Professor John Nachbar
Econ 4111
March 21, 2012

Finite Dimensional Optimization, Part I
The Karush-Kuhn-Tucker Theorem

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 is the same as the problem of maximizing $-f$, so all the results for maximization have easy corollaries for minimization.

Although some of the details are subtle, the basic intuition behind the main result, the Karush-Kuhn-Tucker Theorem (Theorem 3 in Section 4.2), is straightforward. At an unconstrained maximum, you are, metaphorically speaking, at the top of a hill. The gradient, which measures the direction of steepest ascent, is zero. At a constrained maximum with a single binding constraint, you are climbing the hill when you hit a fence, the binding constraint, that keeps you from climbing any further. The gradient is not zero. Instead, it points into the fence, indicating that you could climb higher if the fence weren't in the way. In fact, the gradient of the hill has to be at right angles to the fence. If this were not true then sliding along the fence in one direction or the other would move you a bit further uphill.

Mathematically, the hill is represented by a function, the objective function, and the fence is represented by the level set of another function, the constraint function. In the case of consumer optimization, for example, the objective function is the utility function and the constraint is given by the budget inequality $p \cdot x \leq m$, where p is a vector of prices and m is income. I can represent this constraint in the form $g(x) \leq 0$, where $g(x) = p \cdot x - m$. The gradient of the hill is at right angles to the fence if and only if the gradient of the objective function points in exactly the same direction as the gradient of g. (In consumer problems, there are typically also non-negativity constraints on the x_n. I am ignoring these for this introductory discussion. Example 20 in Section 6 provides an example in which the non-negativity constraints matter.)

The Karush-Kuhn-Tucker (KKT) Theorem generalizes this intuition to two or more binding constraints.[1] The details can be a bit fussy (we have to worry about the possibility of various vectors being linearly dependent and that sort of thing) but the basic idea is simple. Suppose that the constrained maximum occurs at a

[1]The main results were 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 for to see Karush-Kuhn-Tucker referred to simply as Kuhn-Tucker.

spot where two fences intersect. Then the gradient of the hill must point into the corner where the fences meet. Mathematically, this means that the gradient of the objective function must lie between the two constraint gradients. More generally, Kuhn-Tucker says that the gradient of the objective function must lie in the cone spanned by the gradients of the binding constraints.

The results in these notes cover only necessary conditions, conditions that solutions to maximization problems must satisfy. I do not discuss how to guarantee that a candidate solution is a maximum rather than a minimum (or an inflection point, or saddle point, etc.), or that it is a global maximum rather than merely a local maximum. Part II of these notes develop sufficient conditions. The simplest version of these is that the KKT Conditions are sufficient as well as necessary if the objective function is concave and the constraint region is convex.

2 Basic definitions.

Consider a function $f : \mathbb{R}^N \to \mathbb{R}$. Suppose that we are interested in the behavior of f on a non-empty subset $C \subseteq \mathbb{R}^N$. For example, in a competitive demand problem, f would be a utility function and C would be the set of affordable consumption bundles. In general, I refer to f as the *objective function* and C as the *constraint set*, often called the *feasible set* or *opportunity set* in economics applications.

In practice, f is frequently defined over only a subset of $\mathbb{R}^N$. For example, it is common in examples to encounter utility functions of the form $U(x_1, x_2) = \alpha \ln(x_1) + (1 - \alpha) \ln(x_2)$, where $\alpha \in (0, 1)$. This utility function is defined only on $\mathbb{R}^N_{++}$. I ignore this complication in the development here.

Say that x^* is a *local maximum* iff there is an $\varepsilon > 0$ such that $f(x^*) \geq f(x)$ for any $x \in N_\varepsilon(x^*) \cap C$. I write $N_\varepsilon(x^*) \cap C$ because x^* could be at a boundary of C. Recall that x^* is *interior* iff there is an $\varepsilon > 0$ such that $N_\varepsilon(x^*)$ is contained in C. At an interior local maximum, there is an $\varepsilon > 0$ such that $f(x^*) \geq f(x)$ for all $x \in N_\varepsilon(x^*)$. Finally, if $f(x^*) \geq f(x)$ for all $x \in C$ then x^* is a maximum, sometimes called a *gobal* maximum.

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

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

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

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. □

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 $x^* = (1 + \sqrt{17})/8$. □

3 Interior Maxima.

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

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

Proof. Suppose that x^* is an interior local maximum. Let e^n denote the unit vector in which coordinate n is 1 and all other coordinates are zero. Then the partial derivative for coordinate n, $D_n f(x^*)$, is the limit of

$$\frac{f(x^* + \alpha e^n) - f(x^*)}{\alpha}$$

as $\alpha \to 0$. Since x^* is a local maximum, $f(x^* + \alpha e^n) - f(x^*) \leq 0$ for all α sufficiently small. In particular, for all $\alpha > 0$ sufficiently small, the above fraction is weakly negative and hence $D_n f(x^*) \leq 0$. Similarly, for all $\alpha < 0$ sufficiently small, the above fraction is weakly positive and hence $D_n f(x^*) \geq 0$. Combining, $D_n f(x^*)$. This is true for all n and so $\nabla f(x^*) = 0$. ■

Similarly, 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, $f : \mathbb{R} \to \mathbb{R}$ 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 a 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. Let $C \subseteq \mathbb{R}^N$ and consider the maximization problem,

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

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, $\max_{x \in C} u(x)$ where $C = \{x \in \mathbb{R}^N_+ : p \cdot x \leq m\}$. This is often written,

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

In standard form, the problem is,

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

More generally, a maximization problem is in *standard form* iff there are K functions g_k from $\mathbb{R}^N$ to $\mathbb{R}$ such that C is the set of x such that $g_k(x) \leq 0$. More concisely, letting $g = (g_1, \ldots, g_K)$, 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.

Say that constraint k is *binding* at x^* if $g_k(x^*) = 0$. If constraint k is not binding then it is *slack*. Given a feasible x^*, define

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

to be the set of indices of the binding constraints. From the perspective of first order conditions, only the binding constraints matter.

The Karush-Kuhn-Tucker (KKT) Theorem states that if x^* is a local maximum then, subject to a technical condition, there exist K numbers $\lambda_k^* \geq 0$, one λ_k^* for each constraint, such that the following hold.

1. $\nabla f(x^*) = \sum_{k=1}^{K} \lambda_k^* \nabla g_k(x^*)$.
2. $\lambda_k^* g_k(x^*) = 0$ for all k.

Equivalently, there are weights $\lambda_k^* \geq 0$ such that,

1. $\nabla f(x^*) = \sum_{k \in J} \lambda_k^* \nabla g_k(x^*)$,
2. $\lambda_k^* = 0$ for all $k \notin J$.

I refer to these as the *KKT conditions*. The λ_k^* are called *KKT multipliers*. The KKT conditions for a MIN problem (in standard form) are identical.

KKT condition 2, called *complementary slackness*, says that if k is slack then $\lambda_k^* = 0$. As for KKT condition 1, for $J \neq \emptyset$, let W set (positively) spanned by the gradients of the binding constraints:

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

W is a closed, convex cone.[2] KKT Condition 1 states that $\nabla f(x^*) \in W$, is in Figure 1. I gave intuition for this in Section 1.

[2]Convexity is almost immediate. As for closed, consider any sequence $\{w_t\}$ in W and suppose that there is a $w \in \mathbb{R}^N$ such that $w_t \to w$. I must show that $w \in W$. Let A be the matrix whose columns are the elements of S. By CQ, this matrix has full rank. Since $w_t \in W$ then there is a vector, call it λ_t, of multipliers λ_{tk}, $k \in J$, such that $w_t = \sum_{k \in J} \lambda_{tk} \nabla g_k(x^*)$. Write this as $w_t = A\lambda_t$. By CQ, the matrix $A'A$ is invertible. Therefore, since $A'w_t = A'A\lambda_t$, $\lambda_t = (A'A)^{-1}A'w_t$, hence $\lambda_t \to (A'A)^{-1}A'w$. Let $\lambda = (A'A)^{-1}A'w$. Since $\lambda_t \geq 0$, $\lambda \geq 0$. Finally, since $w_t = A\lambda_t$ for all t, $w_t \to w$, and $\lambda_t \to \lambda$, it follows that $w = A\lambda$, which implies that $w \in W$, as was to be shown. Essentially the same argument shows that any cone that is generated by finitely many vectors is closed. In this more general case, Caratheodory's Theorem implies that even if S is not independent, one can find a subset of S that is independent and that spans W, so that there is still a matrix A of full rank such that $w \in W$ iff $w = A\lambda$ for some $\lambda \geq 0$. The proof then proceeds as above.

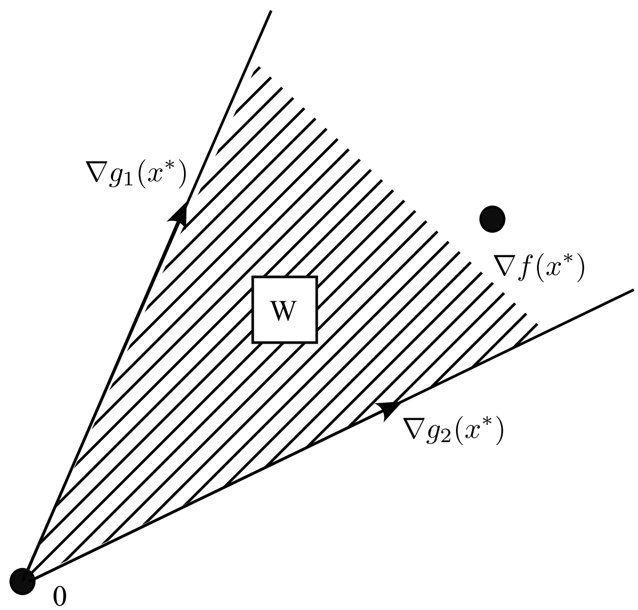

Figure 1: W and KKT Condition 1.

As I discuss in detail in the notes on the Envelope Theorem, λ_k^* measures how much the objective function would be increased if the constraint 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. If a constraint is slack then it does not matter (for local maximization) and its multiplier is zero.

As noted above, the KKT Theorem assumes a technical condition. To see why, consider the following example.

Example 8. Let $f(x) = x$ and let $g(x) = x^2$. Then the constrain set is $C = \{0\}$. The solution is, trivially, $x^* = 0$. Here, however, the conclusion of KKT fails. Since $\nabla f(x^*) = 1$ while $\nabla g(x^*) = 0$, there is no $\lambda \geq 0$ such that $\nabla f(x^*) = \lambda \nabla g(x^*)$. $\square$

A more subtle example, in which the constraint gradients don't vanish, is the following.

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 conclusion of KKT 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 $\nabla f(x^*) = \lambda_1 \nabla g_1(x^*) + \nabla g_2(x^*)$. $\square$

I rule out Example 8 and 9 by assuming either one of the following.

Definition 1. *Assume that the g_k are differentiable, let x^* be feasible, and let $S =$*

$\{\nabla g_k(x^*)\}_{k\in J}$. *S is the set of gradients of the binding constraints.*

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

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

PI rules out Example 8 and 9. In Example 8, PI fails because the gradient of the one binding constraint is 0. In Example 9, PI fails because $\nabla g_1(x^*) + \nabla g_2(x_s) = 0$. The main result, Theorem 3, says that PI is sufficient for existence of KKT multipliers. Moreover, Theorem 3 also says that if LI (which implies PI) holds then the KKT multipliers not only exist but are unique.

Example 10. Let the domain be $\mathbb{R}$, $f(x) = x$, $g_1(x) = x$, $g_2(x) = 2x$. Then $C = \{0\}$ and both constraints are binding. $S = \{1, 1/2\}$. PI holds and the KKT conditions hold with $\lambda_1 = 1$, $\lambda_2 = 0$. But the KKT conditions also hold with $\lambda_1 = 0$, $\lambda_2 = 1/2$. LI fails here, since $1 + (-2)1/2 = 0$. □

PI has a geometric interpretation that plays an important role in the proof of the main result. Say that a cone W is *pointed* iff for any $w \in W$, $w \neq 0$, $-w \notin W$. In Figure 1, W is pointed.

Theorem 2. *If* $J \neq \emptyset$ *then PI holds iff (a)* $\nabla g_k(x^*) \neq 0$ *for all* $k \in J$ *and (b)* W *is pointed.*

Proof. $\Rightarrow$. It is immediate that PI implies $\nabla g_k(x^*) \neq 0$ for all $k \in J$. To show that W is pointed, I argue by contraposition. If W is not pointed then there is a $w \in W$, $w \neq 0$, $-w \in W$. Then there are $\lambda_k, \hat{\lambda}_k \geq 0$, not all zero, such that $w = \sum_{k\in J} \lambda_k \nabla g_k(x^*)$, $-w = \sum_{k\in J} \hat{\lambda}_k \nabla g_k(x^*)$. But then $0 = w + (-w) = \sum_{k\in J}(\lambda_k + \hat{\lambda}_k)\nabla g_k(x^*)$ and $\lambda_k + \hat{\lambda}_k > 0$ for at least some k, violating PI.

$\Leftarrow$. By contraposition. Suppose that $\sum_{k\in J} \lambda_k \nabla g_k(x^*) = 0$, $\lambda_k \geq 0$ with at least some strictly positive. If any $\nabla g_k(x^*) = 0$ for some k then the (a) condition is violated. Assume, therefore, that $\nabla g_k(x^*) \neq 0$ for all $k \in J$. Take any $k \in J$ such that $\lambda_k > 0$ and let $w = \lambda_j \nabla g_j(x^*) \neq 0$. Then $\sum_{k\in J} \lambda_k \nabla g_k(x^*) = 0$ implies $-w = \sum_{k\in J, k\neq j} \lambda_k \nabla g_k(x^*)$, hence $-w \in W$, hence W is not pointed. ■

PI is equivalent to yet another condition, first formulated in Mangasarian and Fromovitz (1967), but I postpone discussing that to Section 4.3. The main result of these notes is the following.

Theorem 3 (Karush-Kuhn-Tucker)**.** *Consider a differentiable MAX problem in standard form. Let* x^* *be a local maximum. Suppose that PI holds at* x^**. Then there are numbers* $\lambda_k^* \geq 0$*, the KKT multipliers, such that*

1. $\nabla f(x^*) = \sum_{k=1}^{K} \lambda_k^* \nabla g_k(x^*)$,

2. $\lambda_k^* g_k(x^*) = 0$ *for every* k.

If LI (which implies PI) holds, then the KKT multipliers are unique.

Proof. If $J = \emptyset$, meaning that no constraints are binding, then KKT Condition 1, becomes $\nabla f(x^*) = 0$. By Theorem 1, this condition holds if x^* is a local maximum and there are no binding constraints.

Otherwise, $J \neq \emptyset$, meaning that at least one constraint is binding. Suppose that $\nabla f(x^*) \in W$, where W, defined above, is the cone spanned by the gradients of the binding constraints. (This holds trivially if $\nabla f(x^*) = 0$.) Then for $k \in J$, simply take $\lambda_k^* \geq 0$ such that $\nabla f(x^*) = \sum_{k \in J} \lambda_k^* \nabla g_k(x^*)$. If LI holds, these λ_k^* are unique. For $k \notin J$, KKT Condition 2 requires $\lambda_k^* = 0$. In summary, if $\nabla f(x^*) \in W$, there are λ_k^* such that the KKT Condition 1 and 2 hold. If LI holds, these λ_k^* are unique.

It remains to show that if x^* is a local maximum, then $\nabla f(x^*) \in W$. I argue by contraposition: if $\nabla f(x^*) \notin W$, then x^* cannot be a local maximum. Suppose that $\nabla f(x^*) \notin W$. As discussed above, PI implies that the closed, convex cone W is pointed. PI also implies that $\nabla g_k(x^*) \neq 0$ for $k \in J$, hence $W \neq \{0\}$. Moreover, the set $\{\nabla f(x^*)\}$ is trivially compact. It then follows from a theorem on separation for pointed cones that there is a $v \in \mathbb{R}^N$ such that $v \cdot \nabla f(x^*) > 0$ while for all $w \in W$, $w \neq 0$, $v \cdot w < 0$. In particular, since PI implies that $\nabla g_k(x^*) \neq 0$ for $k \in J$, $v \cdot \nabla g_k(x^*) < 0$. See Figure 2.

Since $v \cdot \nabla f(x^*) = D_v f(x^*)$ and $v \cdot \nabla g_k(x^*) = D_v g_k(x^*)$, it follows there is an $\overline{\alpha}$ such that for all $\alpha \in (0, \overline{\alpha})$, setting $x_\alpha = x^* + \alpha v$, the following hold.

1. $f(x_\alpha) - f(x^*) > 0$, hence $f(x_\alpha) > f(x^*)$.

2. For $k \in J$, $g_k(x_\alpha) - g_k(x^*) < 0$, hence $g_k(x_\alpha) < g_k(x^*) = 0$.

3. For $k \notin J$, $g_k(x_\alpha) < 0$ (this last follows from continuity of the g_k and the fact that $g_k(x^*) < 0$).

In words, all such x_α are feasible (in fact, interior to the constraint set) and yield a higher value of the objective function than x^*: x^* is not a local maximum, as was to be shown. ■

An explicit check of PI is typically not necessary for reasons given in Section 4.3. In their original formulations, Karush (1939) and Kuhn and Tucker (1951) assume a form of constraint qualification that is weaker than PI. I discuss this briefly in Section 4.6.

4.3 Checking Constraint Qualification: The Slater Condition.

PI is equivalent to a condition that first appeared in Mangasarian and Fromovitz (1967).

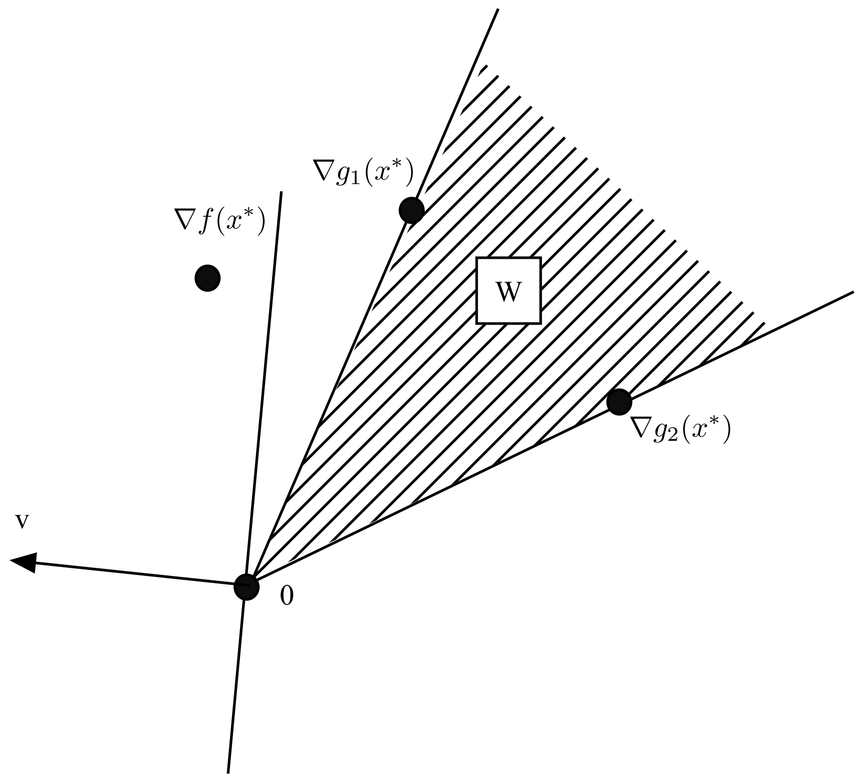

Figure 2: The separation argument.

Definition 2. *Assume that the g_k are differentiable and let x^* be feasible.* Mangasarian Fromovitz constraint qualification (MF) *holds at x^* iff $J = \emptyset$ or there is a $v \in \mathbb{R}^N$ such that $v \cdot \nabla g_k(x^*) < 0$ for all $k \in J$.*

MF says that it is possible to move a small distance away from x^* into the interior of the constraint set C. For the modified MAX problem of Section 4.4, in which equality constraints are added to the problem, MF requires that it is possible to move a small distance away from x^* into the *relative* interior of C (relative to the subset of $\mathbb{R}^N$ defined by the equality constraints).

Theorem 4. *PI and MF are equivalent.*

Proof. It suffices to consider $J \neq \emptyset$.

MF $\Rightarrow$ PI. By Theorem 2, it suffices to show that (a) $\nabla g_k(x^*) \neq 0$ for all $k \in J$, and (b) W is pointed. By MF, there is a v such that $v \cdot \nabla g_k(x^*) < 0$ for all $k \in J$, which implies (a). For (b), for any $w \in W$, $w \neq 0$, let $\lambda_k \geq 0$ be such that $w = \sum_{k \in J} \lambda_k \nabla g_k(x^*)$. Since $w \neq 0$, $\lambda_k > 0$ for at least one k. Then $v \cdot w = \sum_{k \in J} \lambda_k (v \cdot \nabla g_k(x^*)) < 0$. That is, $v \cdot w < 0$ for every $w \in W$, $w \neq 0$. For any such w, $v \cdot (-w) > 0$, hence $-w \notin W$, as was to be shown.

PI $\Rightarrow$ MF. The proof is very similar to the proof of Theorem 3. W is a closed, convex cone. If PI holds then, by Theorem 2, W is pointed and $W \neq \{0\}$. Take any $\hat{w} \in W$, $\hat{w} \neq 0$. Then $-\hat{w} \notin W$. By the same separation theorem used in Theorem 3, there is a v such that $v \cdot (-\hat{w}) > 0$ and $v \cdot w < 0$ for all $w \in W$, $w \neq 0$. By PI, $\nabla g_k(x^*) \neq 0$ for all $k \in J$. Hence $v \cdot \nabla g_k(x^*) < 0$ for all $k \in J$. ■

A sufficient condition for MF, and hence for PI, is that the binding constraint functions are convex and that the following condition holds.

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

For the MAX problem as stated, the Slater Condition is equivalent to requiring that the constraint set C have a non-empty interior. For the modified MAX problem of Section 4.4, 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: if there are points arbitrarily near x^* that are interior to the constraint set then, in particular, the constraint set has a non-empty interior. The converse, that Slater implies MF, is not necessarily true (as illustrated by Example 11 below) but it is true if the constraints are convex.

Theorem 5. *If the binding constraint functions are convex and the Slater condition holds then PI holds and hence, if MAX is differentiable, the KKT conditions hold at a local maximum.*

Proof. If no constraints are binding then there is nothing to show. Suppose then that S is not empty. Let x be as in the Slater condition and let $v = x - x^*$. Then for any $k \in J$, convexity implies

$$g_k(x) \geq Dg_k(x^*)v + g_k(x^*).$$

Since $g_k(x^*) = 0$ by construction of J and since $g_k(x) < 0$ by Slater, $Dg_k(x^*)v < 0$. Since this holds for all $k \in J$, MF follows, and hence so does PI. ■

In Example 8 in Section 4.2, the constraint is convex but Slater fails and PI fails. In the following example, Slater holds but the constraint is not convex and PI fails.

Example 11. The domain is $\mathbb{R}$. $f(x) = x$ and $g(x) = x^2 + x^3$. The constraint set is $(-\infty, -1] \cup \{0\}$ and the solution is $x^* = 0$. Slater holds; for example, $g(-2) = -4 < 0$, but g is not convex. KKT Condition 1 fails. Since $Df(x^*) = 1$ while $Dg(x^*) = 0$, there is no λ such that $Df(x^*) = \lambda Dg(x^*)$. PI fails since $\nabla g(x^*) = 0$. □

In practice, constraints in economic problems are often convex and so checking PI boils down to checking Slater, which is often trivial. For completeness, in Section 4.6, I briefly discuss constraint qualification conditions that are weaker than PI.

4.4 Equality constraints.

I 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 the following.

1. $\nabla f(x^*) = \sum_k \lambda_k \nabla g_k(x^*) + \sum_\ell \mu_\ell \nabla r_\ell(x^*)$.

2. For all inequality constraints k, $\lambda_k g_k(x^*) = 0$ and $\lambda_k \geq 0$.

3. For all equality constraints ℓ, $r_\ell(x^*) = 0$.

Note that 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.

4.5 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 necessary condition for MAX,

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

Suppose that the non-negativity constraint for x_n is constraint k. If $\lambda_k > 0$, then $\lambda_k \nabla g_k(x^*) = (0, \ldots, 0, -\lambda_k, 0, \ldots, 0)$, with $-\lambda_k$ in the nth place, so this constraint is *lowering* the right-hand side of the above equality. Let I be the set of 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 KKT Condition 1 as an *in*equality. I think it is easier, and safer, to remember KKT Condition 1 as an equality, with all constraints explicit.

4.6 Other Forms of Constraint Qualification.

PI is stronger than necessary for existence of KKT multipliers, as illustrated by the following example.

Example 12. 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 $x^* = (1, 1)$. PI fails: $W = \{x \in \mathbb{R}^2 : x_1 + x_2\}$, 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 $\lambda_1 = 1$, $\lambda_2 = 0$. Note that these multipliers are not unique. $\lambda_1 = 0$, $\lambda_2 = -1$ also works. □

In practice, examples in which PI fails are extremely rare. For completeness, however, I briefly discuss weaker conditions for existence of KKT multipliers.

The set W is always a closed, convex cone. Referring to the proof of Theorem 3, if $\nabla f(x^*) \notin W$ then, even if W is not pointed, there is a vector $v \neq 0$ such that $v \cdot \nabla f(x^*) > 0$ while $v \cdot \nabla g_k(x^*) \leq 0$ for all binding k. If in fact $v \cdot \nabla g_k(x^*) < 0$ then the situation is exactly as in the original proof. If, however, $v \cdot \nabla g_k(x^*) = 0$ then matters are more delicate.

If the binding constraints are *concave*, which includes linear constraints as a special case, as in Example 12, then $x^* + \alpha v$ is feasible when $v \cdot \nabla g_k(x^*) = 0$: by concavity, $g_k(x^* + \alpha v) \leq Dg_k(x^*)(x^* + \alpha v - x^*) + g(x^*)$, hence $g(x^* + \alpha v) \leq 0$ since $Dg_k(x^*)v = v \cdot \nabla g_k(x^*) = 0$ and $g(x^*) = 0$.

In the general case, $x^* + \alpha v$ need not be feasible for small α. One might, however, still be approximate $x^* + \alpha v$ with feasible x, in which case the proof of Theorem 3 will still go through. Example 9 in Section 4.2 illustrates one difficulty that can arise with this approach. In this example, at $x^* = 0$, $\nabla f(x^*) = (1, 0)$ and W is the x_2 axis. A v that separates $\nabla f(x^*)$ from W must be (collinear with) (1,0). For each constraint k there is a sequence of points $\{x_t\}$ such that (a) $g_k(x_t) \leq 0$ (x_t satisfies constraint k), (b) $x_t \to x^*$, and (c) $x_t - x^*$ is asymptotically collinear with $v = (1, 0)$:

$$\lim_t \frac{x_t - x^*}{\|x_t - x^*\|} \to v.$$

For example, for g_1, take $x_t = (1/t, -1/t^2)$. The order of quantifiers matters here, however: in Example 9, 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 $v \cdot \nabla f_k(x^*) \leq 0$ for all $k \in J$, there is a sequences of feasible points that satisfy conditions (a) through (c) above for all $k \in J$. Informally, they assume that any direction v that should be feasible, given the derivatives of the binding g_k, is actually feasible. This form of constraint qualification is weaker than PI but examples in which the difference matters tend to be artificial. For more on constraint qualification, see Eustaquio, Karas, and Ribeiro (2008).

4.7 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 13. 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. □

In Example 13, $\lambda = 0$ even though the constraint is binding. Call a constraint k *active* if $\lambda_k > 0$.[3] Although Example 13 provides a counter example, binding constraints will typically be active. Active constraints are always binding, by KKT Condition 2.

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

Example 14. 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. □

The problem illustrated by Example 14 does not occur if the objective function is concave, in which case local maxima are global maxima.

5 Convex Optimization

5.1 Saddle points and the KKT Conditions.

I have presented the KKT Theorem in the form that I find most straightforward, but there is another important approach that I now present. To avoid undue com-

[3]Many references use "active" to mean "binding."

plication, I focus on MAX problems with a single variable, $x \in \mathbb{R}$, and a single constraint, $g : \mathbb{R} \to \mathbb{R}$. It should be evident that the theory below extends easily to $x \in \mathbb{R}^N$ and K constraints.

Given a MAX problem, define $L : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ by

$$L(x, \lambda) = f(x) - \lambda g(x).$$

L is called the *Lagrangian*. Say that (x^*, λ^*) is a *saddle point* of L iff

$$L(x, \lambda^*) \leq L(x^*, \lambda^*) \leq L(x^*, \lambda)$$

for all x, λ. (x^*, λ^*) is a saddle point iff (x^*, λ^*) maximizes L with respect to x and *minimizes* L with respect to to λ.

If (x^*, λ^*) is a saddle point of L, then (x^*, λ^*) satisfies the KKT Conditions for the MAX problem, which I formalize as Theorem 6. The proof is in Section 5.2.

Theorem 6. *Given a MAX problem and associated Lagrangian L, if (x^*, λ^*) is a saddle point of L and L is differentiable then the KKT Conditions hold.*

The remaining results in this section establish that x^* solves the MAX problem iff there is a $\lambda^* \geq 0$ such that (x^*, λ^*) is a saddle point of the associated L. The proofs are again in Section 5.2.

Theorem 7. *Given a MAX problem and associated L, if (x^*, λ^*) is a saddle point of L then x^* solves MAX.*

Recall the *Slater condition* from Section 4.6. With only one constraint, the Slater condition holds iff there is an x such that $g(x) < 0$.

Theorem 8. *Given a MAX problem and associated L, suppose that x^* solves MAX. If f is concave, g is convex, and g satisfies the Slater condition, then there is an $\lambda^* \geq 0$ such that (x^*, λ^*) is a saddle point of L.*

In view of Theorem 6, Theorem 8 is a variant of the KKT Theorem. A related comment is that while Theorem 7 makes no assumptions about the MAX problem, Theorem 8 makes strong assumptions. Some assumptions should be expected: we know from Section 4.6 that the KKT conditions can fail in an arbitrary MAX problem.

If f and g are differentiable then Theorem 8 follows almost immediately from Theorem 5. More explicitly, Theorem 5 implies that there is an $\lambda^* \geq 0$ such that $\nabla f(x^*) - \lambda^* \nabla g(x^*) = 0$, hence

$$D_x L(x^*, \lambda^*) = 0.$$

Moreover, if f is concave and g is convex then L is concave. It follows that for any x, $L(x, \lambda^*) \leq D_x L(x^*, \lambda^*)(x - x^*) + L(x^*, \lambda^*) = L(x^*, \lambda^*)$, which implies that

x^* maximizes $L(x^*, \lambda^*)$. On the other hand, λ^* trivially minimizes $L(x^*, \lambda)$. If the constraint is binding then $g(x^*) = 0$, hence $L(x^*, \lambda) = f(x^*) - \lambda g(x^*) = f(x^*)$. If the constraint is not binding then $L(x^*, \lambda)$ is decreasing in λ, and hence is minimized at $\lambda = 0$.

Theorem 8 does not, however, assume differentiability. It requires, as a consequence, a new proof. This leads to a more general observation. Because Theorem 7 and 8 do not assume differentiability, one can use them to construct a version of KKT that holds even without differentiability, provided f is concave and g is convex. Instead of gradients, one works with *sub*-gradients. For a concave function on $\mathbb{R}$, the sub-gradient at a point x^* is the *set* of numbers d such that $f(x) \leq d(x - x^*) + f(x^*)$. For example, if $f(x) = -|x|$ then the sub-gradient at $x^* = 0$ is $[-1, 1]$. If f is differentiable (and concave) then $d = Df(x^*)$ is the *unique* number such that $f(x) \leq d(x - x^*) + f(x^*)$ for all x. An analogous construction holds for convex functions. I do not pursue the topic of KKT without differentiability here, but it is important in some applications.

In Example 8 in Section 4.2, f is concave, g is convex, but Slater fails and KKT multipliers do not exist. In Example 11 in Section 4.6, f is concave, Slater holds, but g is not convex and KKT multipliers do not exist. In both cases, the saddle point property fails: there is no λ^* such that $L(x^*, \lambda^*) \geq L(x, \lambda^*)$ for all x.

In the next example, Slater holds and g is convex but f is not concave. In contrast to the previous examples, KKT multipliers exist but the saddle point property still fails. There is no λ^* such that $L(x^*, \lambda^*) \geq L(x^*, \lambda)$ for all x.

Example 15. $f(x) = e^x - 1$ and $g(x) = x$. The constraint set is $(-\infty, 0]$ and the solution is $x^* = 0$. Slater holds ($g(-1) = -1 < 0$) and g is convex. In fact, LI holds. $Df(0) = 1$ and $Dg(0) = 1$ so KKT Condition 1 holds with $\lambda^* = 1$. As in the previous examples, $g(0) = 0$ and so $L(x^*, \lambda^*) \leq L(x^*, \lambda)$ for any $\lambda, \lambda^* \geq 0$. The condition $L(x^*, \lambda^*) \geq L(x, \lambda^*)$ reduces to,

$$0 \geq e^x - 1 - \lambda^* x.$$

There is no $\lambda^* \geq 0$ for which this holds for all x; for $\lambda^* = 1$ (as required by KKT Condition 1), the inequality does not hold for *any* $x \neq 0$. □

In the next, and last, example, all conditions are met.

Example 16. $f(x) = -e^{-x} + 1$ and $g(x) = x$. The constraint set is $(-\infty, 0]$ and the solution is, once again, $x^* = 0$. Slater holds ($g(-1) = -1 < 0$) and g is convex so local Slater holds as well, and in fact LI holds. As in Example 15, $Df(0) = 1$ and $Dg(0) = 1$ so KKT Condition 1 holds with $\lambda^* = 1$. Finally, the saddle point property holds. As in the previous examples, the concern is the inequality $L(x^*, \lambda^*) \geq L(x, \lambda^*)$, which here becomes

$$0 \geq -e^{-x} + 1 - \lambda^* x.$$

For $\lambda^* = 1$, this inequality holds for all x (indeed, with strict inequality except for $x = x^*$). □

Finally, recall that these notes discuss only necessary conditions for a maximum, not sufficient conditions. The notes on sufficiency establish that the KKT conditions are sufficient as well as necessary if, as here, f is concave and g is convex.

5.2 Proofs of the saddle point theorems.

Proof of Theorem 6. If (x^*, λ^*) is a saddle point then $D_x L(x^*, \lambda^*) = 0$, hence $\nabla f(x^*) - \lambda^* \nabla g(x^*) = 0$, or

$$\nabla f(x^*) = \lambda^* \nabla g(x^*),$$

which is KKT Condition 1.

For KKT Condition 2, since (x^*, λ^*) is a saddle point, $L(x^*, \lambda^*) \leq L(x^*, \lambda)$ for all $\lambda \geq 0$, which implies that $f(x^*) - \lambda^* g(x^*) \leq f(x^*) - \lambda g(x^*)$ for all $\lambda \geq 0$, or

$$(\lambda - \lambda^*) g(x^*) \leq 0$$

for all $\lambda \geq 0$. Since this must hold for λ arbitrarily large, this implies that

$$g(x^*) \leq 0.$$

Moreover, since this must hold if $\lambda = 0$, this implies that if $g(x^*) < 0$ then $\lambda^* = 0$. Combining,

$$\lambda^* g(x^*) = 0,$$

which is KKT Condition 2. ■

Proof of Theorem 7. The half of the proof of Theorem 6 that addresses KKT Condition 2 (and which does not use differentiability) establishes that $g(x^*) \leq 0$ (x^* is feasible) and $\lambda^* g(x^*) = 0$.

To see that x^* is optimal, note that since (x^*, λ^*) is a saddle point of L, $L(x^*, \lambda^*) \geq L(x, \lambda^*)$ for every x, hence

$$f(x^*) - \lambda^* g(x^*) \geq f(x) - \lambda^* g(x)$$

for every x. Since, as just established, $\lambda^* g(x^*) = 0$, this implies

$$f(x^*) \geq f(x) - \lambda^* g(x),$$

for every x. If x is feasible then, $g(x) \leq 0$, hence $-g(x) \geq 0$, hence (since $\lambda^* \geq 0$), $-\lambda^* g(x) \geq 0$. Therefore, for feasible x, $f(x) - \lambda^* g(x) \geq f(x)$, hence

$$f(x^*) \geq f(x),$$

which establishes that x^* solves MAX. ■

Proof of Theorem 8. As discussed following the statement of Theorem 8, that theorem is almost immediate if f and g are differentiable. The novelty in the following argument is that it does not require differentiability.

Define two sets

$$A = \left\{ a \in \mathbb{R}^2 : \exists x \in \mathbb{R} \text{ such that } a_1 \leq f(x) \text{ and } a_2 \leq -g(x) \right\}$$

and

$$B = \left\{ b \in \mathbb{R}^2 : b_1 > f(x^*) \text{ and } b_2 > 0 \right\}.$$

These sets are convex. In particular, A is convex since f is concave and g is convex (hence $-g$ is concave). B is convex since it is a translation of the strictly positive orthant in $\mathbb{R}^2$.

The sets are also disjoint. To see this, suppose there is a x such that such that $f(x) > f(x^*)$ and hence there is an a_1 such that $f(x^*) < a_1 \leq f(x)$. Since, by assumption, x^* solves MAX, x must be infeasible: $g(x) > 0$, or $-g(x) < 0$. Hence, for any such x, $a_2 < 0$. Then $(a_1, a_2) \notin B$.

By the Separating Hyperplane Theorem (for sets that are convex but not necessarily closed/compact), since A, and B are disjoint convex sets, there is a vector $q = (q_1, q_2) \neq (0, 0)$ such that for any $a \in A$, $b \in B$,

$$a \cdot q \leq b \cdot q \tag{1}$$

Since I can take a_1 and a_2 to be arbitrarily negative, Inequality 1 implies that $q_1, q_2 \geq 0$. I claim that $q_1 > 0$. The argument is by contraposition: if $q_1 = 0$ then Slater cannot hold. Explicitly, taking a sequence in B converging to $(f(x^*), 0)$ establishes that, for any feasible x, since $(f(x), -g(x)) \in A$,

$$q_1 f(x) - q_2 g(x) \leq q_1 f(x^*). \tag{2}$$

If $q_1 = 0$, then, since $(q_1, q_2) \neq 0$ and $q_2 \geq 0$, it follows that $q_2 > 0$. Moreover, if $q_1 = 0$, then from Inequality 2,

$$-q_2 g(x) \leq 0.$$

If $q_2 > 0$, this holds iff $g(x) \geq 0$ for every x. This shows that Slater does not hold. By contraposition, if Slater holds then $q_1 > 0$.

Define

$$\lambda^* = \frac{q_2}{q_1},$$

which is well defined since $q_1 > 0$. Then for any $a \in A$ and any $b \in B$

$$a_1 + \lambda^* a_2 \leq b_1 + \lambda^* b_2.$$

Again taking a sequence in B converging to $(f(x^*), 0)$, this implies that for any x,

$$f(x) - \lambda^* g(x) \leq f(x^*). \tag{3}$$

In particular, for $x = x^*$,

$$f(x^*) - \lambda^* g(x^*) \le f(x^*)$$

or

$$-\lambda^* g(x^*) \le 0.$$

On the other hand, x^* is feasible, hence $g(x^*) \le 0$, hence, since $\lambda^* \ge 0$, $-\lambda^* g(x^*) \ge 0$. Combining

$$\lambda^* g(x^*) = 0.$$

Therefore, for any x,

$$f(x) - \lambda^* g(x) \le f(x^*) = f(x^*) - \lambda^* g(x^*)$$

which shows that $L(x^*, \lambda^*) \ge L(x, \lambda^*)$.

It remains to show that $L(x^*, \lambda) \le L(x^*, \lambda^*)$. Again $\lambda^* g(x^*) = 0$. On the other hand, $g(x^*) \le 0$, hence for any $\lambda \ge 0$, $\lambda g(x^*) \le 0$, hence $-\lambda g(x^*) \ge 0$. It follows that for any $\lambda \ge 0$,

$$-\lambda^* g(x^*) = 0 \le -\lambda g(x^*)$$

or

$$f(x^*) - \lambda^* g(x^*) \le f(x^*) - \lambda g(x^*)$$

which shows that $L(x^*, \lambda) \le L(x^*, \lambda^*)$. Thus (x^*, λ^*) is saddle point of L. ■

6 Using the KKT Conditions.

The bottom line here is bad news: there is no simple procedure for finding points x^* and multipliers λ that satisfy the KKT conditions

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 constraint happens to be labeled 1, then we get as necessary conditions:

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

This gives $n+1$ non-linear equations in $n+1$ unknowns, namely the x and λ_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. If, on the other hand, all solutions lie outside of 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.

Theorem 9. *Consider any function $f : \mathbb{R}^N \to \mathbb{R}$ and any set $C \subseteq \mathbb{R}^N$. Let $h : \mathbb{R} \to \mathbb{R}$ 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 x^* and 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 value of the solution, you will, typically, change the values of the KKT multipliers.

In the following examples, the constraints are linear, and hence, as discussed in Section 4.6, constraint qualification holds.

Example 17. 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:

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

Next, note that any solution x^* will be strictly positive. It is, for example, feasible to take $x_1 = 3/4$, $x_2 = 3/8$, $x_3 = 1$. This yields a positive value for the objective function, whereas the o bjective function is 0 if any $x_n = 0$.

Since at an optimum the $x \geq 0$ constraints do not bind, KKT Condition 2 implies that the KKT multipliers on these constraints are 0, and hence KKT Condition 1 becomes

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

where

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

KKT Condition 2 yields $\lambda_1(4x_1^* + 8x_2^* + 3x_3^* - 9) = 0$. KKT Condition 1 implies $\lambda_1 > 0$ and hence KKT Condition 2 implies $4x_1^* + 8x_2^* + 3x_3^* = 9$. One can also argue more directly that the first constraint must bind since $Df(x) \gg 0$ for any $x \gg 0$.

From KKT Condition 1,

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

Substituting this into KKT Condition 2 (that is, into the budget constraint) yields,

$$\frac{\alpha}{2\lambda_1} + \frac{\alpha}{3\lambda_1} + \frac{\alpha}{6\lambda_1} = \frac{\alpha}{\lambda_1} = 9.$$

Substituting this back into KKT Condition 1 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

$$\lambda_1 = \frac{1}{48} 3^{1/3} 2^{3/4} > 0,$$

and $\lambda_2 = \lambda_3 = \lambda_4 = 0$.

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 9, this yields a new maximization problem with the same solution as the original one (as you can verify by direct calculation):

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

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. For completeness, therefore, define the value of the modified objective function to be $-\infty$ iff any $x_n = 0$. This sort of technicality is common in optimization problems. □

Example 18. Consider the same problem as in Example 17 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 17 is feasible. Since constrain 1 binds at $\tilde{x}$, KKT Condition 1 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 λ_1 that will work. For example, to satisfy the first line, λ_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 $f(x^*) \approx 0.68$. □

Example 19. Once again consider the same problem as in Example 17 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 $\hat{x}$, KKT Condition 1 requires

$$\nabla f(\tilde{x}) = 0.$$

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

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

Example 20. Consider

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

Translating this into standard form,

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

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 constraint $4x_1 + 8x_2 + 3x_3 = 9$ implies $x_3^* = 3$. Since constraint 4 (the $x_3 \geq 0$ constraint) is not binding, it follows that $\lambda_4 = 0$ by KKT Condition 2. 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 KKT Condition 1, 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, $x^* = (0, 0, 3)$ satisfies the necessary conditions. Again, the sufficient conditions of Part II of these notes guarantee that x^* actually is the solution. □

Example 21. Consider the same problem as in Example 20. Suppose that I had instead guessed that only the first constraint was binding. Then by KKT Condition 2, I get $\lambda_2 = \lambda_3 = \lambda_4 = 0$. As for KKT Condition 1, I have

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

Manipulating this (square both sides, solve for $\tilde{x}$ in terms of λ_1, plug into the first constraint, solve for λ_1), yields

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

and

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

Notice that this "solution" satisfies KKT Conditions 1 and 2. BUT, it is not feasible, since constraints 2 and 3 are violated. □

Example 22. Once again, consider the same problem as in Example 20. 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 approximately

$$\begin{bmatrix} \lambda_1 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} \approx \begin{bmatrix} 0.1 \\ -0.4 \\ -1.4 \end{bmatrix}.$$

The critical point to notice here is that $\lambda_3, \lambda_4 < 0$. KKT requires all multipliers to be non-negative. This means that $\hat{x}$ has failed the KKT Conditions and therefore cannot be a maximum. If you had neglected to write the problem in standard form you would instead have found all the multipliers to be positive, and concluded (mistakenly) that x^* satisfied the KKT Conditions. You can compute that $f(\hat{x}) \approx 6.8$ whereas $f(x^*) = 9$. □

References

Eustaquio, R. G., E. W. Karas, and A. A. Ribeiro (2008): "Constraint Qualifications for Nonlinear Programming," Federal University of Parana, Brazil.

Karush, W. (1939): "Minima of Functions of Several Variables with Inequalities as Side Conditions," Masters Thesis, University of Chicago.

Kuhn, H. W., and A. W. Tucker (1951): "Nonlinear Programing," in *Proceedings of the 2nd Berkeley Symposium*, pp. 481–492. University of California Press.

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*, 7, 232–241.