Professor John Nachbar
Econ 4111
April 8, 2012

Finite Dimensional Optimization, Part II Sufficiency and Second Order Conditions

1 Interior Optima.

1.1 A Basic Sufficiency Result.

The following theorem gives sufficient conditions for a point x^* to be a maximum. This result, and the results to follow, is stated assuming that the domain of f is all of $\mathbb{R}^N$; the result can be generalized to allow f to be defined on only a convex subset of $\mathbb{R}^N$.

Theorem 1. *Let $f : \mathbb{R}^N \to \mathbb{R}$ be differentiable. Given x^*, suppose that the following conditions hold.*

1. $Df(x^*) = 0$.
2. *f is concave.*

Then x^* is a maximum. Moreover, if f is strictly concave then x^* is a strict maximum.

Proof. Since f is concave, then for any $x \in \mathbb{R}^N$,

$$f(x) \leq Df(x^*)(x - x^*) + f(x^*),$$

hence $f(x) \leq f(x^*)$, since $Df(x^*) = 0$. If f is strictly concave then the same argument gives $f(x) < f(x^*)$ for $x \neq x^*$. ■

Similarly, if $Df(x^*) = 0$ and f is *convex* then x^* is a minimum; if f is strictly convex then x^* is a strict minimum.

1.2 Second Order Conditions.

If f is $\mathcal{C}^2$ and $D^2f(x^*)$ is negative definite (i.e., f is differentially strictly concave at x^*), then, by continuity, $D^2f(x)$ is negative definite for every x in a sufficiently small ball around x, which implies that f is strictly concave on that ball. This implies the following corollary of Theorem 1.

Theorem 2. *Let $f : \mathbb{R}^N \to \mathbb{R}$ be $\mathcal{C}^2$. Given x^*, suppose that the following conditions hold.*

1. $Df(x^*) = 0$.

2. $D^2f(x^*)$ *is negative definite.*

Then x^* *is a local strict maximum.*

The conditions $Df(x^*) = 0$ and $D^2f(x^*)$ negative definite are called the *first* and *second order* sufficient conditions for a (local) maximum. "First order condition" is often abbreviated as FOC and "second order condition" is often abbreviated as SOC.

The next theorem gives first and second order necessary conditions for a local maximum.

Theorem 3. *Let* $f : \mathbb{R}^N \to \mathbb{R}$ *be* $\mathcal{C}^2$. *If* x^* *is a local maximum then the following conditions hold.*

1. $Df(x^*) = 0$.

2. $D^2f(x^*)$ *is negative semi-definite.*

Proof.

1. This was proved in the notes on necessary conditions.

2. The proof is by contraposition. Suppose that $Df(x^*) = 0$ but $D^2f(x^*)$ is not negative semi-definite. Then for some $v \in \mathbb{R}^N$, $v'D^2f(x^*)v > 0$. Since f is $\mathcal{C}^2$, this implies that along a line through x^* in the direction v, all $v'D^2f(x)v > 0$ for all x near x^*. Thus, the function is locally strictly convex along this line and x^* is a local strict minimum and hence cannot be a local maximum.

■

The sufficient second order condion is stronger than necessary second order condition. As an example of what is at issue, consider $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^4$. $Df(0) = 0$ and $D^2f(0) = 0$, so the necessary conditions are satisfied, but f is actually strictly convex and 0 is a strict minimum, hence not a maximum. As this illustrates, the sufficiency SOC cannot be weakened to semi-definiteness. Conversely, the necessity SOC cannot be strengthened to definiteness. As an example, consider $f(x) = -x^4$. Again $Df(0) = 0$ and $D^2f(0) = 0$. Here 0 is a strict maximum even though $D^2f(0)$ is only negative semi-definite.

A similar but more subtle issue concerns whether Theorem 3 can be strengthened to state that if x^* is a local maximum then f is locally concave. The answer is no. It is possible to construct an example in which, near the maximum, f fails to be locally concave because it oscillates with increasing frequency but decreasing amplitude. But these counter examples are somewhat pathological. As a practical matter, you will not be far wrong if you think of local concavity as necessary at an unconstrained local maximum.

2 Constrained Local Optima.

2.1 Basic Sufficiency with Constraints.

An argument similar to that of Theorem 1 establishes that the Karush-Kuhn-Tucker (KKT) conditions are sufficient as well as necessary provided the g_k are convex and f is concave. Theorem 4 in this section establishes a stronger result: the KKT conditions are sufficient as well as necessary provided the g_k are quasi-convex and either f is concave or $Df(x^*) \neq 0$ (as will typically be the case at a constrained maximum) and f is quasi-concave.

An intuition for why the stronger result is true is the following. Convexity of the g_k can be relaxed to quasi-convexity because the latter is sufficient to guarantee that the constraint set is convex. Concavity of f can be relaxed to quasi-concavity, when $Df(x^*) \neq 0$, for two reasons. First, to check that x^* is a *local* constrained maximum, some directions of movement away from x^* are irrelevant. To take an obvious example, there is no point in checking movement in directions v for which $Df(x^*)v < 0$, since f is decreasing in such directions. Assuming that the constraint set is locally convex and that $Df(x^*) \neq 0$, one can show that x^* will be a local constrained maximum provided the KKT conditions hold and f is only locally quasi-concave, rather than locally concave. I discuss this local intuition in more detail in the next section. Second, still assuming that the constraint set is convex and that $Df(x^*) \neq 0$, one can show that if x^* is a local constrained maximum then x^* is also a (global) constrained maximum provided f is (globally) quasi-concave, even if the KKT conditions fail.

To make the exposition somewhat self-contained, I briefly review quasi-concavity and quasi-convexity. A function f is *quasi-concave* iff for any $\theta \in (0,1)$ and any $x, \hat{x}$ with $f(\hat{x}) \geq f(x)$,

$$f(\theta\hat{x} + (1-\theta)x) \geq f(x).$$

Equivalently, f is quasi-concave iff for any x, the upper contour set

$$\{\hat{x} \in \mathbb{R}^N : f(\hat{x}) \geq f(x)\}$$

is convex. In a utility setting, the upper contour set of x is the set of consumption bundles that are at least as good as x. f is *strictly quasi-concave* iff for any $\theta \in (0,1)$, and any $x, \hat{x}$, with $x \neq \hat{x}$ and $f(\hat{x}) \geq f(x)$,

$$f(\theta\hat{x} + (1-\theta)x) > f(x).$$

If f is strictly quasi-concave then level sets cannot be affine; they must bend. In a utility setting, strict quasi-concavity, together with some form of monotonicity assumption (e.g., $Du(x) > 0$ for all x) implies that indifference curves bend in the usual way, away from the origin.

Similarly, g_k is *quasi-convex* iff for any $\theta \in (0,1)$ and any $x, \hat{x}$ with $f(\hat{x}) \leq f(x)$,

$$g_k(\theta\hat{x} + (1-\theta)x) \leq g_k(x).$$

Equivalently, g_k is quasi-convex iff for any x the *lower contour set*

$$\{\hat{x} \in \mathbb{R}^N : g_k(\hat{x}) \leq g_k(x)\}$$

is convex.

It is easy to show that any concave function is quasi-concave and that any convex function is quasi-convex. It is also easy to show that any increasing transformation function of a quasi-concave function is quasi-concave, and any increasing transformation function of a quasi-convex function is quasi-convex.

Combining these facts, it is often possible to show that a function is quasi-concave by showing that it is an increasing transformation of a concave function. For example, $f : \mathbb{R}^2_+ \to \mathbb{R}$,

$$f(x_1, x_2) = x_1 x_2$$

is not concave. But it is an increasing transformation of

$$\hat{f}(x_1, x_2) = \ln(x_1) + \ln(x_2),$$

which is concave. Hence f is quasi-concave.

Not every quasi-concave function can be represented as an increasing transformation of a concave function. But such a representation is almost invariably possible in applications with explicit functional forms. And one can prove that any quasi-concave function defined on a *compact* set can be represented as an increasing transformation of a concave function. Similar comments apply to quasi-convexity and convexity.

Theorem 4. *Let f and the g_k be differentiable. Let x^* be feasible and suppose that the following hold.*

1. *There is a $\lambda^* \geq 0$ such that the KKT conditions hold at x^*.*

2. *g_k is quasi-convex for all k.*

3. *Either f is concave or f is quasi-concave and $Df(x^*) \neq 0$.*

Then x^* is a constrained maximum.

Moreover, if, in addition, f is strictly quasi-concave then x^* is the strict constrained maximum.

Proof. Consider any x such that $f(x) > f(x^*)$. I argue that x is not feasible.

For ease of notation, let $v = x - x^*$. I claim that $Df(x^*)v > 0$. If f is concave then $f(x) \leq Df(x^*)(x - x^*) + f(x^*)$, which implies the claim. Suppose, on the other hand, that f is merely quasi-concave. Since f is continuous, there is an $\varepsilon > 0$ such that for any such w on the unit sphere in $\mathbb{R}^N$, $f(x + \varepsilon w) > f(x^*)$. For any w, and for any $\theta \in (0, 1)$, quasi-concavity implies that $f(x^*) \leq f(\theta(x + \varepsilon w) + (1 - \theta)x^*) = f(x^* + \theta(x + \varepsilon w - x^*))$, or

$$f(x^* + \theta(x + \varepsilon w - x^*)) - f(x^*) \geq 0$$

Dividing by $\theta > 0$ and taking the limits as $\theta \downarrow 0$ implies,

$$Df(x^*)(x + \varepsilon w - x^*) \geq 0.$$

Hence

$$Df(x^*)(x - x^*) + \varepsilon Df(x^*)w \geq 0.$$

This holds for all w. Since $Df(x^*) \neq 0$, there is a w such that $Df(x^*)w < 0$. The claim follows.

Therefore, $Df(x^*)v > 0$. By KKT Condition 1,

$$\sum_k \lambda_k^* Dg_k(x^*)v > 0.$$

By KKT Condition 2, $\lambda_k^* = 0$ if $k \notin J$ (the set of indices of the binding constraints). Therefore, there is at at least one $k \in J$ such that $Dg_k(x^*)v > 0$. But for this to be true, it must be true that for all $\theta > 0$ sufficiently small, $g_k(x^* + \theta v) > g_k(x^*) = 0$, where the equality comes from the fact that $k \in J$. But quasi-convexity of g_k then implies that $g_k(x) > 0$, implying that x is not feasible. It follows by contraposition that x^* is a constrained maximum.

Moreover, suppose that f is strictly quasi-concave and that there is a feasible x with $f(x) = f(x^*)$. Then, by the definition of strict quasi-concavity, $f(\theta x + (1 - \theta)x^*) > f(x^*)$ if $x \neq x^*$. Since $\theta x + (1 - \theta)$ is feasible (by quasi-convexity of the g_k) and x^* is maximal, it follows by contraposition that $x = x^*$: x^* is the unique maximum. ■

For minimization problems, a similar theorem holds but with f quasi-convex and the g_k quasi-concave.

The conclusion of Theorem 4 can fail if the regularity condition $Df(x^*) \neq 0$ is violated.

Example 1. Suppose $N = K = 1$, $f(x) = x^3$, $g(x) = 0$. Then f is quasi-concave (for functions on $\mathbb{R}$, any monotone function is quasi-concave) and g is convex. At the point $x^* = 0$, KKT is satisfied for any $\lambda \geq 0$ since $\nabla f(x^*) = \nabla g(x^*) = 0$. But x^* is not a solution; in fact, there is no solution. □

2.2 Second Order Sufficiency Conditions with Constraints.

Given x^*, one can prove (this is not obvious) that a sufficient condition for f to be locally strictly quasi-concave is that f is *differentially strictly quasi-concave* at x^*, which, by definition, means that $v'D^2f(x^*)v < 0$ for all $v \in T_f = \{v \in \mathbb{R}^N : v \cdot \nabla f(x^*) = 0\}$. T_f is an $N-1$ dimensional subspace of $\mathbb{R}^N$. Geometrically, f is differentially quasi-concave at x^* if it is differentially strictly concave at x^* for directions that are tangent to the level set of f (hence the T_f notation; T for tangent). f need not be concave in other directions. If $\nabla f(x^*) = 0$ (as at an unconstrained maximum) then $T_f = \mathbb{R}^N$ and differential strict quasi-concavity is the same as differential strict concavity.

Example 2. Define $f : \mathbb{R}^2_{++} \to \mathbb{R}$ by $f(x) = x_1 x_2$. Let $x^* = (1, 1)$. Then

$$\nabla f(x^*) = (x_2^*, x_1^*) = (1, 1)$$
$$D^2 f(x^*) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

The set $\{v \in \mathbb{R}^N : v \cdot \nabla f(x^*) = 0\}$ is spanned by $v = (-1, 1)$ and

$$(-1, 1)' D^2 f(x^*)(-1, 1) = -2 < 0,$$

so f is differentially strictly quasi-concave at x^*. On the other hand, $D^2 f(x^*)$ is not negative definite, or even negative semi-definite. In particular,

$$(1, 1)' D^2 f(x^*)(1, 1) = 2 > 0.$$

□.

Similarly, a sufficient condition for g_k to be locally quasi-convex is that g_k is *differentially strictly quasi-convex* at x^*, which, by definition, means that $v' D^2 g_k(x^*) v > 0$ for all $v \in T_{g_k} = \{v \in \mathbb{R}^N : v \cdot \nabla g_k(x^*) = 0\}$, $v \neq 0$.

Theorem 4 then implies the following result for constrained local maxima.

Theorem 5. *Let $f : \mathbb{R}^N \to \mathbb{R}$ and $g_k : \mathbb{R}^N \to \mathbb{R}$, $k \in \{1, \ldots, K\}$, be $\mathcal{C}^2$. Given a feasible point x^*, suppose that the following conditions hold.*

1. *There is a $\lambda^* \geq 0$ such that the KKT conditions hold.*
2. *g_k is differentially strictly quasi-convex at x^*, for any k binding.*
3. *f is differentially strictly quasi-concave at x^*.*

Then x^* is a local strict maximum.

For intuition for Theorem 5, and also for Theorem 4, suppose that only one constraint is binding, say g_1, and that $\nabla f(x^*) \neq 0$. KKT Condition 1 requires that there is a $\lambda_1^* \geq 0$ such that $\nabla f(x^*) = \lambda_1^* \nabla g(x^*)$; the assumption that $\nabla f(x^*) \neq 0$ implies that $\lambda_1^* > 0$. To check whether x^* is a local maximum, one can ignore movements v away from x^* that are either infeasible or that are known to decrease f. In particular, one can ignore v such that $v \cdot f(x^*) > 0$, even though such a v increases f, since KKT Condition 1 implies that $v \cdot g_1(x^*) > 0$: movement in this direction is not feasible. Hence one can restrict attention to v such that $v \cdot f(x^*) \leq 0$. On the other hand, one can ignore v for which $v \cdot f(x^*) < 0$, since such a v decreases f. This leaves $v \in T_f$. Moreover, by KKT Condition 1, $T_f = T_{g_1}$. Intuitively, then, it should suffice to check differential strict concavity of f and differential strict convexity of g_1 only for $v \in T_f = T_{g_1}$, and this is precisely what Theorem 5 says.

As this example suggests, if more constraints are binding then one can weaken the sufficiency condition even further. For concreteness, suppose that two constraints are binding, g_1 and g_2, that $\nabla f(x^*) \neq 0$, that $\nabla g_1(x^*), \nabla g_2(x^*)$ are linearly independent, and that $\lambda_1^*, \lambda_2^* > 0$. Then any direction v such that $v \cdot \nabla g_1(x^*) > 0$ or $v \cdot g_2(x^*) > 0$ is infeasible. On the other hand, KKT Condition 1 implies that if $v \cdot \nabla g_1(x^*) \leq 0$ and $v \cdot \nabla g_2(x^*) \leq 0$, with at least one inequality strict, then $v \cdot f(x^*) < 0$. Therefore, it should suffice to check differential strict concavity/convexity only for $v \in T_{g_1} \cap T_{g_2}$ which, by KKT Condition 1, is a proper subset of T_f.

And so on. Subject to regularity conditions (e.g., linear independence of the binding $\nabla g_k(x^*)$; the assumption that $\lambda_k^* > 0$ if k is binding), each binding constraint lowers the dimension of the subspace of v that needs to be checked by one. In the extreme case that the regularity conditions hold and there are N binding constraints, the second order sufficiency condition is satisfied vacuously.

One can pursue this type of reasoning to formulate sufficient conditions that are weaker than those of Theorem 5 and are close to necessary, much as the sufficient conditions of Theorem 2 are close to necessary. The precise formulation is somewhat cumbersome. In practice, Theorem 5 is typically strong enough.