Mathematics for Economists: Constrained Optimization

Juuso Välimäki

Spring 2020

Contents

1 Analysis

We start by considering the notions of distance, convergence and continuity in a bit more detail. This will help us understand when optimization problems are well posed in the sense that they have optimal solutions.

1.1 Length and Distance in $\mathbb{R}^n$

The only spaces that we will be interested in these notes are the various Cartesian products of the real line $\mathbb{R}$ denoted by $\mathbb{R}^n$. The exponent n is also called the dimension of the Euclidean space. Hence an element $x \in \mathbb{R}^n$ is an ordered n-tuple $(x_1, ..., x_n)$ where each $x_i \in \mathbb{R}$.

Distance $d(x, y)$ between two vectors $x, y \in \mathbb{R}^n$ is usually based on the Euclidean norm or the length of a vector in $x \in \mathbb{R}^n$ defined by

$$\|x\| = \sqrt{\sum_{i=1}^{n} x_i^2}. \tag{1}$$

This is just the generalization of the Pythagorean theorem to an arbitrary dimension. A distance for $\mathbb{R}^n$ can be derived from this norm as

$$d(x, y) = \|x - y\|.$$

Proposition 1. Let x and y denote points in $\mathbb{R}^n$. Then we have:
(a) $\|x\| \geq 0$ and $\|x\| = 0$ if and only if $x = \mathbf{0}$,
(b) $\|ax\| = a\|x\|$ for every real a,
(c) $\|x - y\| = \|y - x\|$,
(d) $x \cdot y \leq \|x\| \, \|y\|$ (Cauchy-Schwarz inequality),
(e) $\|x + y\| \leq \|x\| + \|y\|$ (Triangle inequality).

Remark 1. To see why the Cauchy-Schwarz inequality is true, consider the sum of squares

$$\sum_{i=1}^{n} (x_i + t y_i)^2.$$

This is a quadratic polynomial in t, and as a sum of squares, it is also non-negative. Hence its discriminant is non-positive, i.e.

$$(2\sum_{i=1}^{n} x_i y_i)^2 \leq 4(\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2).$$

Dividing both sides by 4 and taking square roots on both sides gives Cauchy-Schwarz inequality.

This simple result is one of the most important results in all of mathematics. Equality holds in the result if and only if $x = \lambda y$, i.e. x is proportional to y. We have used this observation to argue that the gradient $\nabla f(\hat{x})$ gives the direction of steepest ascent for a function f at point $\hat{x}$.

From Cauchy-Schwarz, we get easily the triangle inequality:

$$\|x+y\|^2 = (x+y)\cdot(x+y) = \|x\|^2 + \|y\|^2 + 2x \cdot y$$
$$\leq \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| = (\|x\| + \|y\|)^2.$$

The triangle inequality follows by taking square roots on both sides of the inequality. The inequality in the above expression results from the Cauchy-Schwarz inequality.

Exercise In general, any function $\hat{d}(x,y)$ satisfying (a), (c) and (e) in the above list is a distance. It is a good exercise to show that $\hat{d}(x,y) := \max_i |x_i - y_i|$ is a distance in this sense. Are all the other properties above also satisfied by this distance?

By the segment (a,b) we mean the set of all real number x such that $a < x < b$. By the interval $[a,b]$, we mean the set of all real numbers such that $a \leq x \leq b$. If $a_i < b_i$ for $i = 1, ..., n$, the set of all points $\mathbf{x} = (x_1, ..., x_n)$ in $\mathbb{R}^n$ whose coordinates satisfy $a_i \leq x_i \leq b_i$ for $(1 \leq i \leq n)$, is called an n-cell. If $\mathbf{x} \in \mathbb{R}^n$ and $\varepsilon > 0$, the open (or closed) neighborhood $B^\varepsilon(x)$ with center at x and radius ε is defined to be the set of all $y \in \mathbb{R}^n$, such that $\|y - x\| < (\leq)\, r$.

1.2 Open and closed sets

In this subsection, we give some basic definitions on sets in $\mathbb{R}^n$.

Definition 1. A point x is a limit point of the set $E \subset \mathbb{R}^n$ if every neighborhood of x contains a point $y \in E$ with $y \neq x$.

We say that E is *closed* if every limit point of E is an element of E. A point x is an interior point of E if there is a neighborhood $B^\varepsilon(x)$ of x such that $B^{\varepsilon}(x) \subset E$. We say that E is *open* if every point of E is an interior point.

The *complement* of E, denoted by E^c is the set of all points $x \in \mathbb{R}^n$ such that $x \notin E$.

The set E is *bounded* if there is a real number M such that $\|x\| < M$ for all $x \in E$.

Exercise Is the empty set open or closed? Show that $A = \{x : a < x < b\}$ is an open set and that $A = \{x : a \leq x \leq b\}$ is a closed set. Show that the set $\{1, \frac{1}{2}, \frac{1}{3}, ...\}$ is neither open nor closed (hint: is 0 a limit point? Is it in the set?)

Proposition 2. A set $E \subset \mathbb{R}^n$ is open if and only if its complement is closed. A set $F \subset \mathbb{R}^n$ is closed if and only if its complement is open.

A very important property for sets in mathematical analysis is called *compactness*. we give here a definition of compactness for sets in $\mathbb{R}^n$ that should really be derived as a theorem starting from a more fundamental notion, but for practical matters, this is all we need.

Definition 2 (Compact sets)**.** A set $E \subset \mathbb{R}^n$ is called *compact* if it is closed and bounded.

1.3 Sequences

Definition 3. If S is any set, a sequence in S is a function on the set $\mathbb{N} = \{1, 2, 3, ...\}$ of natural numbers and whose range is in S.

Definition 4. A sequence $\{x_n\}$ in $\mathbb{R}^n$ is said to converge if there is a point $x \in \mathbb{R}^n$ with the following property: For every $\epsilon > 0$, there is an integer N such that $n \geq N$ implies that $d(x_n, x) < \epsilon$.

We say that x_n converges to x, x is the limit of $\{x_n\}$ and we write $x_n \to x$,

$$\lim_{n\to\infty} x_n = x.$$

Theorem 1. Let $\{x_n\}$ be a sequence in $\mathbb{R}^n$.

(i) $\{x_n\}$ converges to $x \in \mathbb{R}^n$ if and only if every neighborhood of x contains all but finitely many of the terms of $\{x_n\}$.

(ii) If $x \in \mathbb{R}^n, x' \in \mathbb{R}^n$, and if $\{x_n\}$ converges to x and to x', then $x = x'$.
(iii) If $\{x_n\}$ converges, then $\{x_n\}$ is bounded.
(iv) If $E \subset \mathbb{R}^n$ and x is a limit point of E, then there is a sequence $\{x_n\}$ in E such that $x = \lim_{n\to\infty} x_n$.
(v) $x_n = (x_{1,n}, ..., x_{k,n}) \to x = (x_1, ..., x_k) \Leftrightarrow x_{i,n} \to x_i$ for all $i \in \{1, ..., k\}$.

The last part of the proposition claims that a sequence of vectors converges if and only if all of its coordinates converge.

Definition 5. Given a sequence $\{x_n\}$, consider an infinite sequence $\{n_k\}$ of positive integers, such that $n_1 < n_2 < \cdots$. Then the sequence $\{x_{n_i}\}$ is called a subsequence of $\{x_n\}$. If $\{x_{n_i}\}$ converges, its limit is called a subsequential limit of $\{x_n\}$.

Exercise Show that if $\{x_n\}$ converges to x, then all of its subsequences also converge to x.

Definition 6. A sequence $\{x_n\}$ is said to be a Cauchy sequence if for every $\epsilon > 0$ there is an integer N such that $d(x_n, x_m) < \epsilon$, if $n \geq N$ and $m \geq N$.

Real numbers are constructed in such a way that Cauchy sequences in $\mathbb{R}$ converge, i.e. have limits in $\mathbb{R}$. By part (v) of the previous theorem, the same is true for real vectors.

Theorem 2 (Weierstrass)**.** Every bounded subset $E \subset \mathbb{R}^n$ with infinitely many elements has a limit point in $\mathbb{R}^n$.

Idea of proof for $\mathbb{R}$: Since E is bounded, it is contained in an interval $[-M, M]$ of length $2M$ for some $M < \infty$. Since E has infinitely many elements, either $[-M, 0]$ or $[0, M]$ or both have infinitely many elements. Hence some interval of length M also contains infinitely many elements of E. Continue this process of halving the interval to show that you can come up with a sequence of intervals of length $2^{-k}M$ containing infinitely many elements of E. The midpoints of the sequences form a Cauchy sequence and hence they converge to a point $x \in \mathbb{R}$. This x is a limit point of E. The same construction generalizes easily to $\mathbb{R}^n$

An immediate consequence of this is the following theorem.

Theorem 3. (Bolzano-Weierstrass Theorem)

Every bounded sequence in $\mathbb{R}^n$ contains a convergent subsequence and every sequence in a compact set $E \in \mathbb{R}^n$ has a convergent subsequence whose limit is in E.

1.4 Continuous Functions

Definition 7. Consider a function $f : \mathbb{R}^n \to \mathbb{R}^m$. We write $f(x) \to \hat{y}$ as $x \to \hat{x}$, or

$$\lim_{x \to \hat{x}} f(x) = \hat{y}, \tag{2}$$

if there is a point $y \in \mathbb{R}^m$ with the following property: For every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$x \in B^{\delta}(\hat{x}) \Rightarrow f(x) \in B^{\varepsilon}(\hat{y}).$$

We say that f is *continuous at* $\hat{x}$ if for all $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$x \in B^{\delta}(\hat{x}) \Rightarrow f(x) \in B^{\varepsilon}(f(\hat{x})).$$

Another way of writing this is given in the following simple proposition.

Proposition 3. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous at $\hat{x}$ if for every sequence $\{x_n\}$ that converges to $\hat{x}$, the sequence $\{f(x_n)\}$ converges to $f(\hat{x})$; in symbols,

$$\lim_{n \to \infty} f(x_n) = f\left(\lim_{n \to \infty} x_n\right).$$

A function is said to be continuous if it is continuous at all points in its domain. Continuity of a function f at a point $\hat{x}$ is called a local property of f because it depends on the behavior of f only in the immediate vicinity of $\hat{x}$. A property of f which concerns the whole domain of f is called a global property. Thus, continuity of f on its domain is a global property.

The following proposition gives yet another way of looking at continuity.

Proposition 4. A function f is continuous if and only if the inverse image $f^{-1}(V)$ is open (closed) for every open (closed) set V in Y.

Proposition 5. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}^k$ be continuous functions, and let h be the composite function defined by

$$h(x) = g(f(x)) \qquad \text{for } x \in \mathbb{R}^n.$$

If f is continuous at $\hat{x}$ and if g is continuous at $f(\hat{x})$, then h is continuous at $\hat{x}$.

1.5 Global Properties of Continuous Functions

Definition 8. A function $f : E \to \mathbb{R}$ is said to be bounded if there is a real number M such that $|f(x)| \leq M$ for all $x \in E$.

Recall the definition of the least upper bound and greatest lower bound for a set A of real numbers. We say that $\overline{a}$ is the least upper bound of A if for all $x \in A$, $x \leq \overline{a}$ and for all $a' < \overline{a}$, there is some $x \in A$ such that $x > a'$. Similarly, we say that $\underline{a}$ is the greatest lower bound of A if for all $x \in A$, $x \geq \underline{a}$ and for all $a' > \underline{a}$, there is some $x \in A$ such that $x < a'$.

We write:

$$\overline{a} := \sup A, \underline{a} := \inf A.$$

Theorem 4 (Weierstrass' Theorem)**.** Suppose f is a continuous function on a compact set E, and

$$M = \sup_{x \in E} f(x), \qquad m = \inf_{x \in E} f(x).$$

Then there exists a point $\overline{x}, \underline{x} \in E$ such that $f(\overline{x}) = M$ and $f(\underline{x}) = m$.

Proof. We show this for the supremum. The case for the infimum is analogous. Let $M = \sup_{x \in E} f(x)$. Let $\{m_n\} \to M$ with $m_n < M$ for all n. Then By the definition of supremum, there must be a sequence $\{x_n\} \in E$ with $x_n \geq m_n$. Since E is compact, $\{x_n\}$ has a convergent subsequence $\{x_{n_k}\} \to x \in E$. Since $\{m_n\} \to M$, we also know that $\{m_{n_k}\} \to M$. By continuity of f,

$$M \geq f(x) = \lim f(x_{n_k}) \geq \lim m_{n_k} = M.$$

□

This theorem ensures that our maximization and minimization problems have solutions as long as the objective function is continuous and the feasible set is compact.

Remark 2. To see that E must be closed and bounded and that f has to be continuous, consider the following examples:

1. $f(x) = x$ and $E = \mathbb{R}$.
2. $f(x) = x$ and $E = \{x : 0 < x < 1\}$.
3. $f(x) = x$ for $0 \leq x < 1, f(1) = 0$ and $E = \{x : 0 \leq x \leq 1\}$.

2 Constrained Optimization

We start with some simple examples of constrained optimization to set up expectations for the more general case to follow.

Example 1. Consider finding the maximum for $f(x) = 3 + 2x - x^2$ on the feasible set $F = \{x : -\infty < a \leq x \leq b < \infty$. Since f is continuous and the feasible set F is compact. Therefore Weierstrass' theorem guarantees the existence of a maximizer, i.e. an $x \in F$ such that for all $y \in F$, we have $f(x) \geq f(y)$.

Notice that f is strictly increasing for $x < 1$ and strictly decreasing for $x > 1$. If $a \leq 1 \leq b$, then the function is maximized at its critical point $x = 1$. We say that a direction $(x - x_0)$ is feasible from $x_0 \in F$ if for a small Δ, we have $x_0 + \Delta(x - x_0) \in F$. Linear approximation by the derivative gives:

$$f(x_0 + \Delta(x - x_0)) - f(x_0) = f'(x_0)\Delta(x - x_0).$$

If we have a maximum at x_0, then for all feasible direction

$$f'(x_0)\Delta(x - x_0) \leq 0.$$

If $a < x_0 < b$, then we must have $f'(x_0) = 0$ since both directions $x > x_0, x < x_0$ are feasible. If $f'(x_0) > 0$, then $x > x_0$ cannot be feasible if x_0 is a maximum. Therefore $x_0 = b$ if x_0 is the optimal choice and $f'(x_0) > 0$. Similarly, if $f'(x_0) < 0$ and x_0 is the optimum, then $x_0 = a$.

If all directions are feasible from x_0 and x_0 is a maximum, then just as in the case of unconstrained optimization, we must have $f'(x_0)$. For the other cases, the derivative of the objective function at optimum is closely related to the constraint that binds (i.e. restricts the feasible directions).

In the next subsections, we will generalize our findings to multidimensional optimization problems.

2.1 Optimization with a single equality constraint

We start with local considerations. Let $f : \mathbb{R}^n \to \mathbb{R}$ be the objective function to be maximized and suppose the constraints take the form $g(x) = g(x_1, ..., x_n) = 0$. In other words, $F = \{x : g(x) = 0\}$. We write the maximization problem often as:

$$\max_x f(x)$$
$$\text{subject to } g(x) = 0.$$

A solution to this problem finds point $\hat{x}$ such that $f(\hat{x}) \geq f(x)$ for all $x \in F$. What can we say about such an $\hat{x}$? At this point, we do not know if it exists. If it exists, and f is differentiable, then for small Δ,

$$f(\hat{x} + \Delta(x - \hat{x})) - f(\hat{x}) = Df(\hat{x})(x - \hat{x})\Delta \leq 0$$

for all feasible directions $(x - \hat{x})$. But how do we know which directions are feasible? Assume that the function g defining the constraint is also differentiable. To find the feasible directions, we go back to implicit function theorem. If $\hat{x} \in F$ and $\frac{\partial g}{\partial x_i}(\hat{x}) \neq 0$ for some $i \in \{1, ..., n\}$, then we can find a write $x_i = h(x_1, ..., x_{i-1}, x_{i+1}, ..., x_n) =: h(x_{-i})$ in a neighborhood of $\hat{x}_{-i}$ so that

$$g(h(x_{-i}), x_{-i}) = 0.$$

Notice that it is not possible to use the implicit function theorem if at a critical point of the constraint function. Therefore, we must assume that $Dg(\hat{x}) \neq 0$. We call this the constraint qualification and we will see different versions of this in more complex situations with multiple constraints.

Since the function g is at constant value in the feasible set, we have for all feasible directions $(x - \hat{x})$:

$$\nabla g(\hat{x})(x - \hat{x}) = 0.$$

Notice also that if $(x - \hat{x})$ is feasible, then also $-(x - \hat{x})$ is feasible. From the linear approximation above, this means immediately that for all feasible directions,

$$Df(\hat{x})(x - \hat{x}) = 0.$$

But therefore we have shown that at optimum $\hat{x}$,

$$\nabla f(\hat{x}) = \mu \nabla g(\hat{x}).$$

We have derived the following necessary condition for a constrained optimum at $\hat{x}$: the gradient of the objective function must be a scalar multiple of the gradient of the constraint function at the optimum. The second requirement is that the choice must be feasible, i.e. $g(\hat{x}) = 0$.

2.1.1 The Lagrangean function

The previous discussion motivates the following function that incorporates the constraints into an augmented objective function called the Lagrangean function.

For a constrained optimization problem, we define the following function of $n+1$ variables:

$$\mathcal{L}(x,\mu) = f(x) - \mu g(x).$$

We call the new variable μ the Lagrange multiplier. We will give it a good economic interpretation later in the course. We are interested in the critical points of this augmented function. Therefore we look for $(\hat{x},\hat{\mu})$ such that

$$\frac{\partial \mathcal{L}}{\partial x_i}(\hat{x},\hat{\mu}) = \frac{\partial f}{\partial x_i}(\hat{x}) - \hat{\mu}\frac{\partial g}{\partial x_i}(\hat{x}) = 0 \text{ for all } i,$$
$$\frac{\partial \mathcal{L}}{\partial \mu}(\hat{x},\hat{\mu}) = g(\hat{x}) = 0.$$

As argued above, these are the first-order conditions for the constrained optimization problem. In order to know if we have found a local maximum or a minimum, we need to look at the second-order Taylor -approximations and the definiteness of the Hessian matrix at the critical point.

As before, write the second-order Taylor approximation to $f:\mathbb{R}^n \to \mathbb{R}$ at $\hat{x}$ as:

$$f(x) = f(\hat{x}) + Df(\hat{x})(x-\hat{x}) + (x-\hat{x})\cdot D^2(\hat{x})(x-\hat{x}).$$

If $\hat{x}$ is a maximum, then for all feasible directions $(x-\hat{x})$, we have

$$i) Df(\hat{x})(x-\hat{x}) = 0,$$

$$ii) (x-\hat{x})\cdot D^2(\hat{x})(x-\hat{x}) \leq 0.$$

Since the feasible directions are give by vectors $(x-\hat{x})$ such that

$$\nabla g(\hat{x})\cdot(x-\hat{x}) = 0,$$

the condition for having a local maximum at $\hat{x}$ is equivalent to checking the negative definiteness of the bordered Hessian where we need

1. The Lagrangean $\mathcal{L}$
2. The Equality constraint h

To get the bordered Hessian, start with the derivative of the Lagrangean with respect to the choice variables x at the critical point $\hat{x}$: $D_x^2\mathcal{L}(\hat{x})$ and 'border' it with the derivative of the constraint function (to capture the restriction to feasible directions).

$$D^2\mathcal{L} = \begin{bmatrix} 0 & Dg(\hat{x}) \\ [Dg(\hat{x})]^T & D_x^2\mathcal{L}(\hat{x}) \end{bmatrix}$$

In the special case where we have only two choice variables, I let the variables be x, y for notational ease, we need to examine

$$D^2\mathcal{L} = \begin{bmatrix} 0 & g_x & g_y \\ g_x & \mathcal{L}_{xx} & \mathcal{L}_{xy} \\ g_y & \mathcal{L}_{yx} & \mathcal{L}_{yy} \end{bmatrix}$$

How do we determine the negative definiteness of the bordered Hessian?

1. Leading principal minors must alternate in sign [1]
2. $\det D^2\mathcal{L}(\hat{x})$ must have the same sign as $(-1)^n$.

How many principal minors to examine?

- You need to check the sign of the last $(n-1)$ leading principal minors
- For completeness, I state here that with more constraints, you need to border the Hessian with the derivatives of all binding constraints. If you have k such constraints, then you need to examine the sign of $(n-k)$ leading principal minors.

[1]Recall that a leading principal minor of k^{th} order is obtained from a matrix A by deleting its last k rows and columns.

Bordered Hessians are a bit of a nightmare for me. They are tedious to compute and they tell nothing of significance in the end. We will see later how we can bypass them to a large extent by concentrating on a subset of problems where first-order conditions also turn out to be sufficient.

In any case, here is eventually a concrete example:

Example 2. Find the minima and maxima of $f(x, y, z) = x + y + z^2$ subject to constraints

$$\begin{aligned} x^2 + y^2 + z^2 &= 1 \\ y &= 0 \end{aligned}$$

Start by substituting the second constraint to the objective function and the first constraint to get $f(x, z) = x + z^2$, and

$$x^2 + z^2 = 1$$

Form the Lagrangean

$$\mathcal{L}(x, z, \mu) = x + z^2 - \mu(x^2 + z^2 - 1)$$

Differentiate to get the first-order conditions (FOC):

$$\frac{\partial \mathcal{L}}{\partial x} = 1 - 2\mu x = 0 \tag{3}$$
$$\frac{\partial \mathcal{L}}{\partial y} = 2z - 2\mu z = 0 \tag{4}$$
$$\frac{\partial \mathcal{L}}{\partial \mu} = 1 - x^2 - z^2 = 0 \tag{5}$$

The second FOC gives:

$$z(2 - 2\mu) = 0$$

Therefore either $z = 0$, or $\mu = 1$.

Consider first the possibility that $z = 0$. In that case, (5) implies that $x = \pm 1$. We get two critical points from (3):

$$\left(x = 1, y = 0, z = 0, \mu = \tfrac{1}{2}\right) \text{ and } \left(x = -1, y = 0, z = 0, \mu = -\tfrac{1}{2}\right)$$

If $\mu = 1$, (3) implies that $x = \frac{1}{2}$. By substituting into (5) we get the critical points:

$$\left(x = \tfrac{1}{2}, y = 0, z = \tfrac{\sqrt{3}}{2}, \mu = 1\right) \text{ and } \left(x = \tfrac{1}{2}, y = 0, z = -\tfrac{\sqrt{3}}{2}, \mu = 1\right)$$

As a result, we have four critical points for the Lagrangean. Draw the constraint set and level curves for the objective function to get a guess of the classification of the critical points.

By examining the bordered Hessian, we see that $\left(x = -1, y = 0, z = 0, \mu = -\frac{1}{2}\right)$ and $\left(x = 1, y = 0, z = 0, \mu = \frac{1}{2}\right)$ are local minima, and $\left(x = \frac{1}{2}, y = 0, z = \pm\frac{\sqrt{3}}{2}, \mu = 1\right)$ are local maxima.

Can you show the existence of a maximum? Which of the local maxima is the global maximum?

2.2 Multiple equality constraints

Consider next the case, where we have k equality constraints $g(x) = (g_1(x), ..., g_k(x) : \mathbb{R}^n \to \mathbb{R}^k$. In this case, we have the problem:

$$\begin{aligned} &\max_x f(x) \\ \text{subject to } &g_1(x) = 0, \\ &g_2(x) = 0, \\ &\vdots \\ &g_k(x) = 0. \end{aligned}$$

Form the Lagrangean now with k constraints as a function of $n + k$ variables:

$$\mathcal{L}(x, \mu_1, ..., \mu_k) = f(x) - \sum_{j=1}^{k} \mu_j g_j(x).$$

We can proceed exactly as before to use the linear approximations to characterize the feasible directions from $\hat{x}$ as $\{(x-\hat{x}) : Dg(\hat{x})(x-\hat{x})\} = 0$. Since the objective function cannot increase at the optimum in any feasible direction, we have that

$$Df(\hat{x})(x-\hat{x}) = 0 \text{ whenever } Dg(\hat{x})(x-\hat{x}) = 0.$$

If $Dg(\hat{x})$ has full rank, then this is equivalent to requiring that $Df(\hat{x})$ and $Dg_j(\hat{x})$ must be linearly dependent. Since we assume that $Dg(\hat{x})$ has full rank, this means that there must exist $(\mu_1, ..., \mu_k)$ such that

$$\nabla f(\hat{x}) = \sum_{j=1}^{k} \mu_j \nabla g_j(\hat{x}).$$

Hence we can summarize the three necessary conditions for local maximum:

i) Gradient alignment: $\nabla f(\hat{x}) = \sum_{j=1}^{k} \mu_j \nabla g_j(\hat{x})$,

ii) Constraint holds: $g(\hat{x}) = 0$,

iii) Constraint qualification: $Dg_1(\hat{x}), ..., Dg_k(\hat{x})$ are linearly independent.

The first two can be achieved by requiring that $(\hat{x}, \hat{\mu}_1, ..., \hat{\mu}_k)$ be a critical point of the Lagrangean. The second-order conditions are based on bordered Hessian matrices as explained at the end of the previous subsection.

Let's end this section with another example

Example 3.

Consider the objective function

$$f(x, y, z) = xz + yz$$

and a maximization problem subject to:

$$g_1(x, y, z) = y^2 + z^2 - 1$$
$$g_2(x, y, z) = xz - 3$$

1. Find the critical points of f subject to constraints $g_1(x, y, z) = 0$ and $g_2(x, y, z) = 0$.

2. How would you determine which of the critical points are local minima and which are local maxima? What is the bordered Hessian that you would use?

1. Find first the critical points of the Lagrangean

$$\mathcal{L}(x, y, z, \mu_1, \mu_2) = xz + yz - \mu_1(y^2 + z^2 - 1) - \mu_2(xz - 3)$$

First-order conditions:

$$\frac{\partial \mathcal{L}}{\partial x} = z - \mu_2 z = 0 \tag{6}$$
$$\frac{\partial \mathcal{L}}{\partial y} = z - 2\mu_1 y = 0 \tag{7}$$
$$\frac{\partial \mathcal{L}}{\partial z} = x + y - 2\mu_1 z - \mu_2 x = 0 \tag{8}$$
$$\frac{\partial \mathcal{L}}{\partial \mu_1} = y^2 + z^2 - 1 = 0 \tag{9}$$
$$\frac{\partial \mathcal{L}}{\partial \mu_2} = xz - 3 = 0 \tag{10}$$

We need to solve this system of equations to find the critical points. Start with (6), giving

$$z(1 - \mu_2) = 0, \Leftrightarrow \ z = 0 \text{ or } \mu_2 = 1$$

If $z = 0$, then (10) is not true for any x and as a result, we must have$z \neq 0$. Therefore, we can only have $\mu_2 = 1$ as a candidate solution. The second FOC (7) gives

$$y - 2\mu_1 z = 0, \ \Leftrightarrow y = \frac{z}{2\mu_1}.$$

Plug in the solutions for y and μ_2 into (8) :

$$\frac{z}{2\mu_1} - 2\mu_1 z = 0 \iff z\left(\frac{1}{2\mu_1} - 2\mu_1\right) = 0.$$

We already know that $z \neq 0$, and therefore

$$\frac{1}{2\mu_1} - 2\mu_1 = 0 \iff 4\mu_1^2 = 1 \iff \mu_1 = \pm\frac{1}{2}$$

We have now solved for possible Lagrange multipliers μ_1 ja μ_2, i.e. we have:

$\mu_1 = \pm\frac{1}{2}$ and $\mu_2 = 1$

To get the values of the choice variables, plug in the values of the multipliers into (8) to get:

$$y = \pm z.$$

Substituting into (9), we get (by squaring):

$$2z^2 - 1 = 0 \iff z = \pm\frac{1}{\sqrt{2}}$$

The fifth FOC (10) gives:

$$x = \frac{3}{z},$$

or $x = 3\sqrt{2}$ if $z = \frac{1}{\sqrt{2}}$ and $x = -3\sqrt{2}$ if $z = -\frac{1}{\sqrt{2}}$. We have now found all that we need for the critical points of f subject to the constraints. If $z = \frac{1}{\sqrt{2}}$, then $x = 3\sqrt{2}$, $y = \pm z$. This yields two critical points (x, y, z):

$$1: \left(3\sqrt{2}, \tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}\right)$$

$$2: \left(3\sqrt{2}, -\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}\right)$$

If $z = -\frac{1}{\sqrt{2}}$, then $x = -3\sqrt{2}$, $y = \pm z$. This gives also two critical points (x, y, z):

$$3: \left(-3\sqrt{2}, -\tfrac{1}{\sqrt{2}}, -\tfrac{1}{\sqrt{2}}\right)$$

$$4: \left(-3\sqrt{2}, \tfrac{1}{\sqrt{2}}, -\tfrac{1}{\sqrt{2}}\right)$$

We know that for all critical points, $\mu_2 = 1$, and we can check the sign of μ_1 from FOC (7). After this, we have all the critical points of the problem as:

Critical points for the problem (x, y, z, μ_1, μ_2):

$$\begin{aligned}
1: &\left(3\sqrt{2}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{2}, 1\right)\\
2: &\left(3\sqrt{2}, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, -\frac{1}{2}, 1\right)\\
3: &\left(-3\sqrt{2}, -\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, \frac{1}{2}, 1\right)\\
4: &\left(-3\sqrt{2}, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, -\frac{1}{2}, 1\right)
\end{aligned}$$

We can plug these into the objective function to see which of the critical points could be the true maximum. Do we know now that the problem has a maximum?

2. In order to determine which of the critical points are maxima or minima, we need to consider the bordered Hessian by bordering the Hessian of $\mathcal{L}$ with respect to (x, y, z), i.e. $D_x^2 L$ with the derivatives of the constraints $D_x g_1$ and $D_x g_2$.

$$D^2\mathcal{L} = \begin{bmatrix} 0 & Dg(\hat{x}) \\ [Dg(\hat{x})]^T & D_x^2\mathcal{L} \end{bmatrix}$$

With three dimensional choice variable, we know that $D_x^2 L$ is a $[3 \times 3]$ -matrix. With two constraints $Dg(\hat{x})$ is a $[2 \times 3]$ -matrix. Hence we have a (5×5) square matrix for our bordered Hessian:

$$D^2 L = \begin{bmatrix} 0 & 0 & g_{1x} & g_{1y} & h_{1z} \\ 0 & 0 & g_{2x} & g_{2y} & g_{2z} \\ g_{1x} & g_{2x} & \mathcal{L}_{xx} & \mathcal{L}_{xy} & \mathcal{L}_{xz} \\ g_{1y} & g_{2y} & \mathcal{L}_{yx} & \mathcal{L}_{yy} & \mathcal{L}_{yz} \\ g_{1z} & g_{2z} & \mathcal{L}_{zx} & \mathcal{L}_{zy} & \mathcal{L}_{zz} \end{bmatrix},$$

where we have denoted the partial derivatives by g_{ix}, and second-order partial derivatives by of the Lagrangean by L_{xx} etc. As you can imagine, it is not a pure pleasure to check the definiteness of this bordered Hessian for the four critical points (even though you have to compute a single determinant).