These lecture notes are meant to be used by students entering the University of Mannheim Master program in Economics. They constitute the base for a pre-course in mathematics; that is, they summarize elementary concepts with which all of our econ grad students must be familiar. More advanced concepts will be introduced later on in the regular coursework.

A thorough knowledge of these basic notions will be assumed in later coursework. No prerequisite beyond high school mathematics are required.

Although the wording is my own, the definitions of concepts and the ways to approach them is strongly inspired by various sources, which are mentioned explicitly in the text or at the end of the chapter.

—

*This version: September 5th 2015
†Center for Doctoral Studies in Economic and Social Sciences. Contact: justin.leduc@uni-mannheim.de
Contents

1 Introduction 3

2 A Last Pinch of Vocabulary 4
   2.1 Extrema, Infima, Suprema, and Compactness ................................. 4
   2.2 Level Sets and The Implicit Function Theorem ................................ 7

3 Unconstrained Optimization 12
   3.1 First Order Necessary Conditions .................................................. 12
   3.2 Second Order Necessary Conditions ................................................. 14
   3.3 Sufficient Conditions ...................................................................... 16

4 Constrained Optimization I: Convex Optimization 17
   4.1 Geometrical Interpretation of the Constraints - Convex Optimization .... 18
   4.2 The Karush-Kuhn-Tucker Necessary Conditions ............................... 20
   4.3 Sufficient Conditions ...................................................................... 25
   4.4 Insights From The Associated Unconstrained Problem ....................... 26
   4.5 In Practice ...................................................................................... 27

5 Constrained Optimization II: Non-Convex Optimization 28
   5.1 Quasi-convex Constraints ................................................................. 29
   5.2 Linear Equality Constraints .............................................................. 29
   5.3 Non-Linear Equality Constraints - The Lagrange Conditions .............. 29
1 Introduction

We have seen a lot of new concepts in the last three chapters. Although that represented a lot of work, it was not in vain, and this chapter should convince you of this fact. One of the fascinating aspects of mathematics stems from the richness of the perspectives that it suggests for a given problem. Under one specific perspective, sometimes, the problem to be solved actually appears quite complicated, while under another it appears to be quite simple. Part of your job as an economist, actually, will consist in finding the very perspective which enables you to solve the issue at hand.

There is no need to emphasize the importance of optimization in economics. If you have studied economics, you already have experienced it. If not, you will in the coming months. Throughout this chapter, I focus on real-valued multivariate functions for two reasons: (i) they represent the kind of functions you will mostly be working with; (ii) it allows us not to dig too far into order theory\(^1\). It is important to keep in mind, however, that the methods we will use do not depend specifically on this fact. Rather, they depend on the fact that we are working in a vector space with a specific structure. Namely, a Hilbert space, i.e., a complete vector space with an inner product. But even the requirement of the existence of an inner product may be partially relaxed, and if you are to face such problems, then simple generalizations of our methods are available and nicely described in books such as that of Luenberger\(^2\).

After introducing a last pinch of vocabulary\(^3\), we proceed with optimization techniques, which are an application of the concepts we have studied in the previous chapter. We will start with the unconstrained case, which should make it easier for you to relate the unfamiliar multivariate case to the more familiar univariate case. We will then consider two important classes of constrained optimization, namely, convex programming with inequality constraints and convex programming with equality constraints. Throughout the entire chapter, I will most often consider maximization problems. That will not seem very natural if you studied mathematics or engineering, but it seems most natural to economists. You may consider the material as understood if you feel confident rephrasing everything in terms of minimization problems.

And while we are there, let me restate formally what is meant by a constrained minimization or constrained maximization problem.

---

\(^1\)A theory that formalizes the notion of “bigger” and “smaller” to sets other than the real line.

\(^2\)You may actually hear about duality theory during your master’s years. This theory is specifically concerned about the extension of the methods we present here to vector spaces that cannot be equipped with an inner product. It is also useful in Hilbert spaces, as it captures some of the geometrical intuition lying behind our techniques.

\(^3\)Not that which calls for caution, but that which enhances flavor!
($P_{\text{min}}$) \[ \begin{align*} \text{minimize} & \quad f(x) \\ \text{subject to} & \quad g_i(x) = 0, \ i = 1, \ldots, m. \\ & \quad h_i(x) \leq 0, \ i = 1, \ldots, k. \end{align*} \]

Where $f$ is our objective function, $\text{dom}(f)$ is the domain of $f$, $x$ is the vector of choice variables, $g_i$’s characterize the equality constraints, and $h_i$’s characterize the inequality constraints. Similarly, a standard constrained maximization problem takes the following form:

($P_{\text{max}}$) \[ \begin{align*} \text{maximize} & \quad f(x) \\ \text{subject to} & \quad g_i(x) = 0, \ i = 1, \ldots, m. \\ & \quad h_i(x) \leq 0, \ i = 1, \ldots, k. \end{align*} \]

If $m = k = 0$, we say that the problem is unconstrained.

2 A Last Pinch of Vocabulary

2.1 Extrema, Infima, Suprema, and Compactness

Before starting the whole optimization process, it is wise to (i) formally define what we are looking for (i.e. maximum or minimum) and (ii) evaluate how likely it is that we will find it. The second task is generally achieved by looking at specific properties of the objective function and of its domain. So let us proceed!

**Definition: (Maximum and Minimum)**

Let $X$ be a subset of $\mathbb{R}$. An element $\bar{x}$ in $X$ is called a maximum in $X$ if, for all $x$ in $X$, $\bar{x} \geq x$. An element $x$ in $X$ is called a minimum in $X$ if, for all $x$ in $X$, $x \leq \bar{x}$. If the above inequalities are strict, one speaks of strict maximum (resp. strict minimum). A maximum or a minimum is often simply referred to as an extremum.

**Remark:** Already at this stage one can get a hint at why real-valuedness of the function makes things easier. Assume the function $f$ maps $\mathbb{R}^n$ into $\mathbb{R}^m$ and we were looking for an extremum. Let $x = (x_1, x_2, \ldots, x_m)$ and $y = (y_1, y_2, \ldots, y_m)$ be two vectors in $\mathbb{R}^m$. One says that $x \geq y$ if and only if, for all $i = 1, .., m$, $x_i \geq y_i$. Further, $x > y$ if and only if, for all $i = 1, .., m$, $x_i \geq y_i$ and for at least one $i$ $x_i > y_i$. A symmetric definition applies for $\leq$ and $<$. When $m = 1$, this “order” coincides with the one you usually apply on real numbers. When $m > 1$, it remains intuitive, but, unfortunately, it is not complete in $\mathbb{R}^m$! That is, one can find

---

I have assumed here $\forall i \ \text{dom}(f) \subset \text{dom}(g_i)$ and $\text{dom}(f) \subset \text{dom}(h_i)$, i.e. all constraints are well defined on the domain of $f$. 


elements x and y in \( \mathbb{R}^m \) such that neither \( x \geq y \) nor \( y \geq x \) (find examples!). Yet, if you pay attention to the definition of an extremum, completeness is crucial, as one must be able to compare the candidate to every element of the set \( X \).

**Definition:** *(Local and Global Maximizers)*

Let \( f \) be a real-valued function defined on \( X \subset \mathbb{R}^n \). A point \( \bar{x} \) in \( X \) is:

- A *global maximizer* for \( f \) on \( X \) if and only if:
  \[ \forall x \in X, \ f(\bar{x}) \geq f(x) \]

- A *strict global maximizer* for \( f \) on \( X \) if and only if:
  \[ \forall x \in X, \ x \neq \bar{x}, \ f(\bar{x}) > f(x) \]

- A *local maximizer* for \( f \) on \( X \) if and only if there is an \( \varepsilon > 0 \) such that:
  \[ \forall x \in X \cap B_\varepsilon(\bar{x}), \ f(\bar{x}) \geq f(x) \]

- A *strict local maximizer* for \( f \) on \( X \) if and only if there is an \( \varepsilon > 0 \) such that:
  \[ \forall x \in X \cap B_\varepsilon(\bar{x}), \ x \neq \bar{x}, \ f(\bar{x}) > f(x) \]

**Remark:** Note that, *(Strict) global maximizers are, by definition, (strict) local maximizers.*

(Strict) global and local minimizers can be defined symmetrically. Maximizers and minimizers are sometimes simply referred to as extremizers.

**Fact:** *(Equivalence Between Minimizers and Maximizers)*

Consider a problem of the form \( \mathcal{P}_{\text{min}} \). \( \bar{x} \) is a local (resp. global) extremizer for \( \mathcal{P}_{\text{min}} \) if and only if it is a local (resp. global) extremizer for a problem of the form \( \mathcal{P}_{\text{max}} \) with identical constraints but \(-f(x)\) as an objective function.

Further, the extremum associated is easily obtained by multiplying the image of the extremizer by \(-1\). As a consequence, in the sequel, we can, without loss of insight, restrict our attention to maximization problems.

**PLEASE CONSIDER \( \mathcal{P}_{\text{max}} \) AS A CANONICAL SHAPE AND ALWAYS TRY TO RESHAPE YOUR PROBLEM SO THAT IT FITS IT!**

Let’s now have a look at the conditions which guarantee the existence of an extremum.

---

5Basic facts about order theory can be found in section 3(b) of De la Fuente’s Chapter 1. A very nice and extensive treatment is found in Ok’s book: [https://files.nyu.edu/eo1/public/books.html](https://files.nyu.edu/eo1/public/books.html)
Definition: (Bounded Set)
Let $X$ be a nonempty subset of $\mathbb{R}^n$. We say that $X$ is bounded if and only if $X$ can be contained within a ball centered at the origin, i.e. there exists a positive $M$ in $\mathbb{R}$ such that $X \subset B_M(0)$.

Proposition: (Upper bound, Lower Bound, and Boundedness)
Let $X$ be a nonempty subset of $\mathbb{R}^n$. If $X$ is bounded, then $X$ has an upper bound, i.e. there exists some $u$ in $\mathbb{R}$ such that for all $x$ in $X$ we have: $u \geq x$. Further, $X$ has a lower bound, i.e. there exists some $l$ in $\mathbb{R}$ such that for all $x$ in $X$ we have: $l \leq x$. Reciprocally, if $X$ has both an upper and a lower bound, then $X$ is bounded.

Definition: (Supremum, Infimum)
Let $X$ be a nonempty subset of $\mathbb{R}^n$. The smallest upper bound is called the supremum of $X$, denoted $\sup(X)$. The largest lower bound is called the infimum of $X$, denoted $\inf f(X)$.

Remark: The two previous statements are using inequalities in higher order spaces. The previous remark briefly introduced them. Here, the lack of completeness isn't an issue because the upper bound (resp. lower bound) need not be an element of the set, but can instead be any element in the space!

Theorem: (Heine-Borel - Characterization of Compactness in $\mathbb{R}^n$)
Let $X$ be a nonempty subset of $\mathbb{R}^n$. Then, $X$ is compact if and only if $X$ is bounded and closed.

Remark: This is actually a characterization of compactness valid only in finite dimensional spaces (the Euclidean spaces are finite dimensional spaces) and not a definition. Yet, for our purpose, you may take it as a definition.

Theorem: (Weierstrass - Extreme Value Theorem)
Let $X$ be a nonempty compact subset of $\mathbb{R}^n$ and $f : X \to \mathbb{R}$ be continuous. Then, $f$ is bounded in $X$ and attains both its maximum and its minimum in $X$. That is, there exist points $x_M$ and $x_m$ in $X$ such that $f(x_M) = \sup f(X)$ and $f(x_m) = \inf f(X)$.
2.2 Level Sets and The Implicit Function Theorem

I have mentioned level sets already in the last chapter. They constitute the second classical geometrical representation of a function and are extremely important to understand the geometry of optimization theory. I first provide a formal definition and then discuss the intuition.

**Definition: (Level Sets)**
Let $X$ be a nonempty subset of $\mathbb{R}^n$, $f : X \to \mathbb{R}$, and $c$ be an element of $\mathbb{R}$. The $c$-level set of $f$ is the set $L_c^f := \{x| x \in X, f(x) = c\}$. The $c$-lower level set of $f$ is the set $L_c^{f-} := \{x| x \in X, f(x) \leq c\}$. The strict $c$-lower level set of $f$ is the set $L_c^{f--} := \{x| x \in X, f(x) < c\}$. The $c$-upper level set and the strict $c$-upper level set of $f$ are defined symmetrically.

The central concept is that of the level set here. In words, each level set is associated to a point in the codomain and depicts only those elements of the domain which are mapped into this point. A readily understood advantage of such a representation, therefore, is the fact that it requires less dimensions than the classical graph representation, which lies in the Cartesian product of the domain and the codomain of the function. When the function investigated is real-valued, it is as if we were to “split” the vertical dimension at every point, and associate a level set to every of these points. For instance, let $f$ be an “altitude” function, which takes as an input the geographical coordinates (say $x_1$ is the latitude and $x_2$ the longitude) and gives as an output the altitude of the location. *The level sets are then simply the sets of geographical coordinates which share the same altitude.*

![Figure 1: Level Sets in Geography (source: http://canebrake13.com/fieldcraft/map\'compass.php)](figure)

---

6When $f$ is a bi-dimensional real-valued function (as is our “altitude” function), level sets are often referred to as *contour lines*, so you may sometimes meet this term instead.
Another example, which should ring a bell to those who studied economics before: *indifference curves*! Recall that an indifference curve represents the set of bundles which provide exactly the same utility to our economic agent. In other words, one fixes a utility level $\bar{u}$ and looks for the combinations of good which yield exactly that level of utility.

![Indifference Map of a Cobb-Douglas Utility Function](image)

Figure 2: Indifference Map of a Cobb-Douglas Utility Function

With this example, an illustration of the upper- and lower-level sets is easily reached as well. Namely, the (strict) $\bar{u}$-upper level set is the set of all bundles which yield a utility level (resp. strictly) higher than $\bar{u}$. The (strict) $\bar{u}$-lower level set is the set of all bundles which yield a utility level (resp. strictly) lower than $\bar{u}$.

**Remark:** You may note that I could also have defined level sets as the inverse image of $c$ under $f$, $L^f_c := f^{-1}(c)$. This may help you getting clear with the concept: one fixes a value of interest, $c$, that the function may take. The level set is then the collection of all the antecedents of $c$.

You may have noticed the similarity in the definition of a level set and that of a hyperplane. This similarity isn’t by luck, and is actually the reason that justifies the importance of the characterization of hyperplanes we have seen in chapter 2. Hyperplanes are special instances of level sets, namely, level sets of *linear real-valued functions*. This interpretation of hyperplanes is quite important to keep in mind, as it will be the source of useful geometric intuitions which support our optimization techniques. Indeed, consider a function $f$ with domain $X$ in $\mathbb{R}^n$ and let us get back to the total differential. It is defined given a point $\bar{x}$ in $X$ where $f$ is differentiable, and associates $\nabla f(\bar{x}) \cdot x$ to any point $x$ in $\mathbb{R}^n$. The gradient at $\bar{x}$ is an $n \times 1$ vector. We have seen in chapter 1 that each $n \times 1$ vector could be identified with a linear map from $\mathbb{R}^n$ to $\mathbb{R}$. We now further understand that such a linear map can be geometrically represented via hyperplanes in $\mathbb{R}^n$, namely, hyperplanes of the form \[ \{ x \in \mathbb{R}^n : \nabla f(\bar{x}) \cdot x = c, c \in \mathbb{R} \}. \] And, more specifically, letting $c = \nabla f(\bar{x}) \cdot \bar{x}$, one obtains...
the hyperplane tangent to $L^f_{f(\bar{x})} := \{x \in \mathbb{R}^n : f(x) = f(\bar{x})\}$ at $\bar{x}$.

Optimality conditions, which, as we will soon see, are statements about hyperplanes and level sets.

Figure 3: The hyperplane tangent to $L^f_{f(\bar{x})}$ at $\bar{x}$ is a level set of the total differential at $\bar{x}$!

In reality the control variable may not be as precisely controlled as in the mathematical problem. Therefore, an important exercise, known as comparative statics, consists in testing the robustness of the optimality to small variations around the optimal point. Sufficient conditions for such an exercise to be feasible are given by the implicit function theorem and, given the importance of comparative statics in your curriculum, allow me to open a small parenthesis on that topic now.

Let us start with implicit functions. Very often, when our variable of interest is $y$, we are able to express it as an explicit function of our explanatory variables and parameters. That is, we have a relation of the form $y = f(x)$, where $y$ has been isolated on one side of the equation and $f$ is a rule that associate to every $x$ exactly one $y$. When such an isolation is not possible, we generally only can end up with a relation of the form $F(y, x) = 0$, where $y$ is “pooled together” with the other variables on one side of the equation. This may arise even in very simple cases. For instance, consider the function $y = x^2$ defined on $\mathbb{R}$, but assume $x$, rather than $y$, is our variable of interest. Because such a function is not injective, it is not possible for you to back up a unique value of $x$ if the only information I give you $y = 4$, i.e. you cannot express $x$ as a function of $y$, unless I provide you with more information (e.g. $x \geq 0$). In economics in particular, it is important to make sure that such cases do not arise. Indeed, because we wish to proceed at a certain level of generality, we are usually not willing to specify the utility function $u(x)$ (or the production function $f(x)$). Yet, when one considers the level set of function $u$, i.e. $L^u_{\bar{u}} = \{x | x \in X, u(x) = \bar{u}\}$, one defines an implicit

\footnote{A simple Taylor approximation of the characterizing equation of $L^f_{f(\bar{x})}$ around $\bar{x}$ yields this result. Make sure you can derive that result formally!}

\footnote{Which will also prove useful in the section on non-convex optimization!}
relation between utility levels achieved and goods, \( F(x, \bar{u}) := u(x) - \bar{u} = 0 \), and, given that the optimality conditions will pin down a point on this level set, one needs to make sure that the function allows for an explicit derivation of any \( x \) that would simultaneously satisfy \( F(x, \bar{u}) = 0 \) and the optimality conditions.

If you look at the simple Cobb-Douglas case depicted above, the question as to whether our level set can be seen as a function that relates bananas and apples might appear to be answered by a straightforward and unqualified yes. But consider now the case of satiated preferences. Satiated preferences can be expressed, for instance, by a Gaussian function, e.g. let:

\[
u(x_1, x_2) = \exp \left( - \frac{(x_1 - 2)^2}{1.1} + \frac{(x_2 - 2)^2}{1.1} \right)\]

Figure 4: Indifference Map of Satiated Preferences

Clearly, in this case, our indifference curves can not be seen as expressing good 1 (sodas) as a function of good 2 (burgers) any more, for it sometimes associates two images to a single element of the domain – i.e. it is not injective. Let us look again at the simple case of \( y = x^2 \) to get a hint at how one can possibly deal with the issue. There, we suggested that, with more information, namely, a restriction of the domain to \( \mathbb{R}_+ \), one could achieve injectivity, and, thereby, express \( x \) as a function of \( y: x = y^{\frac{1}{2}}. \) If we are to follow this idea, however, we would wish to know whether it will be successful. The implicit function theorem, which provides sufficient conditions under which this idea will work for at least some neighborhood of our point of interest, is the answer to that last question. In terms of the current example, let \( x_i \) denote the quantity of good \( i \). Given a point of interest \((x_1, x_2)\), provided the theorem’s conditions are satisfied, there will exist a small enough neighborhood of \( x_1 \) for us to picture good 2 as a function of good 1. Further, under additional conditions, the theorem also tells us that this function is continuously differentiable in the selected neighborhood and that its partial derivative can be expressed as a function of our original function’s partial derivatives!
Theorem: (Implicit Function Theorem) Let $X$ be a nonempty subset of $\mathbb{R}^n$ and $f : X \to \mathbb{R}$. Suppose also that $f$ belongs to $C^1(A)$, where $A$ is a neighborhood of $\bar{x}$ in $X$ and that for some $i$ in $\{1, 2, \ldots, n\}$

$$\frac{\partial f(\bar{x})}{\partial x_i} \neq 0$$

Then there exists a function $\phi_i(x_{-i})$ defined on a neighborhood $B$ of $\bar{x}_{-i}$ such that

$$\phi_i(\bar{x}_{-i}) = \bar{x}_i$$

Furthermore, if $x_{-i}$ is in $B$, then $(x_1, \ldots, x_{i-1}, \phi_i(x_{-i}), x_{i+1}, \ldots, x_n)$ is in $A$ and

$$f(x_1, \ldots, x_{i-1}, \phi_i(x_{-i}), x_{i+1}, \ldots, x_n) = f(\bar{x})$$

Finally, the function $\phi_i$ is differentiable at $\bar{x}_{-i}$ and its derivative with respect to $x_j$, $j \neq i$, at $\bar{x}_{-i}$ is

$$\frac{\partial \phi_i(\bar{x}_{-i})}{\partial x_j} = -\frac{\frac{\partial f(x)}{\partial x_j}}{\frac{\partial f(x)}{\partial x_i}}$$

Proof: The interested reader can find an illustrative proof of the two dimensional case on page 207 of De la Fuente [1]. Higher dimensional cases are more delicate (but even more important...). The interested reader will find it in section M.E of Mas Colell, Whinston, and Green [5].

Remark: It is important to note that the obtained derivative is continuous, as it is a composition of continuous functions. Hence, $\phi_i$ belongs to $C^1(B)$, as we wished!

Again, let us go back to the burger-soda example. The conditions of the implicit function theorem tell us where we run into troubles. If any of the $x_i$ is equal to 2, then the partial derivative with respect to $x_i$ vanishes (cf. Chapter 3, Exercise 4), as depicted in the left part of figure 6, and $x_i$ cannot be expressed as a function of $x_j$, $j \neq i$. Geometrically, as the right part of figure 6 shows, with $x_j$’s in the abscissa and $x_i$’s in the ordinate, although the tangent is well defined and unique, it is parallel to the ordinate axis, and no matter how small a neighborhood of $\bar{x}_i$ we pick, we can always find two distinct $x_i$’s which share a
3 Unconstrained Optimization

Let us start with a simple unconstrained maximization problem. To test your understanding of the material, you may want to think what would happen for a minimization problem. This remark will also apply to all the following sections.

\[(\mathcal{P}) \quad \underset{x \in \text{dom}(f)}{\text{maximize}} \quad f(x)\]

where \(\text{dom}(f) \subset \mathbb{R}^n\) and \(f\) is real-valued and continuously differentiable.

3.1 First Order Necessary Conditions

As mentioned in the last chapter, when looking for optima, one is neither able nor willing to investigate every point of the domain of \(f\). The first order necessary conditions are a set of conditions, available for continuously differentiable functions, that single out candidate maximizers or candidate minimizers in their domain. They are local in nature, in the sense that they will pick a point and look at what happens around it, and they rely extensively on the geometrical insights associated with derivatives\(^9\). Namely, consider the case of a local maximizer \(\bar{x}\), and assume the objective function is low dimensional, say, univariate. Then, two situations may arise:

(i) The function is constant in a neighborhood of \(\bar{x}\). In this case, the derivative of \(f\) should be equal to zero at \(\bar{x}\) (as well as in its neighborhood).

\(^9\)Those geometrical insights which we sought to preserve when generalizing the notion of a derivative in chapter 3.
(ii) The function is not constant in a neighborhood of $\bar{x}$. Instead, $\bar{x}$ is at the top of a “hill”.

In this case, the tangent at $\bar{x}$ must be horizontal, i.e., again, the derivative of $f$ should be equal to zero at $\bar{x}$.

![Figure 7: When $f$ is in $C^1(X)$, the derivative vanishes at local extrema!](image)

Analytically – and thinking now again about any finite dimensional domain – given the continuous differentiability of $f$, we can use a first order Taylor approximation, valid in a small enough neighborhood of $\bar{x}$:

$$f(\bar{x} + h) \approx f(\bar{x}) + \nabla f(\bar{x}) \cdot h$$

Assume $\bar{x}$ is a maximizer. Then, $\nabla f(\bar{x})$ can be neither positive nor negative, for, if it were, we would find an appropriate $h$ – namely, an $h$ with $\|h\|$ small enough for the approximation to be valid and such that $\nabla f(\bar{x}) \cdot h$ is positive – such that $f(\bar{x} + h) > f(\bar{x})$. A contradiction.

Hence, the analytical reasoning guarantees the extensibility of our low dimensional geometric intuition to higher dimensional cases. A symmetric reasoning can be applied for the analysis of minima. Therefore, we have the following result:

**Theorem: (Unconstrained Extremum – First Order Necessary Condition)**

Consider $(P)$. Let $\bar{x}$ be an element in the interior of $\text{dom}(f)$. If $\bar{x}$ is a local extremum of $f$, then:

$$\nabla f(\bar{x}) = 0$$

Every global extremum being a local extremum, the First Order Necessary Conditions also apply to global extremizers. Elements of the domain which fulfill this requirement are important enough to get their own label:
Definition: (Critical Point (a.k.a Stationary Point))
Let $X$ be a nonempty subset of $\mathbb{R}^n$, $f : X \to \mathbb{R}$, and $\bar{x}$ be an element of $X$ at which $f$ is differentiable. If $\nabla f(\bar{x}) = 0$, then $\bar{x}$ is called a critical point, or stationary point of $f$.

Elements of the domain where the gradient of $f$ vanishes are critical in the sense that they represent reasonable extremizer candidates. Yet, these necessary conditions are not sufficient, i.e., they need not be actual extremizers. The following figures show an instance of a critical point that is not a maximizer as well as a case where continuous differentiability of the investigated function is violated.

![Figure 8: A null derivative is not sufficient! Nor necessary if $f$ isn’t in $C^1(X)!$](image)

### 3.2 Second Order Necessary Conditions

This section is concerned with second order necessary conditions, i.e., conditions additional to the first order conditions that every extrema must satisfy. Again, these are not sufficient conditions, so beware! Let’s proceed directly with analytical investigations. Let $\bar{x}$ be a local maximizer, $B_\varepsilon(\bar{x})$ denote a neighborhood of $\bar{x}$ where $f(\bar{x}) \geq f(x)$ for all $x \in B_\varepsilon(\bar{x})$. Then, provided $f$ is in $C^3(B_\varepsilon(\bar{x}))$, the second order Taylor approximation writes as follows:

$$f(\bar{x} + h) \approx f(\bar{x}) + \nabla f(\bar{x}) \cdot h + \frac{1}{2} h' \cdot H_f(\bar{x}) h$$

which, given our learning from the last section, can be simplified:

$$f(\bar{x} + h) \approx f(\bar{x}) + \frac{1}{2} h' \cdot H_f(\bar{x}) h$$

Hence, $f(\bar{x}) \geq f(x)$ for all $x \in B_\varepsilon(\bar{x})$ implies that for all $h$ such that $\bar{x} + h \in B_\varepsilon(\bar{x})$,

$$\frac{1}{2} h' \cdot H_f(\bar{x}) h \leq 0$$
And given that only the angle between vectors matter for the sign of an inner product, we have just shown that, if \( \bar{x} \) is a maximizer of \( f \) and \( f \) is in \( C^3(B_\varepsilon(\bar{x})) \), then

\[
\lambda' \cdot H_f(\bar{x}) \lambda \leq 0 \quad \text{for all } \lambda \in \mathbb{R}^n
\]

**Theorem: (Unconstrained Maximum - Second Order Necessary Condition)**

Consider \( (P) \). Let \( \bar{x} \) be an element in the interior of \( \text{dom}(f) \) and \( B_\varepsilon(\bar{x}) \) an open \( \varepsilon \)-ball around \( \bar{x} \). Assume \( f \) is in \( C^2(B_\varepsilon(\bar{x})) \). If \( \bar{x} \) is an extremum of \( f \), then:

\[
\forall \lambda \in \mathbb{R}^n \quad \lambda' \cdot H_f(\bar{x}) \lambda \leq 0
\]

i.e. the Hessian of \( f \) at \( \bar{x} \) is negative semidefinite.

A symmetric statement exists for unconstrained minima. An obvious failure of these conditions is when the Hessian matrix of \( f \) at \( \bar{x} \) is indefinite. Given our smoothness assumptions on \( f \), its Hessian is symmetric and we can relate indefiniteness to statements about its eigenvalues. Namely, an indefinite Hessian has both positive and negative eigenvalues. Along the directions spanned by the eigenvectors associated to negative eigenvalues, \( f(\bar{x}) \) looks like a local maximum. Along the directions spanned by the eigenvectors associated to positive eigenvalues, \( f(\bar{x}) \) looks like a local minimum. To help you picture this out, consider the function \( f(x, y) = x^2 - y^2 \). Its gradient vanishes at \( 0 \) and its Hessian at \( 0 \) is

\[
H_f(0, 0) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}
\]

One can plot \( h' \cdot H_f(0, 0)h \) in a neighborhood of \( 0 \).

![Figure 9: \( f(x, y) = x^2 - y^2 \), \( 0 \) is a Saddle Point](image)

When a matrix is diagonal, each element on the diagonal is an eigenvalue and each column vector is collinear to the eigenvector of the matrix associated to the eigenvalue in that column. Looking at Figure 9 Along the dimension spanned by \( (2, 0)' \), i.e. the \( x \)-axis, \( f(0, 0) \)

---

10Remember this term, it will be used a lot. It is an informal way to say that the investigated function is sufficiently many times continuously differentiable.
looks like a minimum, while along the dimension spanned by \((0, -2)\)', i.e., the y-xis, it looks like a maximum:

Elements of the domain where the gradient vanishes but the Hessian is indefinite are called saddle points. Given the picture, I suppose you can make sense of that name. In figure 8, on the left, the plotted function is \(f(x) = x^3\). Polynomial functions belong to \(C^\infty(X)\), i.e., they can be differentiated at will. The failure of the critical point \(\bar{x} = 0\) to qualify as an extremum actually illustrate why this second order condition is not sufficient. Indeed, \(f''(x)\) can be both positive and negative in a neighborhood of 0, even though \(f''(0) = 0\) does not formally fail the semi-definiteness test. The interested reader can check the first pages of Simon and Blume [7]'s Chapter 16, which introduces quadratic forms, an insightful topic in these matters.

### 3.3 Sufficient Conditions

The second order necessary conditions we have just checked happen to be almost sufficient. Indeed, strengthening the requirement to negative definiteness, rather than negative semidefiniteness, suffices to rule out cases such as \(f(x) = x^3\). The analytical proof is rather easy, its idea is displayed below. Geometrically, one wants \(f(\bar{x})\) to look like a maximum in all of the dimensions spanned by the eigenvectors of the Hessian. Beware, when looking for sufficient conditions, one must distinguish between the case of a local extremum and that of a global one. Because a global extremum is also a local extremum, but not the other way around, sufficient conditions for global extrema are “more demanding” than those for local ones. This is not straightforwardly reflected in the following results, as for a global maximizer we require only negative semidefiniteness but on the whole domain, rather than in an arbitrary neighborhood of \(\bar{x}\).

**Theorem:** (Sufficient Conditions for a Local Unconstrained Maximum)

Consider \((\mathcal{P})\). Let \(\bar{x}\) be an element in the interior of \(\text{dom}(f)\) and \(B_\varepsilon(\bar{x})\) an open \(\varepsilon\)-ball around \(\bar{x}\). Assume \(f\) is in \(C^3(B_\varepsilon(\bar{x}))\). If \(\nabla f(\bar{x}) = 0\) and \(H_f(\bar{x})\) is negative definite, then \(\bar{x}\) is a local maximizer of \(f\).

**Proof:** (idea) Using the exact Taylor expansion (i.e., that which includes the error term, one can show that there exists a \(h\) small enough for the error term to be smaller than \(\frac{1}{2} h' \cdot H_f(x) h\).

\(\square\)

An undeterminate case is thus that of functions with a negative semidefinite but not negative definite Hessian at \(\bar{x}\). As you may have guessed, in such a case, investigations of higher order are necessary. The exception is when \(f\) is globally concave, in which case we have the
Theorem: (Sufficient Conditions for a Global Unconstrained Maximum)
Consider (P). Let $\bar{x}$ be an element in the interior of $\text{dom}(f)$ and $f$ be a concave function in $C^2(\text{dom}(f))$. If $\nabla_f(\bar{x}) = 0$, then $\bar{x}$ is a global maximizer of $f$.

Proof:

We have seen in the previous chapter that a multivariate concave real-valued function $f$ has a convex hypograph, and therefore, using the supporting hyperplane theorem, can be characterized by the following equation:

$$\forall x \in \text{dom}(f) \quad f(x) \leq f(\bar{x}) + \nabla_f(\bar{x}) \cdot (x - \bar{x})$$

Given $\nabla_f(\bar{x}) = 0$, the desired result obtains.

\[
\]

4 Constrained Optimization I: Convex Optimization

I start with convex maximization problems, i.e., problems of the form;

\[
(P_c) \quad \begin{array}{ll}
\text{maximize} & f(x) \\
\text{subject to} & h_i(x) \leq 0, \ i = 1, \ldots, k.
\end{array}
\]

Where $f$ is a concave\footnote{Yes, concave! You should become clear with the terminology “convex optimization” soon!} function, all $h_i$ are convex functions and, as above, $\text{dom}(f) \subset \text{dom}(h_i)$ for all $i$. We further assume that $f$ is in $C^2(\text{dom}(f))$ and all $h_i$ are in $C^2(\text{dom}(h_i))$, respectively.

Remark: Although solved more recently, such problems conceptually simpler and also represent the majority of problems that you will encounter during your studies here. For instance, least square problems and linear programming problem (i.e. problems with a linear objective function and linear inequality constraints) are subsumed in the class of convex optimization problems. For the sake of completeness, in the next section, I present non-convex problems, including the case of equality constraints, where insights are unfortunately harder to get.

AGAIN, THIS IS THE CANONICAL SHAPE OF OUR CONVEX MAXIMIZATION PROBLEMS. IT IS IMPORTANT TO STICK TO IT!
4.1 Geometrical Interpretation of the Constraints - Convex Optimization

It is important to understand that constraints stand for restrictions on the domain of our objective function, and, therefore, define a subset $X$ of the domain over which we are imposed to restrict our search for a maximizer. Such a set is called the feasible set and its elements are called the feasible points. Let us take an easy case to illustrate the point. Assume we are facing simple linear inequality constraints.

$$\text{maximize } f(x) \quad x \in \text{dom}(f)$$

subject to $a_i'x - b_i \leq 0, \ i = 1, \ldots, k.$

Where, $\text{dom}(f)$ is a subset of $\mathbb{R}^n$ and, for all $i$, $a_i$ are elements of $\mathbb{R}^n$ and $b_i$ are elements of $\mathbb{R}$. Such a problem could be rewritten as follows:

$$\text{maximize } f(x) \quad x \in X$$

Where $X := \{x | x \in \text{dom}(f), a_i'x - b_i \leq 0, \ i = 1, \ldots, k\}$ and, again, for all $i$, $a_i$ are elements of $\mathbb{R}^n$ and $b_i$ are elements of $\mathbb{R}$. In the present case, the region defined is an intersection of halfspaces\(^{12}\) and is thus a polyhedron. Figure 10 depicts a case where $k = 5$ and $n = 2$ (note: we depict level sets, we lie in the same space as that of the domain!).

Figure 10: Linear Inequality Constraints Define a Polyhedron of Feasible Points

Remark: This example illustrates a case where $k > n$ and yet $X$ is non empty. You have to keep in mind that this is a major difference between the problems with equality constraints and that with inequality constraints. If we had considered above the hyperplanes and not the halfspaces (i.e. equality and not inequality constrains), then we would not have been able to find a point that satisfies all constraints at once (Make sure you see why in the figure below!).

\(^{12}\)Note that each of the $k$ constraints defines a halfspace!
For this reason the general theorems of the equality constraint section are stated for the case where \( m < n \) (recall, our canonical problem has \( m \) equality constraints), in which we can be sure that \( X \) is neither empty nor discrete (case where \( m = n \)).

Given the analysis in the last section, a natural question to ask is the following: in what way is our problem affected by the switch from the full domain to a restricted feasible set? A crucial difference is that in a constrained problem, if \( \bar{x} \) is a maximizer and \( f \) is all the desired smoothness properties, then it need not be true that the gradient of \( f \) evaluated at \( \bar{x} \) vanishes! Figure 11 depicts the case of \( f : \mathbb{R} \to \mathbb{R} \), which associates \( 2x \) to every \( x \) in \( \mathbb{R} \). When \( x \) is constrained to fulfill \( x \geq -1 \) and \( x \leq 1 \), \( f \) has a unique maximum and a unique minimum while it has neither a maximum nor a minimum in the unconstrained case. Further, at both extrema, the derivative does not vanish!

![Figure 11: Boundary Solution in a Constrained Optimization Problem](image)

An important consequence of this geographical interpretation is that we now understand that we must separate our investigations in two parts: investigate elements on the boundary of our constraint set, where the necessary conditions for the unconstrained case do not apply, and investigate elements on the interior of the constraint set, where we can safely use our previously seen techniques. Checking boundaries cannot and must not be neglected.

Before we move on, let me clarify the terminology of convex optimization problem then. First, mathematicians’ benchmark case is that of a minimization problem. The above problem is equivalent to the following minimization problem:

\[
\begin{align*}
\text{minimize} & \quad -f(x) \\
\text{subject to} & \quad h_i(x) \leq 0, \ i = 1, \ldots, k.
\end{align*}
\]

In such a problem, all functions would be required to be convex, and you would probably be fine with the terminology. Yet, I do not think that this is the reason why such problems are called convex. The reasons are that (i) convex constraints define a convex region in the space.
associated to $f$ domain (lower-level sets of a convex function are convex, and the intersection of convex sets is convex!) and (ii) a convex objective function has a convex lower-level set, as we have seen in last chapter. For a maximization problem, we are interested in the upper-counter set of the objective function, rather than the lower one (you’ll see why later) and, when the objective function is concave, its upper-level set is convex!

### 4.2 The Karush-Kuhn-Tucker Necessary Conditions

**Fact: (Interior Local Maxima)**

Let $X \in \mathbb{R}^n$ denote the feasible set associated to $(P_c)$ and $\bar{x}$ be an element in $\text{Int}(X)$. Then, $\bar{x}$ is a local maximizer of the constrained problem if and only if it is a local maximizer of the unconstrained problem.

This result stems from the fact that our unconstrained analysis relied on the study of local variations in the domain (remember, we used Taylor approximations, i.e., locally valid approximations of the objective function). For a point in the interior of $X$, one can always find a neighborhood small enough for it to be contained in $X$, and the results from our previous section apply. A question that arise though, is whether a local extremum has become a global one once the constraints have been imposed, or whether it is supplanted by (i) some other interior local extremum or (ii) a new extremum generated by the introduction of a boundary on the maximization set. Case (i) is easy to verify and does not change from the unconstrained case, as it amounts to compare the values of the two candidates for “globalism”. Case (ii) is a bit more tricky, let’s turn to it.

For simplicity, let us first consider a two-dimensional convex maximization problem with a single inequality constraint, i.e., a problem of the following type:

$$\begin{align*}
\text{maximize} & \quad f(x) \\
\text{subject to} & \quad h(x) \leq 0
\end{align*}$$

Where $f$ is concave and $h$ is convex, and both have their domain in $\mathbb{R}^2$. Let $X$ denote the feasible set, i.e., $X = \{x | x \in \text{dom}(f), h(x) \leq 0\}$. Because we are about to investigate necessary conditions for a solution on the boundary of $X$, we first assume the existence of such an maximizer, denoted $\bar{x}$ in the sequel, and then consider what properties $\bar{x}$ will necessarily possess. A first necessary condition is that the level set of $f$ at $\bar{x}$ does not intersect the interior of $X$. An easy way for you to picture this fact out is to simply imagine a budget set and to consider a bundle inside the budget set. We know that if our agent has convex preferences, then the solution will lie on the budget line, i.e. the boundary of the delimited set. In this situation, if the indifference curve going through a bundle $x$ located on the

---

13 In a nutshell, one speaks of convex preferences when the upper-level sets of the utility function are
budget line is intersecting the interior of our budget set, then bundle $x$ cannot be optimal, as it is possible to find a point still contained in the budget set lying on a higher indifference curve.

Figure 12: The Level Set of a Boundary Solution Cannot Intersect Int($X$)

Thus, we need only focus on points which belong to a level set tangent to the boundary of $X$.

BEWARE, IMPORTANT INSIGHT ON ITS WAY!

Again, $\bar{x}$ is our assumed maximizer. Denote $L^f_{f(\bar{x})} := \{ x | x \in \mathbb{R}^2, f(x) = f(\bar{x}) \}$ the level set of $f$ going through $\bar{x}$, $L^{++}_f := \{ x | x \in \mathbb{R}^2, f(x) > f(\bar{x}) \}$ the strict $f(\bar{x})$-upper level set, $L^h_0 := \{ x | x \in \mathbb{R}^2, h(x) = 0 \}$ the 0-level set of $h$, and $L^{--}_h := \{ x | x \in \mathbb{R}^2, h(x) < 0 \}$ the strict 0-lower level set of $h$. With these notations, note that the feasible set, $X$ coincides with the 0-lower level set of $h$, i.e., $X = L^h_0 \cup L^{--}_h$. Figure 13 depicts the situation. Now, consider the following: given the concavity of $f$ and the convexity of the $h_i$, we know that $L^{++}_f$ and the lower-level set of $h$ are convex. Given that both such sets have non-empty interiors$^{14}$, the supporting hyperplane theorem (Chapter 2, Appendix 1) tells us that $L^f_{f(\bar{x})}$ and $L^h_0$ – the boundaries of those sets – are smooth curves which can be approximated by their tangent at $\bar{x}$. From the section on level sets, we know that these tangents are characterized by the following hyperplanes:

$$\{ x \in \mathbb{R}^2 : \nabla h(\bar{x}) \cdot x = \nabla h(\bar{x}) \cdot \bar{x} \}$$
$$\{ x \in \mathbb{R}^2 : \nabla f(\bar{x}) \cdot x = \nabla f(\bar{x}) \cdot \bar{x} \}$$

$\nabla h(\bar{x})$ and $\nabla f(\bar{x})$, respectively, are orthogonal to the associated hyperplane. Furthermore, as explained above, if $\bar{x}$ is a solution, then $L^{++}_f \cap X = \emptyset$. An immediate consequence is that these two tangents must coincide (after having had a look at the Figure 13, try out convex, i.e. when the utility function is quasi-concave (see last chapter). In this section, we think of a concave objective function, therefore, thinking of a consumer problem, we are indeed in the case of convex preferences.

$^{14}$Indeed, assume it were not the case for $h$, and then we would simply be facing an equality constraint, that is not the kind of problem we claim to deal with here. The nonemptyness of $L^{++}_f$ is guarantied by the fact that the constraint is binding.
some plots with tangents that do not coincide, you’ll see what happens!)! Let $T$ denote that common tangent. Because $\nabla h(\bar{x})$ and $\nabla f(\bar{x})$ are both orthogonal to $T$, they must be colinear, i.e., there exists a $\lambda$ in $\mathbb{R}$ such that $\nabla f(\bar{x}) = \lambda \nabla h(\bar{x})$. Finally, because $X$ is the set of points such that $h(x) \leq 0$, $\nabla h(\bar{x})$ must be pointing outwards. Also $\nabla f(\bar{x})$ must be pointing inwards, i.e., in the direction of values larger than $\nabla f(\bar{x}) \cdot \bar{x}$. As a consequence, when $h(x) \leq 0$ is binding, $\lambda$ must be non-negative! (See chapter 2, Figure 6).

Figure 13: Fundamental Diagram of Constrained Optimization (Inequality Constraints)

Remark: The case where the constraint is not binding\textsuperscript{15}, which has been dealt with in the ‘Fact’ exposed at the introduction of this section, is easily expressed by the condition $\nabla f(\bar{x}) = \lambda \nabla h(\bar{x})$ too. Indeed, one just has to additionally impose the following condition: one of the two holds true: (i) the constraint bind or (ii) $\lambda = 0$. This implies that, when the constraint is slack, the necessary condition that the gradient of $f$ being null at $\bar{x}$ still holds. It is know as the complementary slackness condition and can be neatly mathematically expressed by the following condition: $\lambda h(\bar{x}) = 0$.

Figure 14: For a slack constraint, keep the same condition but require that $\lambda = 0$!

\textsuperscript{15}You should get used to the following vocable: we say that an inequality constraint is binding if it is satisfied with equality, that it is not binding, or slack, otherwise.
Now, let us have a look at the situation with two constraints, \( h_1(x) \leq 0 \) and \( h_2(x) \leq 0 \), and see if a pattern emerges. Everything else equal and provided the feasible set has a nonempty interior, three types of situations are likely to arise. Type one is the case where both of the constraints are binding. In such a situation, the maximizer necessary lies at the point where the 0-level set of \( h_1 \), \( \mathcal{L}_{h_1}^0 := \{ x \in \mathbb{R}^n, h_1(x) = 0 \} \), and the 0-level set of \( h_2 \), \( \mathcal{L}_{h_2}^0 := \{ x \in \mathbb{R}^n, h_2(x) = 0 \} \), intersect.

![Figure 15: Two constraints, case 1](image)

This time, however, the tangents do not coincide. Yet, we can rely on a similar intuition. We have seen above that if the level set through a point \( x \) on the boundary (of the feasible set) has a non-empty intersection with the interior of the feasible set, then this point \( x \) cannot be optimal. Now, imagine \( \bar{x} \) is an optimal point and \( \nabla f(\bar{x}) \) lies outside the cone delimited by \( \nabla h_1(\bar{x}) \) and \( \nabla h_2(\bar{x}) \). It means that the tangent along the level set of \( f \) is steeper than that along one of the boundaries of our feasible set. As a consequence, it must be that our level set of \( f \) intersect the interior of \( X \). But that would just contradict the optimality of \( \bar{x} \)!

Therefore, at any optimal point, the gradient of \( f \) at \( \bar{x} \) takes the form of a positive linear combination of the two gradients: \( \nabla f(\bar{x}) = \lambda_1 \nabla h_1(\bar{x}) + \lambda_2 \nabla h_2(\bar{x}) \), with \( \lambda_1 \) and \( \lambda_2 \) non-negative.

Type two is the case where only one of the two constraints, say \( h_1 \), is binding. In that case, we naturally ask that \( \lambda_2 \) be equal to zero, and the characterization of \( \lambda_1 \) is very similar to that of \( \lambda \) in the one constraint case. (See Figure 16)

Type three is the situation with an interior solution. In such a situation, none of the constraints are binding, and we unsurprisingly require that both \( \lambda_1 \) and \( \lambda_2 \) be zero! (See Figure 17)
We have just obtained geometrical insights for the following important result:

**Theorem**: *(Karush-Kuhn-Tucker Necessary Conditions)*

Consider \((P_c)\). Let \(\tilde{I} := \{i \in \{1, \ldots, k\} : \ h_i \text{ is not linear}\}\). Assume the Slater condition is satisfied, i.e.:

\[
(S) \quad \exists \tilde{x} \in X, h_i(\tilde{x}) < 0 \ \forall i \in \tilde{I}
\]

Then, a necessary condition for an element \(\tilde{x}\) of \(X\) to be a solution of \((P_c)\) is the KKT condition:

\[
\exists! \tilde{\lambda} := (\tilde{\lambda}_1, \ldots, \tilde{\lambda}_k) \in \mathbb{R}_+^k
\]

such that:

\[
\begin{align*}
\nabla f(\tilde{x}) &= \sum_{i=1}^{k} \tilde{\lambda}_i \nabla h_i(\tilde{x}) \quad (\text{KKT 1}) \\
\tilde{\lambda}_i h_i(\tilde{x}) &= 0, \quad i = 1, \ldots, k \quad (\text{KKT 2})
\end{align*}
\]

**Remark 1**: The geometrical discussion exposed above is strongly inspired by Rochet [6] and captures the idea of the proof. For those willing to see the analytical version (which is shorter
but not as insightful!), a good reference is Kreps [3]'s Appendix 5. In the analytical version, the Slater condition is replaced by a more demanding “constraint qualification”. This is due to the fact that they do not focus on convex problems. I discuss this point in the next section.

(KKT 1) follows from the above intuition and simply states that the gradient of the objective function at $\bar{x}$ is a linear combination with non-negative coefficients (a.k.a. conic combination) of the gradients of the constraints at $\bar{x}$. (KKT 2) has also been intuited above and is known as the complementary slackness condition: for all $i$, either $\lambda_i = 0$ or the $i$th constraint is satisfied with equality. Put differently, if our candidate solution is not “effectively” constrained by inequality $h_i$, then the gradient of $h_i$ should not enter the linear combination of gradients forming the gradient of $f$, which we have seen above too.

Remark 2: Remember that our reasoning has relied on the existence of a nonempty interior of the feasible set, $X$. The Slater condition precisely makes sure that the interior of $X$ is nonempty.

4.3 Sufficient Conditions

Actually, provided the problem faced is convex, the conditions we have just detailed happen to be sufficient too. The necessity was using the supporting hyperplane theorem. The sufficiency will simply use our second version of the separating hyperplane theorem (Again, see Chapter 2, Appendix 1).

**Theorem:** (Sufficient Conditions for an Convex Maximization Problem)
Consider $(P_c)$. If the Slater condition holds and $\bar{x}$ fulfills the KKT conditions, then $\bar{x}$ is a global maximizer of $f$.

**Proof:** (Idea)

We have already argued above that the upper-level set of $f$ and the feasible set $X$ are convex sets with non empty interior (provided the Slater condition hold). Hence we can apply the separating hyperplane theorem, which guarantees the existence of a hyperplane separating $L_f^+\bar{x}$ and $X$.

This hyperplane separates the space in two halfspaces. More precisely, all feasible points (i.e. all $x$ in $X$) are such that $\nabla f(\bar{x}) \cdot x \leq \nabla f(\bar{x}) \cdot \bar{x}$ which implies $\nabla f(\bar{x}) \cdot (x - \bar{x}) \leq 0$. And as one of the characterizations of concavity states that any tangent plane lies weakly above the graph of $f$, we have that:

$$\forall x \in X \ f(x) \leq f(\bar{x}) + \nabla f(\bar{x}) \cdot (x - \bar{x}) \leq f(\bar{x})$$
Remark: To summarize, solving convex problem is, by now, a well known technology. That means two things: (i) you should exercise to master this KKT result, and (ii) once the technique is mastered, the difficult task is not solving the problem, but, rather, it is finding out whether or not the problem you face can be expressed. Point (ii) will, I wish you, come with time. We turn to point (i) soon.

4.4 Insights From The Associated Unconstrained Problem

Before convex analysis was developed, mathematicians were using another approach to solve constrained problems. The idea was, as for so many problems in mathematics, to attempt to go back to the type of problem one knows how to solve. In our case, unconstrained problems are the kind of problems we find easy to deal with: one just has to take first and second order derivatives and then check necessity and sufficiency. Looking for an unconstrained problem, the solution of which would coincide with that of the constrained problem, Lagrange came up with the following multivariate real-valued function:

$$
\mathcal{L}(x, \lambda) := f(x) - \sum_{i=1}^{k} \lambda_i h_i(x)
$$

Because of this function, the $\lambda_i$'s are usually referred to as the Lagrange multipliers of the problem. Although Lagrange initially developed this technique for problems with equality constraints, there are two benefits of presenting this technique: (i) it provides a nice insight about the meaning of our $\lambda_i$'s, and (ii) the approach is fairly general and is therefore still used today for solving non-convex problems, such as problems with non-linear equality constraints. I postpone the discussion on such problems to the next section, but wish to present the interpretation of the Lagrangian multipliers now, as it will matter much in your studies.

Assume you were to maximize the above Lagrangian (with $x$ as the control variable) and not abide by constraint $h_i$, i.e., assume you were to pick a $\tilde{x}$ such that $h_i(\tilde{x}) > 0$. Then, looking at the Lagrangian, and given that your multipliers are positive, you can note that the second part of the Lagrange function operates as a kind of “punishment device”: you pay a fee $\lambda_i h_i(\tilde{x}) > 0$ for selecting a $\tilde{x}$ that lies outside of the feasible set. And because rewards should not be awarded for going inside the constraint set, the complementary slackness conditions kick in. Further, assume you are ready to pay that punishment, it must be because it is worth it, i.e., it must be that the value of the Lagrangian increases when doing so. But we don’t want you to go outside the constraint set, so the $\lambda_i$’s are chosen precisely so as to make you indifferent between staying on the boundary of the constraint set and moving outside of it. In other words, the $\lambda_i$’s indicate the value by which $f(\tilde{x})$ would increase if one was to slightly relax the constraint. Considering an economic example, if $h_i$ is, say, a government budget constraint (with $x_i$ in, say, million Euros), then your $\lambda_i$ stands for the shadow price of raising more money: the consumer would be ready to pay $\lambda_i$ for a million
Euros more in his pocket.

Note that, taking $x$ as the control variable maximizing the Lagrangian, the first order conditions exactly yield (KKT 1):

$$\nabla f(\bar{x}) - \sum_{i=1}^{k} \bar{\lambda}_i \nabla h_i(\bar{x}) = 0 \quad \text{(KKT 1)}$$

Some people find easier to remember the Lagrangian rather than the (KKT 1) condition. Hence, when proceeding, they usually build it and take the first order conditions with respect to $x$. I do not advise to proceed in this way, as I think it may lead to mistakes. The Lagrangian was developed for equality constraints where the sign of the multipliers do not matter and remains best suited for such problems. In any case, proceed in the way you find the most intuitive, but remember, go back to canonical shape, and think about this penalty story when you lie outside the feasible set, this will help avoiding mistakes.

4.5 In Practice

Practice with the KKT conditions will require a bit of experience. But before you start with exercises, it is useful to know some hints about the way to proceed. First, put your problem in the canonical shape. Second, make sure you face a convex problem. Third, make sure the Slater condition holds. If the problem is non-convex, go to the next section. If the problem is convex but the Slater condition fails, then the KKT theorem does not apply, good luck! We know how the solution to our problem is characterized. Namely, by (KKT 1) and (KKT 2), i.e.

$$\nabla f(\bar{x}) = \sum_{i=1}^{k} \bar{\lambda}_i \nabla h_i(\bar{x}) \quad \text{(KKT 1)}$$

$$\bar{\lambda}_i h_i(\bar{x}) = 0 \quad i = 1, \ldots, k \quad \text{(KKT 2)}$$

Where the $\bar{\lambda}_i$’s are non-negative. One first tries to deal with the complementary slackness condition (KKT 2). At this point, the experience kicks in: you try to guess which constraints will be binding, which will be slack, and after guessing, you show it formally:\footnote{If you’re really sure and time is super constrained, then at least state in one sentence what makes you so sure!} For all the slack constraints you write down that the associated $\bar{\lambda}_i$ is zero, and forget about them.

Then, the (KKT 1) condition must be considered. Simply write it down. (At this stage, folks who prefer to set up the Lagrangian do it. Just do what’s most comfortable to you!) Given the conditions, using a bit of algebra should enable you to pin down the element(s) of the domain that is a (are) maximizer(s).
In Practice - A Summary (Implementing KKT’s Theorem)

(i) Put your problem in the canonical shape ($P_c$)

(ii) Verify that your problem is convex!!

(iii) Verify the Slater condition is satisfied.

(iv)* Try to guess which constraints will not be binding. After guessing, show it formally.
Try to guess which constraints will be binding. After guessing, show it formally.

(v) For those constraints which you have shown not to be binding, write down the fact that the associated multiplier is equal to 0.

(vi) Build the Lagrangian $L(x, \lambda)$ (no need to plug in the slack constraints!) and take the FOCs w.r.t. $x$, or simply write down the (KKT 1) condition.

(vii) Small algebraic manipulations should be sufficient to characterize the maximizer(s).

* What if one has issues in step (iv)? I.e. what if one can’t show whether some constraints are binding or slack? Then, unfortunately, a laborious case by case investigation of each possibility is required. (Think, e.g. when you were asked to check the “corner” conditions in a consumer problem!)

5 Constrained Optimization II: Non-Convex Optimization

Without convexity, except for two important cases I am about to discuss, the Lagrange technique is the only one we have at hand. The issue with this technique is that a proper understanding of it requires a proper understanding of duality theory, which is far beyond the scope of this lecture. Fortunately, as I already mentioned earlier, most of the problems you will have to face are convex. Some of the problems you will face are non-convex, but may be turned into convex problems or simply solved by very similar techniques. I first expose you to these environments, where the Karush-Kuhn-Tucker conditions remain applicable. Then I move to more “upsetting” problems you may have to face: optimization problems with non-linear equality constraints.

17And also fairly demanding. The interested reader can find a good first approach in Luenberger. You will see there that convex analysis also relies on duality, but all of it is nicely hidden behind the hyperplane theorems.
5.1 Quasi-convex Constraints

All that has been said about convex problems also apply if the constraints are only quasi-convex, *provided they also are continuously differentiable*. The reason is that our reasoning relied on the convexity of the lower-level sets of our constraints, which is preserved by quasi-convexity, as seen in chapter 3. It also relied on the the smoothness of the boundary of the feasible set. This is guaranteed by the continuous differentiability.

5.2 Linear Equality Constraints

Linear equality constraints can be handled with our theorem too! In fact, one just has to split them into two parts. Namely, let \( g_i(x) = 0 \) denote a linear equality constraint. Denote \( g_{i,1}(x) := g_i(x) \) and \( g_{i,2}(x) := -g_i(x) \). You can then replacing the constraint \( g_i(x) = 0 \) by the two following constraints:

\[
g_{i,1}(x) \leq 0 \quad \text{and} \quad g_{i,2}(x) \leq 0
\]

Your problem is now a convex optimization problem. Further, you also know that both of these constraints must be binding! That will come in handy!

**NON-LINEAR EQUALITY CONSTRAINT CANNOT BE SPLIT INTO TWO!**

The reason is that, for non-linear constraint, splitting them makes you run into a conflict with the Slater condition. Indeed, let \( g_j(x) = 0 \) be a non-linear equality constraint. Denote \( g_{j,1}(x) := g_j(x) \) and \( g_{j,2}(x) := -g_j(x) \). Then, the Slater condition would require that there exists an \( \tilde{x} \) such that:

\[
g_{i,1}(\tilde{x}) = g_j(\tilde{x}) < 0 \quad \text{and} \quad g_{i,2}(\tilde{x}) = -g_j(\tilde{x}) < 0.
\]

A contradiction. Thus, we are safe with linear constraints only because they are not concerned by the Slater condition.

5.3 Non-Linear Equality Constraints - The Lagrange Conditions

In this section we have a look at the following kind of problem:

\[
\begin{align*}
(P_l) \quad \text{maximize} \quad & f(x) \\
\text{subject to} \quad & g_i(x) = 0, \ i = 1, \ldots, m.
\end{align*}
\]

Where \( f \) is differentiable, the \( g_i \) are continuously differentiable, and at least one \( g_i \) is non-linear.

*Remark*: if we want to make sure that our feasible set is neither empty nor discrete, we must impose that \( m \), the number of constraints, be strictly smaller than \( n \) the dimension of our
vectors in the domain of $f$.

If we consider again the geometrical interpretation via level sets, the problem exhibits a crucial difference with the case of inequality constraints. Namely, the feasible set is non-convex and has an empty interior. This implies that we cannot use our previous techniques. The idea, as argued above, is to get back to an unconstrained problem that has a solution set identical to the one of the constrained problem. A noticeable difference is that nothing restricts the sign of the multipliers any more, as the penalty should now be applied to any excursion outside of the 0-level set of the constraint, no matter the sense. In Figure 18, two examples are plotted.

![Figure 18: Fundamental Diagram of Constrained Optimization (Non-Linear Equality Constraints)](image)

Instead of relying on convex analysis, one can rely on the implicit function theorem, provided some smoothness conditions are satisfied. Thereby one guarantees that approximating our nonlinear constraints by linear constraints (their tangents) is feasible in a neighborhood of the maximizer. We have only seen the version of the implicit function theorem for real valued function, which is not sufficient for a case of multiple constraints, for they add up into a function that is not real-valued. Indeed, considering the set of $m$ constraints as a unique constraint $g : \mathbb{R}^n \to \mathbb{R}^m$ yields:

$$g(x) := \begin{pmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{pmatrix} = 0$$

As mentioned above, a generalized implicit function theorem exists and, rather than resorting to non-vanishing gradients, as in the bivariate case, it requires that $\det(J_g(\bar{x})) \neq 0$ (where
The Jacobian of \( g \) at \( \bar{x} \), i.e., that the set of gradients forms an independent family of vectors. Then, for optimization problems with non-linear equality constraints, the following result obtains:

**Theorem: (Lagrange’s First Order Necessary Condition)**

Consider \((P_l)\). If \( \bar{x} \) satisfies the Lagrange regularity condition, i.e.,

\[
(\mathcal{L}) \ \{\nabla g_1(\bar{x}), ..., \nabla g_m(\bar{x})\}
\]

are linearly independent vectors,

then a necessary condition for \( \bar{x} \) to be an extremum is the following:

\[
(\text{NC 1}) \ \exists \bar{\lambda} \in \mathbb{R}^m, \nabla f(\bar{x}) = \sum_{i=1}^{m} \bar{\lambda}_i \nabla g_i(\bar{x})
\]

**Proof:** See e.g. section M.K of Mas Colell, Whinston, and Green [5]. There, the Lagrange regularity condition is called the “constraint qualification”. The latter term is the most general of the two. Lagrange regularity condition is the specific “constraint qualification” associated to problems of type \( P_l \), in the same way that the Slater condition is the specific “constraint qualification” associated to convex problems. Mas-Colell, Whinston, and Green [5] provides further details about this constraint, which might be of interest to you.

**Remark:** Note that the Lagrange regularity condition implies that the \( \bar{\lambda}_i \)'s are unique! Note also that it is very demanding in practice, as it must be checked at every single candidate point \( \bar{x} \)!

Finally, to derive the appropriate sufficient conditions, one can simply build the Lagrangian and use the unconstrained local second order sufficient conditions presented in an earlier section. Beware though, for problems of type \( P_l \), the Lagrangian is a function of both \( x \) and \( \lambda \)! Hence, let:

\[
\mathcal{L}(x, \lambda) := f(x) - \sum_{i=1}^{m} \lambda_i g_i(x)
\]

The following result obtains:

**Theorem: (Sufficient Conditions for a Local Equality Constrained Maximum)**

Consider \( P_l \). Let \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) be twice differentiable at \( \bar{x} \in \mathbb{R}^n \) and, for all \( i = 1, ..., m \), let \( g_i : \mathbb{R}^n \rightarrow \mathbb{R} \) be twice differentiable at \( \bar{x} \in \mathbb{R}^n \). If there exists \( \bar{\lambda} \) in \( \mathbb{R}^m \) such that

\[
\nabla \mathcal{L}(\bar{x}, \bar{\lambda}) = 0 \quad \text{and} \quad H_{\mathcal{L}}(\bar{x}, \bar{\lambda}) \text{ is negative definite},
\]

then \( \bar{x} \) is a local maximizer of our constrained problem.

**Proof:** It is a second order sufficient condition for unconstrained problems. We have already proved it.
Note that the Hessian of the Lagrange function at \((\bar{x}, \bar{\lambda})\) writes as follows:

\[
\mathbf{H}_L(\bar{x}, \bar{\lambda}) = \begin{pmatrix}
\frac{\partial^2 L}{\partial \lambda^2}(\bar{x}, \bar{\lambda}) & \frac{\partial L}{\partial x \partial \lambda}(\bar{x}, \bar{\lambda}) \\
\frac{\partial L}{\partial \lambda}(\bar{x}, \bar{\lambda}) & \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})
\end{pmatrix} = \begin{pmatrix}
0_{m \times m} & -\mathbf{J}_g(\bar{x}) \\
-\mathbf{J}_g(\bar{x})' & \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})
\end{pmatrix}
\]

Where \(\mathbf{J}_g(\bar{x})\) denotes the Jacobian of the aggregate constraint function \(g\) at \(\bar{x}\). Checking the negative definiteness of this matrix happens to be equivalent to showing the negative definiteness of the following matrix (see exercises):

\[
\tilde{\mathbf{H}}_L(\bar{x}, \bar{\lambda}) := \begin{pmatrix}
0_{m \times m} & \mathbf{J}_g(\bar{x}) \\
\mathbf{J}_g(\bar{x})' & \frac{\partial L}{\partial x^2}(\bar{x}, \bar{\lambda})
\end{pmatrix}
\]

This matrix is called the bordered Hessian and it is the one people think of when speaking of second order conditions for a Lagrange approach. (The reason, I suppose, is that this is the Lagrangian’s second order matrix if you set up your Lagrangian adding the linear combination of the constraints – instead of subtracting them. When constraints are equality constraints, the sign of the lambda does not matter, and therefore, constructing the Lagrangian by adding the linear combination of the constraints also makes sense. In the case of a convex problem, however, given the canonical form we chose, the Lagrangian must be constructed by subtracting the linear combination of the constraints).

**THAT \(f\) IS CONCAVE IS **NOT** SUFFICIENT TO ENSURE THAT A POINT SATISFYING (NC 1) BE AN EXTREMUM! NOT EVEN A LOCAL ONE!**

That is because the concavity of \(f\) does not ensure that of the Lagrangian! Concavity of \(f\) together with convexity of the constraints, however, ensures it! Then global second order sufficiency conditions also apply!

*Remark: For the case which combines equality and inequality constraints, see Mas-Colell, Whinston, and Green [5] section M.K.*
The approach in these notes has been mostly geometrical. For those who prefer an analytic argumentation, there is an excellent and short book by Avinash K. Dixit [2] which I advise you to read.

References


