\infty}$ in $f[C]$ such that. $y_n \to \sup f[C]$. Since $f[C]$ is compact, then it is closed. Therefore, $\sup f[C] \in f[C]$. Thus, there is $\overline{x} \in C$ such that $f(\overline{x}) = \sup f[C]$. By def. of sup, $f(\overline{x}) \geq f(x)$ for all $x \in C$. Existence of $\underline{x}$ is left as an exercise. *Q.E.D.*

The importance of this result is that when the domain of a continuous function is closed and bounded, then the function does attain its maxima and minima within its domain.

4 Characterizing maximizers

Even though maximization is not a differential problem, when one has differentiability there are results that make it easy to find maximizers. For this section, we take set D to be open.

4.1 Problems in $\mathbb{R}$

For simplicity, we first consider the case $K = 1$ and $D \subset \mathbb{R}$.

LEMMA 1. *Suppose that $f : D \to \mathbb{R}$ is differentiable. If $f'(\bar{x}) > 0$, then there is some $\delta > 0$ such that for every $x \in B_\delta(\bar{x}) \cap D$ we have $f(x) > f(\bar{x})$ if $x > \bar{x}$, and $f(x) < f(\bar{x})$ if $x < \bar{x}$.*

Proof: Let $\varepsilon \equiv \frac{f'(\bar{x})}{2} > 0$. Then, $f'(\bar{x}) - \varepsilon > 0$. By def. of f', there exists $\delta > 0$ such that

$$|\tfrac{f(x)-f(\bar{x})}{x-\bar{x}} - f'(\bar{x})| < \varepsilon, \forall x \in B'_\delta(\bar{x}) \cap D.$$

Hence, $\frac{f(x)-f(\bar{x})}{x-\bar{x}} > f'(\bar{x}) - \varepsilon > 0$. *Q.E.D.*

The analogous result for the case of a negative derivative is presented, without proof, as the following corollary.

COROLLARY 2. *Suppose that $f : D \to \mathbb{R}$ is differentiable. If $f'(\bar{x}) < 0$, then there is some $\delta > 0$ such that for every $x \in B_\delta(\bar{x}) \cap D$ we have $f(x) < f(\bar{x})$ if $x > \bar{x}$, and that $f(x) > f(\bar{x})$ if $x < \bar{x}$.*

THEOREM 5. *Suppose that $f : D \to \mathbb{R}$ is differentiable. If $\bar{x} \in D$ is a local maximizer of f, then $f'(\bar{x}) = 0$.*

Proof: Suppose not: $f'(\bar{x}) \neq 0$. If $f'(\bar{x}) > 0$, then, by Lemma 1, there is $\delta > 0$ such that for all $x \in B_\delta(\bar{x}) \cap D$ satisfying $x > \bar{x}$ we have that $f(x) > f(\bar{x})$. Since $\bar{x}$ is a local maximizer of f, then there is $\varepsilon > 0$ such that $f(x) \leq f(\bar{x})$ for all $x \in B_\varepsilon(\bar{x}) \cap D$. Since $\bar{x} \in \text{int}(D)$, there is $\gamma > 0$ such that $B_\gamma(\bar{x}) \subseteq D$. Let $\beta = \min\{\varepsilon, \delta, \gamma\} > 0$. Clearly, $(\bar{x}, \bar{x} + \beta) \subset B'_\beta(\bar{x}) \neq \varnothing$ and $B'_\beta(\bar{x}) \subseteq D$. Moreover, $B'_\beta(\bar{x}) \subseteq B_\delta(\bar{x}) \cap D$ and $B'_\beta(\bar{x}) \subseteq B_\varepsilon(\bar{x}) \cap D$. This implies that there exists x such that $f(x) > f(\bar{x})$ and $f(x) \leq f(\bar{x})$, an obvious contradiction. A similar contradiction appears if $f'(\bar{x}) < 0$, by Corollary 2. *Q.E.D.*

THEOREM 6. *Let $f : D \to \mathbb{R}$ be $\mathbb{C}^2$. If $\bar{x} \in D$ is a local maximizer of f then $f''(\bar{x}) \leq 0$.*

Proof: Since $\bar{x} \in \text{int}(D)$, there is $\varepsilon > 0$ for which $B_\varepsilon(\bar{x}) \subseteq D$. Fix $h \in B_\varepsilon(0)$. Since f is twice differentiable, Taylor's Theorem implies there is some x_h^* in the interval joining $\bar{x}$ and $\bar{x}+h$, such that

$$f(\bar{x}+h) = f(\bar{x}) + f'(\bar{x})h + \frac{1}{2}f''(x_h^*)h^2$$

Since $\bar{x}$ is a local maximizer, there is a $\delta > 0$ such that $f(x) \leq f(\bar{x})$ for all $x \in B_\delta(\bar{x}) \cap D$. Let $\beta = \min\{\varepsilon, \delta\} > 0$. By construction, for any $h \in B'_\beta(0)$ one has that

$$f'(\bar{x})h + \frac{1}{2}f''(x_h^*)h^2 = f(\bar{x}+h) - f(\bar{x}) \leq 0.$$

By Theorem 5, since f is differentiable and $\bar{x}$ is a local maximizer, $f'(\bar{x}) = 0$. Since $h \in B'_\beta(0)$, then $f''(x_h^*)h^2 \leq 0$ and, hence, $f''(x_h^*) \leq 0$. Now, letting $h \to 0$, we obtain that $\lim_{h\to 0} f''(x_h^*) \leq 0$. It follows hence that $f''(\bar{x}) \leq 0$ because f'' is continuous and each x_h lies in the interval joining $\bar{x}$ and $\bar{x}+h$. *Q.E.D.*

Notice that the last theorems only give us necessary conditions:[3] this is not a tool that tells us which points are local maximizers, but it tells us what points are not. A complete characterization requires both necessary and sufficient conditions. We now find sufficient conditions.

THEOREM 7. *Suppose that $f : D \to \mathbb{R}$ is twice differentiable. If $f'(\bar{x}) = 0$ and $f''(\bar{x}) < 0$, then $\bar{x}$ is a local maximizer.*

Proof: Since $f : D \to \mathbb{R}$ is twice differentiable and $f''(\bar{x}) < 0$, by Corollary 2 there is $\delta > 0$ such that for $x \in B_\delta(\bar{x}) \cap D$, (i)$f'(x) < f'(\bar{x}) = 0$ if $x > \bar{x}$; and (ii)$f'(x) > f'(\bar{x}) = 0$, when $x < \bar{x}$. Since $x \in \text{int}(D)$, there is an $\varepsilon > 0$ such that $B_\varepsilon(\bar{x}) \subseteq D$. Let $\beta = \min\{\delta, \varepsilon\} > 0$. By the Mean Value Theorem, we have that,

$$f(x) = f(\bar{x}) + f'(x^*)(x - \bar{x}) \text{ for all } x \in B_\beta(\bar{x}),$$

for some x^* in the interval between $\bar{x}$ and x (why?). Thus, if $x > \bar{x}$, we have $x^* \geq \bar{x}$, and, therefore, $f'(x^*) \leq 0$, so that $f(x) \leq f(\bar{x})$. On the other hand, if $x < \bar{x}$, then $f'(x^*) \geq 0$, so that $f(x) \leq f(\bar{x})$. *Q.E.D.*

Notice that the sufficient conditions are stronger than the set of necessary conditions: there is a little gap that the differential method does not cover.

EXAMPLE 1. *Consider $f(x) = x^4 - 4x^3 + 4x^2 + 4$. Note that*

$$f'(x) = 4x^3 - 12x^2 + 8x = 4x(x-1)(x-2).$$

Hence, $f'(x) = 0$ if and only if $x \in \{0, 1, 2\}$. Since $f''(x) = 12x^2 - 24x + 8$,

$$f''(0) = 8 > 0, f''(1) = -4 < 0, \text{ and } f''(2) = 8 > 0.$$

Note that $x = 0$ and $x = 2$ are local min of f and $x = 1$ is a local max. $x = 0$ and $x = 2$ are global min but $x = 1$ is not a global max.

[3] And there are further necessary conditions.

4.2 Higher-dimensional problems

We now allow for functions defined on higher-dimensional domains (namely $K \geq 2$ and $D \subset \mathbb{R}^K$). The results of the one-dimensional case generalize as follows.

THEOREM 8. *If $f : D \to \mathbb{R}$ is differentiable and $\overline{x} \in D$ is a local maximizer of f, then $Df(\overline{x}) = 0$.*

THEOREM 9. *If $f : D \to \mathbb{R}$ is $\mathbb{C}^2$ and $\overline{x} \in D$ is a local maximizer of f, then $D^2 f(\overline{x})$ is negative semidefinite.*

As before, these conditions do not tell us which points are maximizers, but only which ones are not. Before we can argue sufficiency, we need to introduce the following lemma.

THEOREM 10. *Suppose that $f : D \to \mathbb{R}$ is $\mathbb{C}^2$ and let $\bar{x} \in D$. If $Df(\bar{x}) = 0$ and $D^2 f(\bar{x})$ is negative definite, then $\bar{x}$ is a local maximizer.*

EXAMPLE 2. *Consider $f(x, y) = x^3 - y^3 + 9xy$. Note that*

$$\begin{aligned} f'_x(x, y) &= 3x^2 + 9y \\ f'_y(x, y) &= -3y^2 + 9x \end{aligned}$$

Hence, $f'_x(x, y) = 0$ and $f'_y(x, y) = 0 \iff (x, y) \in \{(0, 0), (3, -3)\}$.

$$D^2 f(x) = \begin{pmatrix} f''_{xx} & f''_{yx} \\ f''_{xy} & f''_{yy} \end{pmatrix} = \begin{pmatrix} 6x & 9 \\ 9 & -6y \end{pmatrix}.$$

Note that

$$f''_{xx} = 6x, |D^2 f(x, y)| = -36xy - 81.$$

At $(0, 0)$ the two minors are 0 and -81. Hence, $D^2 f(0, 0)$ is indef. At $(3, -3)$ the two minors are 18 and 243. Hence, $D^2 f(3, -3)$ is positive definite and $(3, -3)$ is a local min. $(3, -3)$ is not a global min since $f(0, n) = -n^3 \to -\infty$ as $n \to \infty$.

5 Global Maxima

Note that the results that we obtained in the previous sections hold only locally. We now study the extent to which local extrema are, in effect, global extrema.

THEOREM 11. *Suppose $D \subset \mathbb{R}$ and $f : D \to \mathbb{R}$ is $\mathbb{C}^1$ in the interior of D and:*

1. *the domain of f is an interval in $\mathbb{R}$.*
2. *$\overline{x}$ is a local maximum of f, and*
3. *$\overline{x}$ is the only solution to $f'(x) = 0$ on D.*

Then, $\overline{x}$ is the global maximum of f.

Proof: In order to get a contradiction, suppose there exists x^* such that $f(x^*) > f(\overline{x})$. Without loss in generality, assume $x^* > \overline{x}$. Since $\overline{x}$ is a local maximum, there exists $\varepsilon > 0$ such that $f(x) < f(\overline{x})$ for all $x \in B_\varepsilon(\overline{x}) \cap D$. Hence, f is strictly decreasing to the right of $\overline{x}$ and so $f'(x) < 0$ for some $x' \in B_\varepsilon(\overline{x}) \cap D$. Since $f(x^*) > f(\overline{x})$, there must also exist $x' < x'' < x^{**}$ such that $f'(x'') > 0$. But then, by continuity of f' there exists $x' < x < x''$ such that $f'(x) = 0$, a contradiction. *Q.E.D.*

To obtain further results, we need to introduce some definitions.

Definition 7. *Let D be a convex subset of $\mathbb{R}^K$. Then, $f : U \to \mathbb{R}$ is a*

- *concave function if for all $x, y \in U$, and for all $\theta \in [0, 1]$,*

$$f(\theta x + (1-\theta)y) \geq \theta f(x) + (1-\theta)f(y)$$

- *strictly concave function if for all $x, y \in U$, $x \neq y$, and for all $\theta \in (0, 1)$,*

$$f(\theta x + (1-\theta)y) > \theta f(x) + (1-\theta)f(y)$$

Definition 8. *Let D be a convex subset of $\mathbb{R}^K$. Then, $f : U \to \mathbb{R}$ is a*

- *quasi-concave function if for all $x, y \in U$, and for all $\theta \in [0, 1]$,*

$$f(x) \geq f(y) \implies f(\theta x + (1-\theta)y) \geq f(y)$$

- *strictly quasi-concave function if for all $x, y \in U$, $x \neq y$, and for all $\theta \in (0, 1)$,*

$$f(x) \geq f(y) \implies f(\theta x + (1-\theta)y) > f(y)$$

The following Theorem shows that quasi-concavity and strict quasi-concavity are ordinal properties of a function.

Theorem 12. *Suppose $D \subset \mathbb{R}^K$, $f : D \to \mathbb{R}$ is quasi-concave and $g : f(D) \to \mathbb{R}$ is nondecreasing. Then $g \circ f : D \to \mathbb{R}$ is quasi-concave. If f is strictly quasi-concave and g is strictly increasing, then $g \circ f$ is strictly quasi-concave.*

Proof: Consider any $x, y \in U$. If f is quasi-concave, then $f(\theta x+(1-\theta)y) \geq \min\{f(x), f(y)\}$. Therefore, g nondecreasing implies

$$g(f(\theta x + (1-\theta)y)) \geq g(\min\{f(x), f(y)\}) = \min\{g(f(x)), g(f(y))\}.$$

If f is strictly quasi-concave, then for $x \neq y$, we have $f(\theta x + (1-\theta)y) > \min\{f(x), f(y)\}$. Therefore, if g is strictly increasing, we have

$$g(f(\theta x + (1-\theta)y)) > g(\min\{f(x), f(y)\}) = \min\{g(f(x)), g(f(y))\}.$$

Q.E.D.

Theorem 13. *Suppose that $D \subset \mathbb{R}^K$ is convex and $f : D \to \mathbb{R}$ is a concave function. If $\bar{x} \in D$ is a local maximiser of f, it is also a global maximiser.*

Proof: We argue by contradiction: suppose that $\bar{x} \in D$ is a local maximizer of f, but it is not a global maximizer. Then, the proof follows from four steps:
1. There is $\varepsilon > 0$ such that $f(x) \leq f(\bar{x})$ for all $x \in B_\varepsilon(\bar{x}) \cap D$;
2. There is $x^* \in D$ such that $f(x^*) > f(\bar{x})$. Clearly, $x^* \notin B_\varepsilon(\bar{x})$, and so $\|x^* - \bar{x}\| \geq \varepsilon$.
3. Now, since D is convex and f is concave, we have that for $\theta \in [0, 1]$,

$$f(\theta x^* + (1-\theta)\bar{x}) \geq \theta f(x^*) + (1-\theta)f(\bar{x}),$$

but, since $f(x^*) > f(\bar{x})$, we further have that if $\theta \in (0,1]$, then $\theta f(x^*) + (1-\theta)f(\bar{x}) > f(\bar{x})$, so that $f(\theta x^* + (1-\theta)\bar{x}) > f(\bar{x})$.
4. Now, consider $\theta^* \in (0, \varepsilon/\|x^* - \bar{x}\|)$. Clearly, $\theta^* \in (0,1)$, so $f(\theta^* x^* + (1-\theta^*)\bar{x}) > f(\bar{x})$. However, by construction,

$$\|(\theta^* x^* + (1-\theta^*)\bar{x}) - \bar{x}\| = \theta^*\|x^* - \bar{x}\| < (\frac{\varepsilon}{\|x^* - \bar{x}\|})\|x^* - \bar{x}\| = \varepsilon,$$

which implies that $(\theta^* x^* + (1-\theta^*)\bar{x}) \in B_\varepsilon(\bar{x})$, and, moreover, by convexity of D, we have that $(\theta^* x^* + (1-\theta^*)\bar{x}) \in B_\varepsilon(\bar{x}) \cap D$. This contradicts the fact that $f(x) \leq f(\bar{x})$ for all $x \in B_\varepsilon(\bar{x}) \cap D$. *Q.E.D.*

An analogous reasoning let us show that strict quasi concavity implies that a local maximiser is also a global maximiser. Note that the result does not hold if we just assume quasi-concavity (why?).

Theorem 14. *Suppose that $D \subset \mathbb{R}^K$ is convex and $f : D \to \mathbb{R}$ is a strictly quasi-concave function. If $\bar{x} \in D$ is a local maximizer of f, then it is also a global maximizer.*

Proof: The proof follows by steps 1, 2 and 4 in the proof of Theorem 13. *Q.E.D.*

For the sake of practice, it is a good idea to work out an exercise like Exercise 3.9 in pages 57 and 58 of Simon and Blume.

6 Uniqueness

Theorem 15. *Suppose $U \subset \mathbb{R}^K$, $f : U \to \mathbb{R}$ attains a maximum on U.*
(a) If f is quasi-concave, then the set of maximisers is convex.
(b) If f is strictly quasi-concave, then the maximiser of f is unique.

Proof: (a) Suppose x and y are maximisers of f. Let $z \in U$ and $\theta \in [0,1]$ be arbitrary. Since $f(x) = f(y)$, f is quasi-concave and x is a maximiser, then

$$f(\theta x + (1-\theta)y) \geq f(x) \geq f(z)$$

and so it follows that $\theta x + (1-\theta)y$ is a maximiser of f.
(b) Suppose x and y are maximisers of f but $x \neq y$. Then,

$$f\left(\frac{1}{2}x + \frac{1}{2}y\right) > \min\{f(x), f(y)\} = f(x)$$

which implies that x is not a maximiser of f, a contradiction. *Q.E.D.*