

# Finite Dimensional Optimization

## Part I: The KKT Theorem<sup>1</sup>

### 1 Introduction

These notes characterize maxima and minima in terms of first derivatives. I focus primarily on maximization. The problem of minimizing a function  $f$  has the same solution (or solutions) as the problem of maximizing  $-f$ , so all of the results for maximization have easy corollaries for minimization.

The main result of these notes is the Karush-Kuhn-Tucker (KKT) Theorem, recorded as Theorem 2 in Section 4.4. The KKT Theorem was formulated independently, first in Karush (1939) and later in Kuhn and Tucker (1951). Karush's contribution was unknown for many years and it is common to see the KKT Theorem referred to as Kuhn-Tucker (and I still sometimes do this in my own notes).

These notes cover only necessary conditions, conditions that solutions to maximization problems must satisfy. Part II of these notes discuss how to guarantee that a candidate solution is indeed a maximum (or a minimum, or an inflection point, or saddle point, etc.). Part III of these notes develops some of the complementary machinery of convex programming. One of the main attractions of convex programming is that it extends to situations where the functions are not differentiable.

### 2 Basic definitions.

Consider a function  $f : X \rightarrow \mathbb{R}$ , where  $X \subseteq \mathbb{R}^N$  is open and convex.

In economic applications, the economically meaningful domain is often  $\mathbb{R}_+^N$ , which is not open. In such cases, I assume that  $f$  can be extended or restricted to an open set. For example, consider a competitive demand problem with utility  $u : X \rightarrow \mathbb{R}$ .

1. If  $u(x) = \sum_n a_n x_n$  with all  $a_n > 0$ , then this  $u$  is defined on all of  $\mathbb{R}^N$ . In this case, the condition  $x_n \geq 0$  will be expressed as an explicit constraint, in addition to the budget constraint, in the formal statement of the problem.
2. If  $u(x) = \sqrt{x_1} + x_2$ , I take the domain to be  $\mathbb{R}_{++} \times \mathbb{R}$ . There will not, in fact, be a solution with  $x_1 = 0$ . And, again, the condition  $x_2 \geq 0$  will be expressed as an explicit constraint in the formal statement of the problem.

---

<sup>1</sup>©️🔗📄🔄. This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License.

Suppose that we are interested in the behavior of  $f$  on a non-empty subset  $C \subseteq X$ . For example, in a competitive demand problem,  $f$  is a utility function, and  $C$  is the set of *feasible* consumption bundles: non-negative and affordable. In general, I refer to  $f$  as the *objective function* and  $C$  as the *constraint set*, also called the *feasible set* or *opportunity set*.

**Definition 1.** Let  $C \subseteq X \subseteq \mathbb{R}^N$ ,  $f : X \rightarrow \mathbb{R}$ , and  $x^* \in C$ .

1.  $x^*$  is a local maximum iff and there is an  $\varepsilon > 0$  such that  $f(x^*) \geq f(x)$  for any  $x \in N_\varepsilon(x^*) \cap C$ .
2.  $x^*$  is an interior local maximum iff there is an  $\varepsilon > 0$  such that  $N_\varepsilon(x^*) \subseteq C$  and  $f(x^*) \geq f(x)$  for all  $x \in N_\varepsilon(x^*)$ .
3.  $x^*$  is a (global) maximum iff  $f(x^*) \geq f(x)$  for all  $x \in C$ .

The definitions for minima are analogous, but with inequalities reversed.

The following examples illustrate these concepts. In all of the examples,  $f : \mathbb{R} \rightarrow \mathbb{R}$ .

*Example 1.*  $f(x) = -x^2$ .  $C = \mathbb{R}$ . Then  $x^* = 0$  is the maximum.  $\square$

*Example 2.*  $f(x) = x^2$ .  $C = \mathbb{R}$ . Then  $x^* = 0$  is the *minimum*. There is no maximum.  $\square$

*Example 3.*  $f(x) = x$ .  $C = [-1, 1]$ . Then  $x^* = 1$  is the maximum. It is not interior. The restriction to  $C$  is binding, meaning that the solution is at a boundary of  $C$ .  $\square$

*Example 4.* Let  $f(x) = -x(x-1)(x+1)^2$ .  $C = \mathbb{R}$ . Here,  $x = -1$  is a local maximum but not a global maximum. The global maximum is at  $x^* = (1 + \sqrt{17})/8$ .  $\square$

### 3 Interior Maxima.

I begin by recording the fact, which should already be familiar, that at an interior local maximum, the gradient of the objective function  $f$  is zero. The intuition is that the gradient of  $f$  gives the direction of fastest increase in  $f$ , and at an interior maximum,  $f$  is not increasing in any direction.

**Theorem 1.** Consider  $f : X \rightarrow \mathbb{R}$ . If  $x^*$  is an interior local maximum of  $f$  and  $f$  is differentiable at  $x^*$  then

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

**Proof.** By contraposition. Suppose that  $\nabla f(x^*) \neq 0$ . Then  $\nabla f(x^*) \cdot \nabla f(x^*) > 0$ , which implies that the directional derivative of  $f$  in the direction  $\nabla f(x^*)$  is strictly positive. Therefore, there is  $\bar{\gamma}$  such that for all  $\gamma \in (0, \bar{\gamma})$ ,

$$f(x^* + \gamma \nabla f(x^*)) - f(x^*) > 0,$$

which implies that  $x^*$  is not a local maximum. ■

By an analogous argument, if  $x^*$  is a local *minimum* then, again,  $\nabla f(x^*) = 0$ . I often refer to  $\nabla f(x^*) = 0$  as the (*unconstrained*) *first-order condition (FOC)*. It is “first order” in the sense that it involves only first derivatives.

The condition  $\nabla f(x^*) = 0$  is necessary for a local maximum but not sufficient for a local, let alone a global, maximum. In the examples below,  $f : \mathbb{R} \rightarrow \mathbb{R}$  and  $C = \mathbb{R}$ .

*Example 5.* Let  $f(x) = x^3$ . Then  $\nabla f(0) = 0$  but  $x^* = 0$  is neither a local maximum nor a local minimum.

*Example 6.* Let  $f(x) = x^2$ . Then  $\nabla f(0) = 0$  but  $x^* = 0$  is not a local maximum; it is, however, a global minimum

*Example 7.* Recall Example 4. The gradient is zero at  $-1$ ,  $(1 - \sqrt{17})/8$ , and  $(1 + \sqrt{17})/8$ . The first is a local, but not global maximum, the second is a local but not global minimum (there is no minimum), and the last is the global maximum.

## 4 Constrained Maxima – The KKT Theorem.

### 4.1 The constrained maximization problem in standard form.

To proceed further I need to establish some notational conventions. As before, let  $C \subseteq X \subseteq \mathbb{R}^N$  and consider the maximization problem,

$$\max_{x \in C} f(x).$$

The main result of these notes, the Karush-Kuhn-Tucker (KKT) Theorem, assumes that  $C$  is written in a particular manner, which I refer to as *standard form*. If  $C$  is not expressed in standard form then KKT is still true *provided* its conclusion is reformulated. This is not difficult to do, but from a purely practical standpoint, one can avoid error in using KKT by remembering to write  $C$  in standard form.

To illustrate standard form, consider the consumer utility maximization problem,  $\max_{x \in C} f(x)$  where  $C = \{x \in \mathbb{R}_+^N : p \cdot x \leq m\}$ . This is often written,

$$\begin{aligned} \max_x \quad & f(x) \\ \text{s.t.} \quad & p \cdot x \leq m, \\ & x \geq 0. \end{aligned}$$

In standard form, the problem is,

$$\begin{aligned} \max_x \quad & f(x) \\ \text{s.t.} \quad & p \cdot x - m \leq 0, \\ & -x_1 \leq 0, \\ & \vdots \\ & -x_N \leq 0. \end{aligned}$$

More generally, a maximization problem is in *standard form* iff there are  $K$  functions  $g_k : X \rightarrow \mathbb{R}$  such that  $C$  is the set of  $x$  such that  $g_k(x) \leq 0$  for all  $k$ . More concisely, letting  $g = (g_1, \dots, g_K)$ , the canonical maximization problem in standard form, MAX, is

$$\begin{aligned} \max_x \quad & f(x) \\ \text{s.t.} \quad & g(x) \leq 0. \end{aligned}$$

In particular,  $C = g^{-1}((-\infty, 0])$ .

Similarly, the canonical minimization in standard form, MIN, is

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & g(x) \geq 0. \end{aligned}$$

Note that constraints in MAX use  $\leq$  while the constraints in MIN use  $\geq$ . As a mnemonic, in MAX problems think of constraints as imposing “ceilings” blocking further increases in  $f$  via increases in  $x$ . This motivates the “less than or equal” form of the standard constraint. Conversely, for MIN problems think of constraints as establishing “floors” preventing further decreases in  $f$  via decreases in  $x$ , hence the “greater than or equal” form of the standard constraint. This caricature of the MAX and MIN problems should not be taken too seriously. For example, the constraints  $x \geq 0$  in the consumer maximization problem are really “floors” even though in the standard form they are expressed as “ceilings” ( $-x \leq 0$ ).

## 4.2 The KKT Theorem – An Informal Statement.

Let  $x^*$  be feasible, so that  $g_k(x^*) \leq 0$  for all  $k$ . Say that constraint  $k$  is *binding* at  $x^*$  if  $g_k(x^*) = 0$ . If constraint  $k$  is not binding, so that  $g_k(x^*) < 0$ , say that  $k$  constraint is *slack*. Define  $J$  to be the set of indices of the binding constraints:

$$J = \{k \in \{1, \dots, K\} : g_k(x^*) = 0\}.$$

From the perspective of first order conditions, only the binding constraints matter.

KKT states that if  $x^*$  is a local maximum, and if  $J \neq \emptyset$ , then, subject to a technical condition called *constraint qualification* discussed in Section 4.3, for each  $k \in J$  there is a number  $\lambda_k^* \geq 0$  such that,

$$\nabla f(x^*) = \sum_{k \in J} \lambda_k^* \nabla g_k(x^*).$$

I refer to this as the *KKT condition*. (If  $J = \emptyset$  then the KKT condition is  $\nabla f(x^*) = 0$ , as in Theorem 1.) The  $\lambda_k^*$  are called *KKT multipliers*. The KKT condition for a MIN problem (in standard form) is identical.

For geometric intuition, suppose that  $J \neq \emptyset$ , and let  $A$  be the set that is positively spanned by the gradients of the binding constraints:

$$A = \left\{ x \in \mathbb{R}^N : \exists \lambda_k \geq 0 \text{ s.t. } x = \sum_{k \in J} \lambda_k \nabla g_k(x^*) \right\}.$$

Then  $A$  is a *cone*: for any  $a \in A$  and any  $\lambda \geq 0$ ,  $\lambda a \in A$ . KKT thus says that, if  $J \neq \emptyset$ , then  $\nabla f(x^*)$  lies in the cone  $A$ . See Figure 1. Very informally, the intuition behind KKT is that if  $x^*$  is a local maximum but  $\nabla f(x^*) \neq 0$  then moving in a direction that increases  $f$  must violate one of the binding constraints, which means moving in a direction that increases at least one of the binding  $g_k$  functions. The proof of KKT verifies that  $\nabla f(x^*) \in A$  formalizes this intuition.

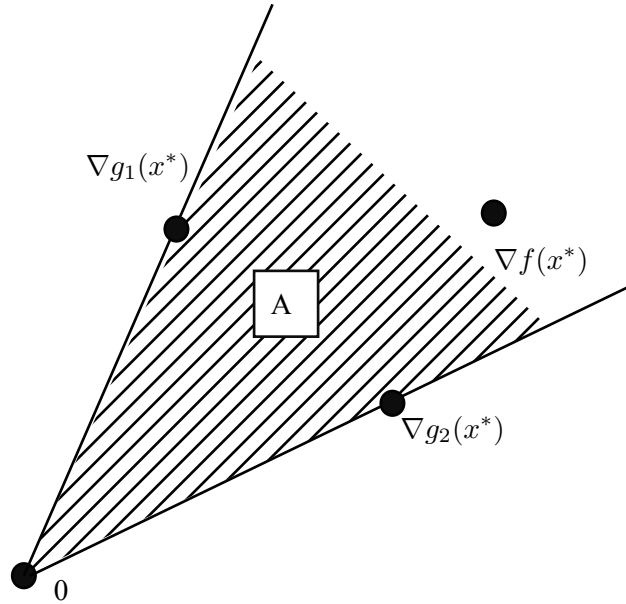


Figure 1: The cone  $A$  and the KKT condition.

As I discuss in the notes on the Envelope Theorem,  $\lambda_k^*$  measures how much the objective function would be increased if constraint  $k$  were relaxed slightly. For example, in the standard utility maximization problem, the KKT multiplier on the budget constraint measures how much utility would increase if the consumer had a bit more income.

### 4.3 MF Constraint Qualification.

I give a formal statement and proof of KKT in Section 4.4. Before doing so, I need to discuss the technical condition called Constraint Qualification mentioned in Section 4.2.

To see that *some* additional condition may be needed, consider the following example, in which the KKT condition does not hold at the solution.

*Example 8.* The domain is  $\mathbb{R}$ . Let  $f(x) = x$  and let  $g(x) = x^3$ . Then the constraint set is  $C = (-\infty, 0]$  and the solution is  $x^* = 0$ . The KKT condition fails. Since  $\nabla f(x^*) = 1$  while  $\nabla g(x^*) = 0$ , there is no  $\lambda \geq 0$  such that  $\nabla f(x^*) = \lambda \nabla g(x^*)$ .  $\square$

Here is a more subtle example of the KKT condition failing, in which the constraint gradients don't vanish.

*Example 9.* Let  $f(x_1, x_2) = x_1$ . Let the constraints be  $g_1(x_1, x_2) = x_1^2 + x_2$  and  $g_2(x_1, x_2) = x_1^2 - x_2$ . The boundaries of the constraints are parabolas in  $x_2$  in terms of  $x_1$ . The constraint set is again  $C = \{0\}$  and the solution is again  $\{0\}$ . Once again the KKT condition fails.  $\nabla f(x^*) = (1, 0)$  while  $\nabla g_1(x^*) = (0, 1)$  and  $\nabla g_2(x^*) = (0, -1)$ . There are no  $\lambda_1, \lambda_2 \geq 0$  such that  $\nabla f(x^*) = \lambda_1 \nabla g_1(x^*) + \lambda_2 \nabla g_2(x^*)$ .  $\square$

I rule out Example 8 and 9 by assuming the following condition, an example of *constraint qualification*. Here and below, assume that  $x^*$  is feasible and that  $g_k$  is differentiable at  $x^*$  for all  $k \in J$ .

**Definition 2.** Mangasarian-Fromovitz Constraint Qualification (MF) *holds at  $x^*$  iff there exists a  $v \in \mathbb{R}^N$  such that for all  $k \in J$ ,  $\nabla g_k(x^*) \cdot v < 0$ .*

MF was introduced in Mangasarian and Fromovitz (1967). It is arguably the most useful form of constraint qualification but it is not the only one and it is not the weakest. Section 4.8 surveys other forms of constraint qualification. In Section 4.5, I show that a sufficient condition for MF is that (a) the  $g_k$  are convex functions and (b) the constraint set  $C$  has a non-empty interior. In practice, these conditions are typically very easy to check.

In Example 8, MF fails because  $\nabla g(x^*) = 0$ , hence  $\nabla g(x^*) \cdot v = 0$  for any  $v$ . In Example 9, MF fails because  $\nabla g_1(x^*) = -\nabla g_2(x^*)$ , and hence for any  $v$  such that  $\nabla g_1(x^*) \cdot v < 0$ ,  $\nabla g_2(x^*) \cdot v > 0$ .

Whether MF holds can depend on the details of how the constraint set is described, rather than on the constraint set per se. Consider, for example, the following variant of Example 8.

*Example 10.* The domain is  $\mathbb{R}$ . Let  $f(x) = g(x) = x$ . Then the constraint set is  $C = (-\infty, 0]$  and the solution is  $x^* = 0$ , as in Example 8. Although this problem has the same objective function as in Example 8 and the same constraint set, the KKT condition holds at the solution here, whereas it did not in 8. Explicitly, here  $\nabla f(0) = \nabla g(0) = 1$  and so  $\nabla f(0) = \lambda \nabla g(0)$  with  $\lambda = 1$ , which is positive.  $\square$

#### 4.4 The Karush-Kuhn-Tucker Theorem (KKT).

Given a MAX problem, a *local solution* is a local maximum for the given objective function and constraint set. The main result of these notes is the following, which uses MF for constraint qualification. In their original formulations, Karush (1939) and Kuhn and Tucker (1951) used an even weaker form of constraint qualification. I discuss this briefly in Section 4.8.

**Theorem 2** (Karush-Kuhn-Tucker). *Let  $x^*$  be a local solution to a differentiable MAX in standard form. If  $J = \emptyset$ , then  $\nabla f(x^*) = 0$ . If  $J \neq \emptyset$  and MF holds at  $x^*$ , then for every  $k \in J$  there is a  $\lambda_k^* \geq 0$  such that*

$$\nabla f(x^*) = \sum_{k \in J} \lambda_k^* \nabla g_k(x^*).$$

*This equality is called the KKT condition and the  $\lambda_k$  are called the KKT multipliers.*

**Proof.** I claim that if  $x^*$  is feasible but the KKT condition does not hold then there is a  $v \in \mathbb{R}^N$  such that  $\nabla f(x^*) \cdot v > 0$  and, for any  $k \in J$ ,  $\nabla g_k(x^*) \cdot v < 0$ . Suppose for the moment that this claim is true.

Since  $f$  is differentiable,  $\nabla f(x^*) \cdot v$  equals the directional derivative of  $f$  at  $x^*$  in the direction  $v$ . Since  $\nabla f(x^*) \cdot v > 0$ , this implies that there is a  $\gamma_f > 0$  such that for all  $\gamma \in (0, \gamma_f)$ ,  $f(x^* + \gamma v) - f(x^*) > 0$ , hence  $f(x^* + \gamma v) > f(x^*)$ .

By a similar argument, since  $\nabla g_k(x^*) \cdot v < 0$  for any  $k \in J$ , there is a  $\gamma_k > 0$  such that for all  $\gamma \in (0, \gamma_k)$ ,  $g_k(x^* + \gamma v) - g_k(x^*) < 0$ , hence  $g_k(x^* + \gamma v) < 0$  (since “= 0”  $k \in J$ , hence  $g_k(x^*) = 0$ ).

Finally, for any  $k \notin J$ , since  $g_k(x^*) < 0$  and  $g_k$  is continuous (since it is differentiable), there is  $\gamma_k > 0$  such that for all  $\gamma \in (0, \gamma_k)$ ,  $g_k(x^* + \gamma v) < 0$ .

Let  $\bar{\gamma}$  be the minimum of  $\gamma_f$  and the  $\gamma_k$ ; note that  $\bar{\gamma} > 0$  since it is a minimum of a finite set of strictly positive numbers. Then for all  $\gamma \in (0, \bar{\gamma})$ ,

1.  $f(x^* + \gamma v) > f(x^*)$ ,
2.  $g_k(x^* + \gamma v) < 0$  for all  $k$  (if there are any constraints at all).

Thus, for all  $\gamma \in (0, \bar{\gamma})$ ,  $x^* + \gamma v$  is feasible and yields a higher value for the objective function. Therefore  $x^*$  is not a local solution.

It remains to show that the above claim is true.

1.  $J = \emptyset$ . In this case, the KKT condition failing means that  $\nabla f(x^*) \neq 0$ . Then, since  $\nabla f(x^*) \neq 0$ ,  $\nabla f(x^*) \cdot \nabla f(x^*) > 0$ , which proves the claim with  $v = \nabla f(x^*)$ . This is essentially the same argument as in the proof of Theorem 1.
2.  $J \neq \emptyset$ . As already discussed in Section 4.2, the KKT condition is equivalent to requiring that  $\nabla f(x^*) \in A = \{x \in \mathbb{R}^N : \exists \lambda_k \geq 0 \text{ s.t. } x = \sum_{k \in J} \lambda_k \nabla g_k(x^*)\}$ , where  $A$  is the cone positively spanned by the gradients of the binding constraints. I now argue by contraposition: if  $\nabla f(x^*) \notin A$  then there exists a  $v \in \mathbb{R}^N$  such that  $\nabla f(x^*) \cdot v > 0$  and for all  $k \in J$ ,  $\nabla g_k(x^*) \cdot v < 0$ . As already discussed, this will complete proof.

As proved in the notes on Cones, the cone  $A$  is closed and convex. By the Separating Hyperplane Theorem for Cones in the notes on Cones, there is a  $\hat{v} \in \mathbb{R}^N$  such that  $\nabla f(x^*) \cdot \hat{v} > 0$  and for all  $a \in A$ ,  $a \cdot \hat{v} \leq 0$ .<sup>2</sup>

By MF, there is a  $\tilde{v} \in \mathbb{R}^N$  such that  $\nabla g_k(x^*) \cdot \tilde{v} < 0$  for all  $k \in J$ . Take  $\theta \in (0, 1)$  and let  $v = \theta \tilde{v} + (1 - \theta) \hat{v}$ .

- $\nabla f(x^*) \cdot v = \theta \nabla f(x^*) \cdot \tilde{v} + (1 - \theta) \nabla f(x^*) \cdot \hat{v}$ , which is strictly positive for  $\theta$  small even if  $\nabla f(x^*) \cdot \tilde{v}$  is negative.
- $\nabla g_k(x^*) \cdot v = \theta \nabla g_k(x^*) \cdot \tilde{v} + (1 - \theta) \nabla g_k(x^*) \cdot \hat{v}$ , which is strictly negative for any  $\theta \in (0, 1)$ .

This completes the proof. The picture is as in Figure 2.

■

The KKT multipliers need not be unique.

*Example 11.*  $N = 1$ ,  $f(x) = x$ ,  $g_1(x) = x$  and  $g_2(x) = 2x$ . Then the constraint set is  $(-\infty, 0]$ , the solution is at  $x^* = 0$ , MF holds (take  $v = -1$ ), and the KKT condition is  $1 = \lambda_1 + 2\lambda_2$ . Any  $\lambda_1, \lambda_2 \geq 0$  that satisfy this expression will work. For example, both  $\lambda_1 = 1$ ,  $\lambda_2 = 0$  and  $\lambda_1 = 0$ ,  $\lambda_2 = 1/2$  are valid.  $\square$

The KKT multipliers are unique if the gradients of the binding constraints are linearly independent, a condition that I call LI. As I discuss in Section 4.8, LI implies MF.

The KKT Theorem is sometimes expressed in an alternate form, which I use in the notes on Convex Optimization.

**Theorem 3** (KKT Theorem - Alternate Form). *Let  $x^*$  be a local solution to a differentiable MAX in standard form. If MF holds at  $x^*$ , then for every  $k$  there is a  $\lambda_k^* \geq 0$  such that,*

$$\nabla f(x^*) = \sum_k \lambda_k^* \nabla g_k(x^*),$$

---

<sup>2</sup>The notes, A Basic Separation Theorem for Cones, provides a separation theorem that is slightly weaker but is adequate for this application and has a more self-contained proof.



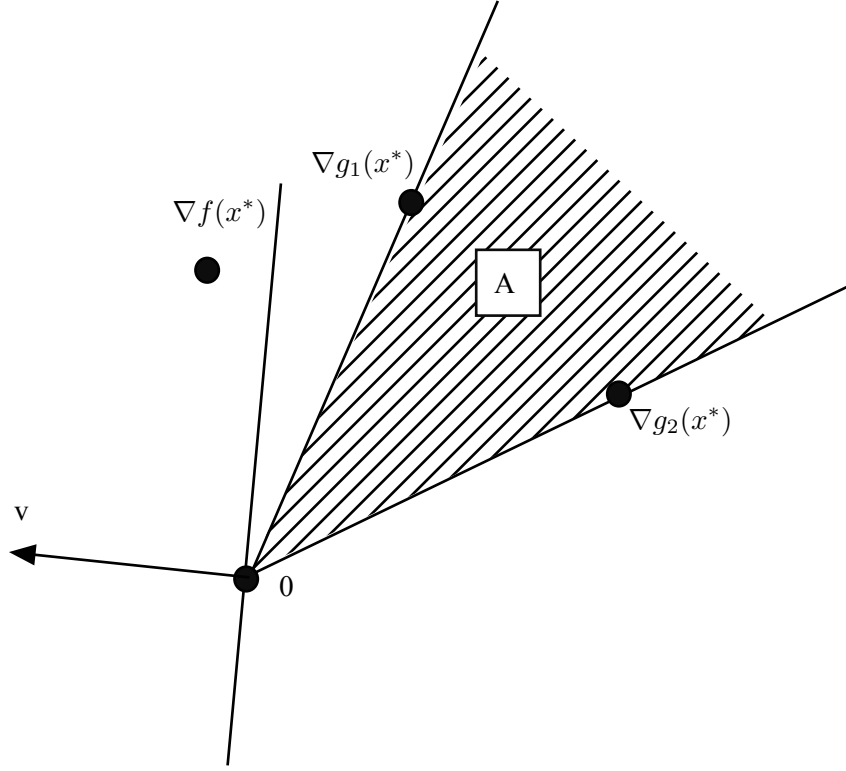


Figure 2: The separation argument behind KKT.

with  $\lambda_k^* = 0$  if  $k \notin J$ . In particular, if  $J = \emptyset$  then  $\nabla f(x^*) = 0$ .

**Proof.** This is just a reinterpretation of Theorem 2. ■

*Remark 1.* Thus, in Theorem 3, for every  $k$ ,  $\lambda_k^* = 0$  if  $g_k(x^*) < 0$  and  $g_k(x^*) = 0$  if  $\lambda_k^* > 0$ . This is called *complementary slackness* and can also be written  $\lambda_k^* g_k(x^*) = 0$ . □

#### 4.5 Checking MF: The Slater Condition.

Theorem 4 below shows that a sufficient condition for MF is that the constraint functions are convex and that the following condition holds.

**Definition 3.** A MAX problem in standard form satisfies the Slater Condition iff there is a point  $x$  such that  $g_k(x) < 0$  for all  $k$ .

The Slater Condition is equivalent to requiring that the constraint set  $C$  has a non-empty interior. For the modified MAX problem of Section 4.6, in which equality

constraints are added to the problem, the Slater Condition is equivalent to requiring that  $C$  have a non-empty *relative* interior (relative to the subset of  $\mathbb{R}^N$  defined by the equality constraints).

Informally, Slater is a non-local version of MF. MF implies Slater: MF implies that there are points arbitrarily near  $x^*$  that are interior to the constraint set, which implies, in particular, that the constraint set has a non-empty interior. The converse, that Slater implies MF, is not true in general but is true with additional conditions.

**Theorem 4.** *In a differentiable MAX problem, at any feasible  $x^*$ , Slater is equivalent to MF if either*

1. *each binding constraint function  $g_k$  is convex,*
2. *each binding constraint function  $g_k$  is quasi-convex and  $\nabla g_k(x^*) \neq 0$ .*

*An analogous result holds for MIN problems, with “concave” exchanged for “convex.”*

**Proof.** As already noted above, if MF holds then Slater holds.

Conversely, MF holds vacuously if no constraints are binding. Suppose then that at least one constraint is binding. Let  $x$  be as in the Slater condition and let  $v = x - x^*$ . I claim that for any  $k \in J$ ,  $\nabla g_k(x^*) \cdot v < 0$ . Consider any  $k \in J$ .

1. Suppose that  $g_k$  is convex. Then  $g_k(x) \geq \nabla g_k(x^*) \cdot (x - x^*) + g_k(x^*)$ . Since  $g_k(x) < 0$  (by assumption) and  $g_k(x^*) = 0$  (since  $k \in J$ ), and since  $v = x - x^*$ , this implies  $\nabla g_k(x^*) \cdot v < 0$ , as was to be shown.
2. Suppose that  $g_k$  is quasi-convex and that  $\nabla g_k(x^*) \neq 0$ . By continuity, since  $g_k(x) < 0$ , there is an  $\varepsilon > 0$  such that for all  $w$  on the unit sphere in  $\mathbb{R}^N$ ,  $g_k(x + \varepsilon w) < 0$ . Since  $g_k(x^*) = 0$  (since  $k \in J$ ), it follows that  $g_k(x + \varepsilon w) < g_k(x^*)$ . Hence for any  $\theta \in (0, 1)$ , by quasi-convexity,  $g_k(\theta(x + \varepsilon w) + (1 - \theta)x^*) \leq g_k(x^*)$ . Rewriting,  $g_k(x^* + \theta(v + \varepsilon w - x^*)) \leq g_k(x^*)$ , or

$$g_k(x^* + \theta(v + \varepsilon w)) - g_k(x^*) \leq 0.$$

Dividing by  $\theta$  and taking the limit as  $\theta \downarrow 0$  gives the directional derivative of  $g_k$  at  $x^*$  in the direction  $v + \varepsilon w$ . Since  $g_k$  is differentiable, this directional derivative is equal to  $\nabla g_k(x^*) \cdot (v + \varepsilon w)$ , hence  $\nabla g_k(x^*) \cdot (v + \varepsilon w) \leq 0$ , or

$$\nabla g_k(x^*) \cdot v + \varepsilon \nabla g_k(x^*) \cdot w \leq 0.$$

This holds for any  $w$  on the unit sphere. Since  $\nabla g_k(x^*) \neq 0$ , there is a  $w$  for which  $\nabla g_k(x^*) \cdot w > 0$ , which implies  $\nabla g_k(x^*) \cdot v < 0$

■

In practice, in economic applications, *all* constraints are often convex in MAX problems (or concave in MIN problems) and so checking MF constraint qualification boils down to checking Slater, which is often trivial.

In the following example, the constraint function is convex but Slater does not hold and MF fails.

*Example 12.* The domain is  $\mathbb{R}$ .  $f(x) = x$ .  $g(x) = x^2$ , which is convex. The constraint set is  $C = \{0\}$ , which has an empty interior. The solution is (trivially)  $x^* = 0$ . MF fails since  $\nabla g(0) = 0$  and hence there is no  $v \in \mathbb{R}$  such that  $\nabla g(0) \cdot v < 0$ .  $\square$

On the other hand, in the next example, Slater holds but the constraint function condition fails and MF fails, again because the gradient of the constraint function vanishes at  $x^*$ .

*Example 13.* Recall Example 8 in Section 4.3, with domain  $\mathbb{R}$ ,  $f(x) = x$ ,  $g(x) = x^3$ , and  $x^* = 0$ . MF fails. In this example, Slater holds (take  $x = -1$ ) but the constraint violates both of the two conditions in Theorem 4;  $g$  is not convex in the constraint region; it *is* quasi-convex but  $\nabla g(0) = 0$  and so, once again, there is no  $v \in \mathbb{R}$  such that  $\nabla g(0) \cdot v < 0$ .  $\square$

Finally, here is a more elaborate example, a variation on Example 9 in in Section 4.3, in which Slater holds, the gradients of the constraint functions do not vanish at  $x^*$ , but MF still fails.

*Example 14.* The domain is  $\mathbb{R}^2$ . Let  $f(x_1, x_2) = x_1$ . Let  $g_1(x_1, x_2) = x_1^2(x_1 + 1) - x_2$ ,  $g_2(x_1, x_2) = -x_1^2(x_1 + 1) + x_2$ . The constraint set is the union of the origin and points  $(x_1, x_2)$  with  $x_1 \leq -1$  and  $x_2 \in [x_1^2(x_1 + 1), -x_1^2(x_1 + 1)]$ . The solution is  $x^* = (0, 0)$ . But  $\nabla f(0, 0) = (1, 0)$  while  $\nabla g_1(0, 0) = (0, -1)$  and  $\nabla g_2(0, 0) = (0, 1)$  and hence there is no  $\lambda_1, \lambda_2$  such that  $\nabla f(0, 0) = \lambda_1 \nabla g_1(0, 0) + \lambda_2 \nabla g_2(0, 0)$ . In this example, neither constraint function is quasi-convex.  $\square$

## 4.6 Equality constraints.

Economists sometimes want to consider constraints that must hold with equality. A standard economic example is the budget constraint  $p \cdot x = m$  or  $p \cdot x - m = 0$ . Requiring  $p \cdot x = m$  is equivalent to requiring that both  $p \cdot x \leq m$  and  $p \cdot x \geq m$  hold simultaneously. There is thus a sense in which equality constraints are just special cases of inequality constraints, and accordingly equality constraints ought to fit within the KKT framework. Without belaboring the issue, a minor modification of the KKT theorem (which I will not prove or even state formally) says that if the equality constraints are given by  $r_\ell(x) = 0$  then the first order conditions for a local minimum are that there are multipliers  $\lambda_k^* \geq 0$  and  $\mu_\ell^* \in \mathbb{R}$  such that,

$$\nabla f(x^*) = \sum_{k \in J} \lambda_k^* \nabla g_k(x^*) + \sum_{\ell} \mu_\ell^* \nabla r_\ell(x^*).$$

Note that the  $\mu_\ell^*$ , the multipliers on the equality constraints, could be either positive or negative.

What economists model as equality constraints are often just binding inequality constraints. For example, in the case of utility maximization, the constraint  $p \cdot x = m$  is really  $p \cdot x \leq m$ . The consumer is physically allowed to set  $p \cdot x < m$  (i.e., to spend less than her income) but it is not optimal to do so (the constraint is binding at the solution) under standard assumptions. It is accordingly understood that the sign of the KKT multiplier on the budget constraint cannot be negative. I will not dwell further on equality constraints.

#### 4.7 Non-negativity constraints.

It is common in economic applications to require that  $x \geq 0$ . This generates  $N$  constraints of the form:

$$-x_n \leq 0.$$

From the KKT condition,

$$\nabla f(x^*) = \sum_{k \in J} \lambda_k^* \nabla g_k(x^*).$$

Suppose that the non-negativity constraint for  $x_n$  is constraint  $k$ , and that, moreover, this constraint is binding. If  $\lambda_k^* > 0$ , then  $\lambda_k^* \nabla g_k(x^*) = (0, \dots, 0, -\lambda_k^*, 0, \dots, 0)$ , with  $-\lambda_k^*$  in the  $n$ th place, so this constraint is *lowering* the right-hand side of the above equality. Let  $I$  be the set of binding constraints *other than* the non-negativity constraints. Then

$$\nabla f(x^*) \leq \sum_{k \in I} \lambda_k \nabla g_k(x^*).$$

I mention this because many authors do not treat the conditions  $x_n \geq 0$  as explicit constraints. Because of this, these authors state the KKT condition as an inequality. I think it is easier, and safer, to remember the KKT condition as an equality, with all binding constraints explicit.

#### 4.8 Other forms of constraint qualification.

MF is just one of many forms of constraint qualification, any of which suffice to imply that the KKT condition holds at a local solution. I will not be exhaustive but here are some of the more important alternatives.

Let  $S$  be the set of the gradients of the binding constraints,  $S = \{\nabla g_k(x^*)\}_{k \in J}$ . Recall (from the notes on Cones) that a cone  $A$  is *pointed* iff for any  $a \in A$ ,  $a \neq 0$ ,  $-a \notin A$ . Here are three other forms of constraint qualification.

**Definition 4.** Linearly independent constraint qualification (LI) holds at  $x^*$  iff  $S$  is linearly independent: if  $\sum_{k \in J} \lambda_k \nabla g_k(x^*) = 0$  then  $\lambda_k = 0$  for all  $k \in J$ . LI holds vacuously if  $J = \emptyset$ .

**Definition 5.** Positive linearly independent constraint qualification (PI) *holds at*  $x^*$  *iff*  $S$  *is positive linearly independent: if*  $\sum_{k \in J} \lambda_k \nabla g_k(x^*) = 0$  *and*  $\lambda_k \geq 0$  *for all*  $k \in J$  *then*  $\lambda_k = 0$  *for all*  $k \in J$ . *PI holds vacuously if*  $J = \emptyset$ .

**Definition 6.** Pointed cone constraint qualification (PC) *holds at*  $x^*$  *iff*  $\nabla g_k(x^*) \neq 0$  *for all*  $k \in J$  *and*  $A$  *is pointed. PC holds vacuously if*  $J = \emptyset$ .

It is, I hope, obvious that LI implies PI. Here are examples showing that the converse is not true in general.

*Example 15.*  $N = 2$  and there are three binding constraints, with  $\nabla g_1(x^*) = (1, 0)$ ,  $\nabla g_2(x^*) = (0, 1)$ , and  $\nabla g_3(x^*) = (1, 1)$ . The cone  $A$  is the non-negative orthant. LI fails here, since  $\nabla g_1(x^*) + \nabla g_2(x^*) - \nabla g_3(x^*) = 0$ , but PI can be shown to hold (this will follow from Theorem 5 below). Note here that the cone  $A$  is positively spanned by the independent vectors  $\nabla g_1(x^*)$  and  $\nabla g_2(x^*)$ ;  $\nabla g_3(x^*)$  is, in a sense, redundant.  $\square$

*Example 16.*  $N = 3$  and there are four binding constraints, with

- $\nabla g_1(x^*) = (1, 0, 0)$ ,
- $\nabla g_2(x^*) = (0, 1, 0)$ ,
- $\nabla g_3(x^*) = (1, 0, 1)$ ,
- $\nabla g_4(x^*) = (0, 1, 1)$ .

The cone  $A$  has a square cross-section. LI fails here, since  $-\nabla g_1(x^*) + \nabla g_2(x^*) + \nabla g_3(x^*) - \nabla g_4(x^*) = 0$ , but PI can be shown to hold (again, this will follow from Theorem 5 below). Note that all four binding constraint gradients are needed to positively span  $A$ , even though the gradients are not independent. No gradient is redundant in the same sense as in Example 15.  $\square$

On the other hand, the three conditions MF, PI, and PC are all equivalent. Thus, if MF holds then the cone  $A$  is pointed, as drawn in Figure 2.

**Theorem 5.** *MF, PI, PC are equivalent.*

**Proof.** If  $J = \emptyset$  then the claim holds vacuously. Assume, therefore, that  $J \neq \emptyset$ .

- $\text{MF} \Rightarrow \text{PI}$ . By contraposition. Suppose that PI fails. Then there are  $\lambda_k \geq 0$ , with at least one  $\lambda_k > 0$ , such that  $\sum_{k \in J} \lambda_k \nabla g_k(x^*) = 0$ . Then for any  $v \in \mathbb{R}^N$ ,  $0 = 0 \cdot v = (\sum_{k \in J} \lambda_k \nabla g_k(x^*)) \cdot v = \sum_{k \in J} \lambda_k (\nabla g_k(x^*) \cdot v)$ , which implies that MF cannot hold.
- $\text{PI} \Rightarrow \text{PC}$ . By contraposition. If  $\nabla g_k(x^*) = 0$  for any  $k \in J$  then PI fails (take  $\lambda_k > 0$  for this  $k$  and zero otherwise). On the other hand, suppose that  $A$  is not pointed. Then there is a  $a \in A$  such that  $a \neq 0$  and  $-a \in A$ . By definition of  $A$ , there are weights  $\lambda_k \geq 0$  such that  $a = \sum_{k \in J} \lambda_k \nabla g_k(x^*)$ , and

weights  $\hat{\lambda}_k \geq 0$  such that  $-a = \sum_{k \in J} \hat{\lambda}_k \nabla g_k(x^*)$ . But then  $0 = a - a = \sum_{k \in J} (\lambda_k + \hat{\lambda}_k) \nabla g_k(x^*)$ . Since  $a \neq 0$ ,  $\lambda_k + \hat{\lambda}_k \neq 0$  for at least one  $k$ , hence PI fails.

- PC  $\Rightarrow$  MF. It is almost immediate that  $A$  is a convex cone. Because  $A$  is finitely generated, it is closed; see the notes on Cones. By assumption,  $A$  is pointed. Then the Supporting Hyperplane Theorem for Pointed Cones in the notes on Cones implies that there is a  $v$  such that  $a \cdot v < 0$  for all  $a \in A$ ,  $a \neq 0$ . By assumption,  $\nabla g_k(x^*) \neq 0$  for every  $k \in J$ , and the result follows (take  $a = \nabla g_k(x^*)$ ).

■

Taking stock, LI is stronger than PI and hence stronger than MF. Theorem 4 is false for LI (Slater plus convexity of the constraints does not imply LI) and, as a consequence, LI is potentially more difficult to verify, which is the main reason why I have emphasized MF rather than LI. But LI has some advantages. First, if LI holds then the KKT multipliers not only exist but are unique. In this context, look back at Example 11, where MF holds but LI fails and the multipliers are not unique. Second, if LI holds then there is an existence proof for KKT multipliers that is elementary in the sense of not relying on a separation argument. I combine both of these observations in the next theorem and its proof.

**Theorem 6.** *Let  $x^*$  be a local solution to a differentiable MAX in standard form. If  $J = \emptyset$ , then  $\nabla f(x^*) = 0$ . If  $J \neq \emptyset$  and LI holds at  $x^*$ , then for every  $k \in J$  there are unique  $\lambda_k^* \geq 0$  such that*

$$\nabla f(x^*) = \sum_{k \in J} \lambda_k^* \nabla g_k(x^*).$$

**Proof.** Since LI implies MF, existence of the KKT multipliers are given by Theorem 2. Uniqueness is guaranteed by a standard linear algebra argument.

As noted above, if LI holds then one can prove existence of the KKT multipliers without invoking a separation argument. The following proof is attributed in Kreps (2012) to Ben Porath.

Suppose that  $J \neq \emptyset$ . For ease of notation (and this is only for ease of notation), suppose that all  $K$  constraints are binding at  $x^*$ .

Suppose first that  $\nabla f(x^*)$  is not in the vector space spanned by the  $\nabla g_k(x^*)$ . LI then implies that the set  $\{\nabla f(x^*), \nabla g_1(x^*), \dots, \nabla g_K(x^*)\}$  is independent, in which case the matrix, call it  $B$ , formed by using these  $K + 1$  vectors as rows, has rank  $K + 1$ . Then there is a  $v \in \mathbb{R}^N$  such that  $Bv = (1, -1, \dots, -1) \in \mathbb{R}^{K+1}$ , hence, in particular,  $\nabla f(x^*) \cdot v > 0$  and  $\nabla g_k(x^*) \cdot v < 0$  for all  $k$ . As in the above proof of KKT, this implies that  $x^*$  is not a local maximum. By contraposition, the above set cannot be independent, and hence there exist  $\lambda_k^*$  (not necessarily non-negative) such that  $\nabla f(x^*) = \sum_k \lambda_k^* \nabla g_k(x^*)$ .

It remains to show that  $\lambda_k^* \geq 0$ . Suppose that  $\nabla f(x^*) = \sum_k \lambda_k^* \nabla g_k(x^*)$  but with some  $\lambda_k^* < 0$ . For ease of notation, suppose  $\lambda_1^* < 0$ . Choose  $M < 0$ , hence  $M\lambda_1^* > 0$ , such that  $M\lambda_1^* - \sum_{k>1} \lambda_k^* > 0$ . Let  $\hat{B}$  be the matrix formed by the using the  $K$  constraint gradients as rows. Then by LI,  $\hat{B}$  has rank  $K$  and hence there is a  $v$  such that  $\hat{B}v = (M, -1, \dots, -1) \in \mathbb{R}^K$ . Since  $\nabla f(x^*) = \sum_k \lambda_k^* \nabla g_k(x^*)$ ,  $\nabla f(x^*) \cdot v = \sum_k \lambda_k^* \nabla g_k(x^*) \cdot v = M\lambda_1^* - \sum_{k>1} \lambda_k^* > 0$ . On the other hand,  $\nabla g_k(x^*) \cdot v < 0$  for all  $k$ . Once again, this implies that  $x^*$  is not a local maximum. Hence, by contraposition,  $\lambda_k^* \geq 0$  for all  $k$ .  $\square$  ■

*Remark 2.* Continuing in the spirit of the proof of Theorem 6, suppose that, as in Example 15, one can find a subset of the binding constraint gradients that (a) positively spans  $A$  and (b) is independent. In this case, one can still use the linear algebra argument in the proof of Theorem 6 even if LI is violated: work just with an independent set of binding constraints that positively spans  $A$  and set  $\lambda_k^* = 0$  for all of the other binding constraint gradients. But if, as in Example 16, there is no subset of binding constraint gradients that (a) spans  $A$  and (b) is independent, then this proof of KKT no longer applies.  $\square$

In practice, examples in which MF fails, or even where LI fails, are extremely rare. For completeness, however, I discuss briefly here weaker conditions for existence of KKT multipliers.

The set  $A$  formed by the gradients of the binding constraints is always a closed, convex cone. Even without constraint qualification, the Separating Hyperplane Theorem for Cones in the notes on Cones implies that there is a vector  $v \neq 0$  such that  $\nabla f(x^*) \cdot v > 0$  while  $\nabla g_k(x^*) \cdot v \leq 0$  for all binding  $k$ . Constraint qualification was introduced to guarantee that there is a  $v$  such that, in fact,  $\nabla g_k(x^*) \cdot v < 0$ , which ensures that for small  $\gamma > 0$ ,  $g_k(x^* + \gamma v) < 0$ , so that  $x^* + \gamma v$  is feasible. But this is sometimes stronger than necessary.

If the binding constraints are *concave*, which includes linear constraints as a special case, then  $x^* + \gamma v$  is feasible when  $\nabla g_k(x^*) \cdot v = 0$ : by concavity,  $g_k(x^* + \gamma v) \leq \nabla g_k(x^*) \cdot (x^* + \gamma v - x^*) + g(x^*)$ , hence  $g(x^* + \gamma v) \leq 0$  since  $Dg_k(x^*)v = \nabla g_k(x^*) \cdot v = 0$  and  $g(x^*) = 0$ .

*Example 17.* The domain is  $\mathbb{R}_+^2$ .  $f(x) = x_1 x_2$ .  $g_1(x) = x_1 + x_2 - 2$ .  $g_2(x) = 2 - x_1 - x_2$ . The constraint set is the diagonal line segment  $C = \{(x_1, x_2) \in \mathbb{R}_+^2 : x_1 + x_2 = 2\}$ . The solution is  $x^* = (1, 1)$ . MF fails since PC fails:  $A = \{x \in \mathbb{R}^2 : x_1 = x_2\}$ , which is not pointed. But KKT multipliers exist.  $\nabla f(x^*) = \nabla g_1(x^*) = (1, 1)$ .  $\nabla g_2(x^*) = (-1, 1)$ . So take  $\lambda_1 = 1$ ,  $\lambda_2 = 0$ . As one would expect since LI fails, these multipliers are not unique.  $\lambda_1 = 0$ ,  $\lambda_2 = -1$  also works.  $\square$

But in the general case, if  $\nabla g_k(x^*) \cdot v = 0$ , then  $x^* + \gamma v$  need not be feasible even for small  $\gamma$ . Essentially the same proof will, however, continue to work if I can find feasible points arbitrarily near  $x^*$  in (approximately) the direction  $v$ . Example 9 in Section 4.3 illustrates one difficulty that can arise. In this example,

at  $x^* = 0$ ,  $\nabla f(x^*) = (1, 0)$  and  $A$  is the  $x_2$  axis. A  $v$  that separates  $\nabla f(x^*)$  from  $A$  must lie along the  $x_1$  axis. Take  $v = (1, 0)$ . For each constraint  $k$ , there is a sequence of feasible points  $\{x_t\}$  that approximates  $x^* + \gamma v$  in the sense that  $(x_t - x^*)/\|x_t - x^*\| \rightarrow v$ . For example, for  $k = 1$ , take  $x_t = (1/t, -1/t^2)$ . But the order of quantifiers matters: one needs a different sequence for each constraint. In particular, the sequence just given for  $g_1$  violates  $g_2$  and hence the sequence is not feasible.

Both [Karush \(1939\)](#) and [Kuhn and Tucker \(1951\)](#) assume that, for any  $v$  such that  $\nabla g_k(x^*) \cdot v \leq 0$  for all  $k \in J$ , there is a sequences of feasible points that approximates  $x^* + \gamma v$  in the sense just discussed. Informally, they assume that any direction  $v$  that should be feasible, given the derivatives of the binding  $g_k$ , actually is feasible. This form of constraint qualification is strictly weaker than MF but it is difficult to check directly and examples in which the difference matters tend to be artificial.

#### 4.9 Binding constraints, active constraints, and slack constraints.

Recall that a constraint  $k$  is binding if  $g_k(x^*) = 0$ . It is *slack* otherwise. This terminology is somewhat misleading.

First, a constraint can be binding but irrelevant.

*Example 18.* Let  $f(x) = -x^2$  and let the constraint be  $x \geq 0$ ; thus  $g(x) = -x$ . The solution is at  $x^* = 0$ . The constraint is binding but  $\lambda^* = 0$  because  $\nabla f(x^*) = 0$ . Relaxing the constraint does not affect the solution.  $\square$

In [Example 18](#),  $\lambda^* = 0$  even though the constraint is binding. Call a binding constraint  $k$  *active* if  $\lambda_k^* > 0$ .<sup>3</sup> Although [Example 18](#) provides a counter example, binding constraints will typically be active.

Second, slack constraints can affect the global solution, as the next example illustrates.

*Example 19.* Let the constraints be  $x \geq -1$  and  $x \leq 1$ , hence  $g_1(x) = -x - 1$  and  $g_2(x) = x - 1$ . Let

$$f(x) = -x^2 + x^4.$$

The graph of  $f$  looks like a W. There are constrained maxima at -1, 0, and 1.

At  $x^* = 0$ ,  $g_1(x^*) < 0$  and  $g_2(x^*) < 0$ , so the constraints are not binding. Nevertheless, the restriction to the constraint set matters. If either constraint were relaxed then the objective function would increase. For example, if the constraint  $x \leq 1$  is changed to  $x \leq 3$  then the unique constrained maximum is  $x^* = 3$ . In particular, 0 is no longer a maximum.

The underlying issue here is that KKT is a result about local, rather than global, maximization. Even if the constraint is relaxed,  $x^* = 0$  remains a *local* maximum. The fact that constraints are slack at  $x^* = 0$  correctly reflects this.  $\square$

---

<sup>3</sup>Many references use “active” to mean “binding;” here I am taking active to be more restrictive than binding.



The problem illustrated by Example 19 does not occur if the objective function is concave, in which case local maxima are global maxima.

## 5 Using KKT.

The bottom line here is bad news: there is no simple procedure for finding points  $x^*$  and multipliers  $\lambda^*$  that satisfy KKT.

A systematic procedure would be to try to solve the unconstrained problem. If you can do so without violating a constraint, then you are done. If not, pick a constraint and look for solutions to the problem in which this one constraint is binding. For example, if the binding constraint happens to be labeled 1, then we get,

$$\begin{aligned}\nabla f(x^*) &= \lambda_1^* \nabla g_1(x^*) \\ g_1(x^*) &= 0.\end{aligned}$$

This gives  $n + 1$  equations (not necessarily linear) in  $n + 1$  unknowns, namely the  $x^*$  and  $\lambda_1^*$ . With good fortune, you may be able to solve it analytically. Having gone through all  $K$  constraints one by one, if you find solutions that do not violate other constraints, choose the one that maximizes the value of  $f$  (there may be more than one). If, on the other hand, there are no solutions within the constraint set  $C$ , then start looking at the constraints two at a time. And so on. You may be able to cut down on the tedium if you can be clever and figure out which constraints are likely to be binding.

Another useful fact to remember when doing maximization problems is the following. In the theorem below,  $C$  could be any set at all, not necessarily a subset of  $\mathbb{R}^N$ .

**Theorem 7.** *Consider any set  $C$ , any function  $f : X \rightarrow \mathbb{R}$ . Let  $h : \mathbb{R} \rightarrow \mathbb{R}$  be any strictly increasing function. Then  $x^*$  maximizes  $f$  on  $C$  iff it maximizes  $\hat{f} = h \circ f$  on  $C$ .*

**Proof.** Suppose  $f(x^*) \geq f(x)$  for any  $x$  in  $C$ . Then, since  $h$  is strictly increasing,  $h(f(x^*)) \geq h(f(x))$ . And conversely. ■

One can sometimes simplify calculations enormously by a clever choice of  $h$ . Note that if you modify the objective function in this way, while you won't change the solution, you will, typically, change the KKT multipliers.

In the following examples, the constraints are linear, and hence, as discussed in Section 4.8, constraint qualification holds.

*Example 20.* Consider the problem

$$\begin{aligned} \max \quad & f(x) = x_1^{1/2} x_2^{1/3} x_3^{1/6} \\ \text{s.t.} \quad & 4x_1 + 8x_2 + 3x_3 \leq 9 \\ & x \geq 0 \end{aligned}$$

This could, for example, be a utility maximization problem with utility function  $f$  and budget constraint given by prices  $p = (4, 8, 3)$  and income 9.

I need to translate this into standard form:

$$\begin{aligned} \max \quad & f(x) = x_1^{1/2} x_2^{1/3} x_3^{1/6} \\ \text{s.t.} \quad & 4x_1 + 8x_2 + 3x_3 - 9 \leq 0 \\ & -x \leq 0 \end{aligned}$$

All constraint functions are linear (hence convex) and Slater holds (set  $x_1 = 3/8$ ,  $x_2 = 3/16$ , and  $x_3 = 1/2$ , for example). Therefore, by Theorem 4, MF constraint qualification holds and hence any solution must satisfy the KKT condition. The check on constraint qualification in the other examples in this section is so similar that I won't bother to make those checks explicit.

Next, note that any solution  $x^*$  will be strictly positive. As already noted, for example, it is feasible to take  $x_1 = 3/8$ ,  $x_2 = 3/16$ ,  $x_3 = 1/2$ . This yields a positive value for the objective function, whereas the objective function is 0 if any  $x_n = 0$ . Thus, the non-negativity constraints do not bind.

On the other hand, at any solution, the first constraint must bind: if  $4x_1 + 8x_2 + 3x_3 - 9 < 0$ , then one could increase the objective function which is increasing in all of its arguments, by increasing one or all of the  $x_n$ , if only a little, without violating the constraint.

Since at an optimum only the first constraint binds, the KKT condition is

$$\begin{bmatrix} \frac{\gamma}{2x_1^*} \\ \frac{\gamma}{3x_2^*} \\ \frac{\gamma}{6x_3^*} \end{bmatrix} = \lambda_1^* \begin{bmatrix} 4 \\ 8 \\ 3 \end{bmatrix},$$

where

$$\gamma = x_1^{*1/2} x_2^{*1/3} x_3^{*1/6}$$

Hence

$$\begin{bmatrix} \frac{\gamma}{2\lambda_1^*} \\ \frac{\gamma}{3\lambda_1^*} \\ \frac{\gamma}{6\lambda_1^*} \end{bmatrix} = \begin{bmatrix} 4x_1^* \\ 8x_2^* \\ 3x_3^* \end{bmatrix}.$$

Substituting this into the binding constraint,  $4x_1^* + 8x_2^* + 3x_3^* - 9 = 0$ , yields,

$$\frac{\gamma}{2\lambda_1^*} + \frac{\gamma}{3\lambda_1^*} + \frac{\gamma}{6\lambda_1^*} = 9,$$

or

$$\frac{\gamma}{\lambda_1^*} = 9.$$

Substituting this back into the KKT condition yields,

$$\begin{bmatrix} \frac{9}{2} \\ 3 \\ \frac{3}{2} \end{bmatrix} = \begin{bmatrix} 4x_1^* \\ 8x_2^* \\ 3x_3^* \end{bmatrix}$$

or

$$x^* = \left( \frac{9}{8}, \frac{3}{8}, \frac{1}{2} \right) \gg 0,$$

while

$$\lambda_1^* = \frac{1}{48} 3^{1/3} 2^{3/4}.$$

I finish with two remarks. First, note that I have not actually shown that  $x^*$  is a solution. This will follow from Part II of these notes, which give sufficient conditions.

Second, I could have made the calculations tidier by working with the log of the objective function. By Theorem 7, this yields a new maximization problem with the same solution as the original one (as you can verify by direct calculation):

$$\begin{aligned} \max \quad & \hat{f}(x) = \frac{1}{2} \ln(x_1) + \frac{1}{3} \ln(x_2) + \frac{1}{6} \ln(x_3) \\ \text{s.t.} \quad & 4x_1 + 8x_2 + 3x_3 - 9 \leq 0 \\ & -x \leq 0 \end{aligned}$$

Strictly speaking,  $\ln(f(x))$  is not defined if any  $x_n = 0$ . But I have just argued that no such  $x$  can be a solution. This sort of technicality is common in economics optimization problems and typically are simply ignored.  $\square$

*Example 21.* Consider the same problem as in Example 20 but now suppose that I had guessed that

$$\tilde{x} = \left( \frac{3}{4}, \frac{3}{8}, 1 \right),$$

which is not actually a solution but which, as I noted in the course of Example 20, is feasible. Since constraint 1 binds at  $\tilde{x}$ , the KKT condition requires,

$$\nabla f(\tilde{x}) = \lambda_1 \nabla g_1(\tilde{x}),$$

or (approximately and after some tedious calculation),

$$\begin{bmatrix} 0.41 \\ 0.56 \\ 0.10 \end{bmatrix} = \lambda_1 \begin{bmatrix} 4 \\ 8 \\ 3 \end{bmatrix}.$$

There is no  $\lambda_1$  that will work. For example, to satisfy the first line,  $\lambda_1$  would have to be about  $1/10$ , but to satisfy the third it would have to be about  $1/30$ . Geometrically, we are detecting the fact that the vectors  $\nabla f(\tilde{x})$  and  $\nabla g(\tilde{x})$  point in different directions, indicating that  $\tilde{x}$  is not a solution. And, in fact, you can compute that  $f(\tilde{x}) \approx 0.62$  whereas  $f(x^*) \approx 0.68$ .  $\square$

*Example 22.* Once again consider the same problem as in Example 20 but now suppose that I had guessed that

$$\hat{x} = \left( \frac{1}{4}, \frac{1}{8}, \frac{1}{3} \right).$$

This is not a solution but it is feasible. Since *no* constraints bind at  $\hat{x}$ , the KKT condition requires

$$\nabla f(\hat{x}) = 0,$$

which direct calculation verifies is not true. You can compute that  $f(\hat{x}) \approx 0.32$  whereas, again,  $f(x^*) \approx 0.68$ .  $\square$

In the next example, more than one constraint binds at the solution.

*Example 23.* Consider

$$\begin{aligned} \max \quad & \sqrt{x_1 + 1} + 2\sqrt{x_2 + 1} + 3\sqrt{x_3 + 1} \\ \text{s.t.} \quad & 4x_1 + 8x_2 + 3x_3 \leq 9 \\ & x \geq 0 \end{aligned}$$

Translating this into standard form,

$$\begin{aligned} \max \quad & \sqrt{x_1 + 1} + 2\sqrt{x_2 + 1} + 3\sqrt{x_3 + 1} \\ \text{s.t.} \quad & 4x_1 + 8x_2 + 3x_3 - 9 \leq 0 \\ & -x \leq 0 \end{aligned}$$

The objective function is strictly increasing in all arguments so I again conclude that the first constraint will be binding at the solution. It is no longer obvious, however, whether the non-negativity constraints bind.

Let's guess (correctly as it turns out) that at the solution only  $x_3$  is positive. This is intuitive, since  $x_3$  gets the most weight in the objective function but the least weight in the first constraint. Don't place too much confidence in this sort of intuition, however; we're just making an educated guess and we could have been wrong.

If  $x_1^* = x_2^* = 0$  then the binding constraint  $4x_1 + 8x_2 + 3x_3 = 9$  implies  $x_3^* = 3$ .

I have,

$$\begin{aligned}\nabla f(x^*) &= (1/2, 1, 3/4) \\ \nabla g_1(x^*) &= (4, 8, 3) \\ \nabla g_2(x^*) &= (-1, 0, 0) \\ \nabla g_3(x^*) &= (0, -1, 0)\end{aligned}$$

From the KKT condition, I get

$$\begin{bmatrix} 1/2 \\ 1 \\ 3/4 \end{bmatrix} = \begin{bmatrix} 4 & -1 & 0 \\ 8 & 0 & -1 \\ 3 & 0 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1^* \\ \lambda_2^* \\ \lambda_3^* \end{bmatrix},$$

which can be solved to yield,

$$\begin{bmatrix} \lambda_1^* \\ \lambda_2^* \\ \lambda_3^* \end{bmatrix} = \begin{bmatrix} 1/4 \\ 1/2 \\ 1 \end{bmatrix} > 0.$$

Thus, I can satisfy the KKT condition at  $x^* = (0, 0, 3)$ . Again, the sufficient conditions of Part II of these notes guarantee that  $x^*$  actually is the solution.  $\square$

*Example 24.* Consider the same problem as in Example 23. Suppose that I had instead guessed that only the first constraint was binding. Then the KKT condition is,

$$\begin{bmatrix} \frac{1}{2} \frac{1}{\sqrt{\tilde{x}_1+1}} \\ \frac{1}{\sqrt{\tilde{x}_2+1}} \\ \frac{3}{2} \frac{1}{\sqrt{\tilde{x}_1+1}} \end{bmatrix} = \lambda_1 \begin{bmatrix} 4 \\ 8 \\ 3 \end{bmatrix}.$$

Manipulating this yields

$$\tilde{x} = \begin{bmatrix} -3/5 \\ -3/5 \\ 27/5 \end{bmatrix}.$$

and

$$\lambda_1 = \frac{1}{16} \sqrt{10} > 0.$$

Notice that this “solution” satisfies the KKT condition. BUT, it is not feasible, since constraints 2 and 3 are violated.  $\square$

*Example 25.* Once again, consider the same problem as in Example 23. Suppose that I had instead guessed the solution to be  $\hat{x} = (9/4, 0, 0)$ , which is feasible. Working through a similar calculation to the one above, but with different binding constraints, yields

$$\begin{bmatrix} 1/\sqrt{13} \\ 1 \\ 3/2 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 0 \\ 8 & -1 & 0 \\ 3 & 0 & -1 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_3 \\ \lambda_4 \end{bmatrix}.$$

Solving this yields

$$\begin{bmatrix} \lambda_1 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} = \begin{bmatrix} \frac{1}{4\sqrt{13}} \\ \frac{2}{\sqrt{13}} - 1 \\ \frac{3}{4\sqrt{13}} - \frac{3}{2} \end{bmatrix} \approx \begin{bmatrix} 0.1 \\ -0.4 \\ -1.3 \end{bmatrix}.$$

The critical point to notice here is that  $\lambda_3, \lambda_4 < 0$ . KKT requires all multipliers on binding constraints to be non-negative. This means that  $\hat{x}$  has failed the KKT condition and therefore cannot be a maximum. If you had neglected to write the problem in standard form you might instead have found all the multipliers to be positive, and concluded (mistakenly) that  $x^*$  satisfied the KKT conditions. You can compute that  $f(\hat{x}) \approx 1.8$  whereas  $f(x^*) = 6$ .  $\square$

## References

- Karush, William. 1939. “Minima of Functions of Several Variables with Inequalities as Side Conditions.” Masters Thesis, University of Chicago.
- Kreps, David M. 2012. *Microeconomic Foundations I: Choice and Competitive Markets*. Princeton University Press.
- Kuhn, H. W. and A. W. Tucker. 1951. Nonlinear Programming. In *Proceedings of the 2nd Berkley Symposium*. University of California Press pp. 481–492.
- Mangasarian, O. L. and S. Fromovitz. 1967. “The Fritz John Necessary Optimality Conditions in the Presence of Equality and Inequality Constraints.” *Journal of Mathematical Analysis and Applications* 17:37–47.