ARE211, Fall 2009

NPP2: TUE, NOV 10, 2009 PRINTED: AUGUST 27, 2009 (LEC# 22)

Contents

6. Nonlinear Programming Problems and the Kuhn Tucker conditions (cont)

Key points:

(1) Understanding the problem of the vanishing gradient

(2) Defn of pseudo-concavity: f is pseudo-concave if

$$\forall \mathbf{x}, \mathbf{x}' \in X, \text{ if } f(\mathbf{x}') > f(\mathbf{x}) \text{ then } \bigtriangledown f(\mathbf{x}) \cdot (\mathbf{x}' - \mathbf{x}) > 0.$$

(3) Pseudo-concavity and its relationship to quasi-concavity:

if f is $\mathbb{C}^2$ then f is pseudo-concave iff f is quasi-concave and if $\bigtriangledown f(\cdot) = 0$ at $\mathbf{x}$ implies $f(\cdot)$ attains a global max at $\mathbf{x}$.

(4) The principal minor representation of negative definiteness subject to constraint:

for all $\mathbf{x}$, and all $k = 1, ..., n$, the sign of the k'th leading principal minor of the following bordered matrix must have the same sign as $(-1)^k$: $\begin{bmatrix} 0 & \triangledown f(\mathbf{x})' \\ \triangledown f(\mathbf{x}) & \mathrm{Hf}(\mathbf{x}) \end{bmatrix}$. where the k'th leading principal minor of this matrix is the determinant of the top-left $(k+1) \times (k+1)$ submatrix.

(5) Sufficient conditions for a solution to the NPP:

If f is pseudo-concave and the g^j's are quasi-convex, then a sufficient condition for a solution to the NPP at $\bar{\mathbf{x}} \in \mathbb{R}^m_+$ is that there exists a vector $\bar{\boldsymbol{\lambda}} \in \mathbb{R}^m_+$ such that

$$\triangledown f(\bar{\mathbf{x}})^T = \boldsymbol{\lambda}^T J g(\bar{\mathbf{x}})$$

and $\bar{\boldsymbol{\lambda}}$ has the property that $\bar{\lambda}_j = 0$, for each j such that $g^j(\bar{\mathbf{x}}) < b_j$.

(6) understanding the role of second order conditions in the sufficiency argument

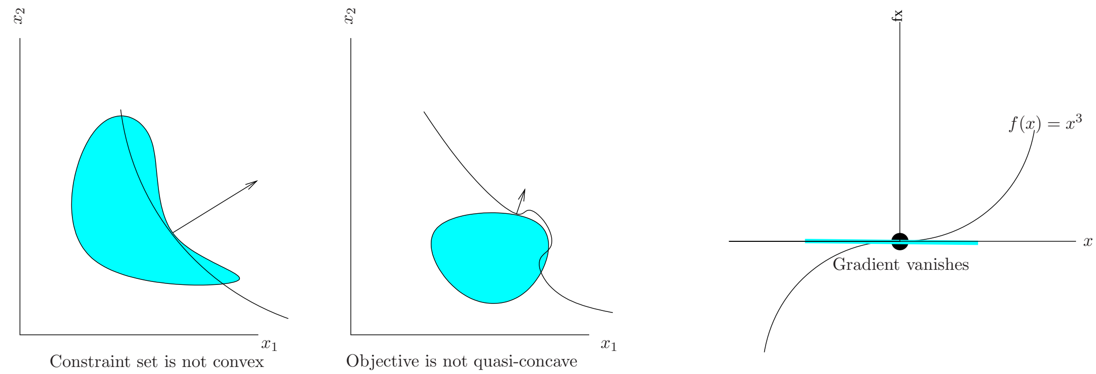

Figure 1. Three examples where KKT conditions are not sufficient for a soln

6.2. **Sufficient conditions for a solution to an NPP: Preliminaries**

So far we've only established necessary conditions for a solution to the NPP. Not surprisingly, without further restrictions, the KTT conditions aren't sufficient for a solution. They may be satisfied at a *minimum* on the constraint set, or else at a local but not global max. In this lecture we focus on identifying restrictions we can impose on the objective and constraint functions which ensure that the KKT conditions will be both necessary *and* sufficient for a solution. A good place to start, in our search for restrictions is to assume that objective function f is strictly quasi-concave while the constraint functions are quasi-convex. (Since the lower contour sets of quasi-convex functions are convex, and the intersection of convex sets is convex, and the constraint set is an intersection of lower contour sets, the condition that the constraint functions are quasi-convex implies that the constraint *set* is a convex set.) This isn't quite good enough, as we will see.

Figure Fig. 1 illustrates why quasi-concavity of the objective and convexity of the constraint set are required as restrictions. It also illustrates why they are not enough to ensure that the a solution to the KKT is sufficient for a solution to the NPP.

- the left panel has a constraint set that isn't generated by quasi-convex functions; level set represents a strictly quasi-concave objective function; the KKT conditions are satisfied at a local tangency, but it's a local *minimum* on the north east boundary of the constraint set. there are points further away that give higher values for the objective function.

- the middle panel has a convex constraint set, but the objective isn't quasi-concave. In this case we have a local max on the constraint set that isn't a solution to the problem. By a solution we mean a *global* max on the constraint set.
- the right panel is a more subtle problem. The constraint set, $[-1, 1]$ is convex and compact. It's given by $g^1(x) = x \le 1$, $g^2(x) = -x \le -1$. The objective function $f(x) = x^3$ is strictly quasi-concave, but at the origin (represented by the big dot, the KKT is satisfied: i.e., $f'(x) = 0 = [0, 0] \cdot [1, -1] = 0$. But zero is clearly not a solution to the NPP. The problem in this example is referred to as the "vanishing gradient" problem, because the gradient vanishes at an x value that is not a global maximum.

These examples make clear that we cannot say that if the objective and constraint functions have the right "quasi" properties, then satisfying the KKT conditions is sufficient for a max. We will have to strengthen quasi-concavity just enough so that it excludes function such as $f(x) = x^3$.

6.2.1. *First preliminary: the problem of the vanishing gradient.* To avoid the problem in the right panel of Fig. 1, we could simply assume that f has a non-vanishing gradient. But this restriction throws the baby out with the bath-water: e.g., the problem max $\mathbf{x}(1 - \mathbf{x})$ s.t. $\mathbf{x} \in [0, 1]$ has a global max at 0.5, at which point the gradient vanishes. More generally, the non-vanishing gradient assumption excludes *any* differentiable function that attains a global maximum.

So we need a condition that implies quasi-concavity and also has the property that the gradient vanishes at $\mathbf{x}$ *only* if $\mathbf{x}$ is a global maximizer of the function. The following condition on f—called *pseudoconcavity* in S&B (the original name) and M.K.9 in MWG—does just precisely this

$$\forall \mathbf{x}, \mathbf{x}' \in X, \text{ if } f(\mathbf{x}') > f(\mathbf{x}) \text{ then } \triangledown f(\mathbf{x}) \cdot (\mathbf{x}' - \mathbf{x}) > 0. \tag{1}$$

Note that (1) says a couple of things. First, it says that a *necessary* condition for $f(\mathbf{x}') > f(\mathbf{x})$ is that $dx = (\mathbf{x}' - \mathbf{x})$ makes an acute angle with the gradient of f. (This looks very much like quasi-concavity). Second, it implies that

$$\text{if } \triangledown f(\cdot) = 0 \text{ at } \mathbf{x} \text{ then } f(\cdot) \text{ attains a global max at } \mathbf{x} \tag{2}$$

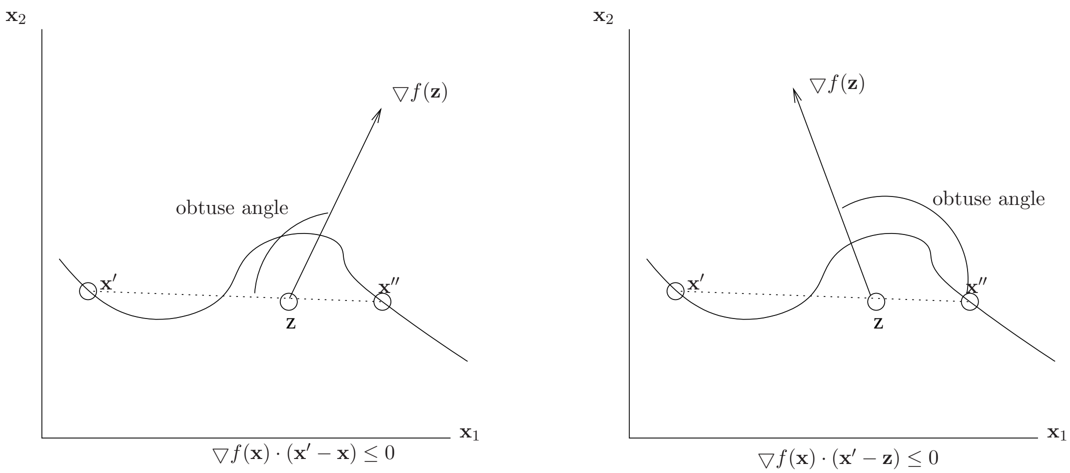

FIGURE 2. pseudo-concavity implies quasi-concavity

since if not then there would necessarily exist $\mathbf{x}, \mathbf{x}' \in X$ s.t. $f(\mathbf{x}') > f(\mathbf{x})$, and $\bigtriangledown f(\mathbf{x}) \cdot (\mathbf{x}' - \mathbf{x}) = \mathbf{0} \cdot (\mathbf{x}' - \mathbf{x}) = 0$, violating (1).

Our next result establishes precisely the relationship between pseudo-concavity and quasi-concavity:

$$\text{if } f \text{ is } \mathbb{C}^2 \text{ then } f \text{ is pseudo-concave iff } f \text{ is quasi-concave and satisfies (2)} \tag{3}$$

To prove the $\Longrightarrow$ direction of (3), we'll show that (2) together with ($\neg$ quasi-concavity) implies ($\neg$ pseudo-concavity).

Suppose that f satisfies (2) but is not quasi-concave, i.e., $\exists \mathbf{x}', \mathbf{x}'', \mathbf{z} \in X$ such that $\mathbf{z} = \lambda \mathbf{x}' + (1-\lambda)\mathbf{x}''$, for some $\lambda \in (0,1)$ and $f(\mathbf{x}'') \geq f(\mathbf{x}') > f(\mathbf{z})$. We will show that f violates (1) at a point $\mathbf{x}$ near $\mathbf{z}$. Since f does not obtain a global maximum at $\mathbf{z}$, (2) implies that $\bigtriangledown f(\mathbf{z}) \neq 0$. Therefore by continuity, we can pick $\epsilon > 0$ sufficiently small and $\mathbf{x} = \mathbf{z} + \epsilon \bigtriangledown f(\mathbf{z})$ such that $f(\mathbf{x}) < f(\mathbf{x}')$ and $\bigtriangledown f(\mathbf{x}) \cdot \bigtriangledown f(\mathbf{z}) > 0$. As Fig. 2 indicates, there are now two cases to consider: the angle between $\bigtriangledown f(\mathbf{x})$ and $\mathbf{x}' - \mathbf{x}''$ is either $\leq 90°$ or $> 90°$.

First suppose that $\bigtriangledown f(\mathbf{x}) \cdot (\mathbf{x}' - \mathbf{x}'') \leq 0$. In this case, we have

$$\begin{aligned} \bigtriangledown f(\mathbf{x}) \cdot (\mathbf{x}' - \mathbf{x}) &= \bigtriangledown f(\mathbf{x}) \cdot (\mathbf{x}' - (\mathbf{z} + \epsilon \bigtriangledown f(\mathbf{z}))) = \bigtriangledown f(\mathbf{x}) \cdot ((1-\lambda)(\mathbf{x}' - \mathbf{x}'') - \epsilon \bigtriangledown f(\mathbf{z}))) \\ &= (1-\lambda) \underbrace{\bigtriangledown f(\mathbf{x}) \cdot (\mathbf{x}' - \mathbf{x}'')}_{\leq 0 \text{ by assumption}} - \epsilon \underbrace{\bigtriangledown f(\mathbf{x}) \bigtriangledown f(\mathbf{z})}_{>0 \text{ by construction}} < 0 \end{aligned}$$

On the other hand, if $\bigtriangledown f(\mathbf{x}) \cdot (\mathbf{x}'' - \mathbf{x}') < 0$, repeat the above argument to conclude that $\bigtriangledown f(\mathbf{x}) \cdot (\mathbf{x}'' - \mathbf{x}) < 0$. We have thus identified a point $\mathbf{x}$ such that $f(\mathbf{x}'') \geq f(\mathbf{x}') > f(\mathbf{x})$ such that *either* $\bigtriangledown f(\mathbf{x}) \cdot (\mathbf{x}' - \mathbf{x}) < 0$ or $\bigtriangledown f(\mathbf{x}) \cdot (\mathbf{x}'' - \mathbf{x}) < 0$, verifying that (1) is violated.

To prove the $\Leftarrow$ direction of (3), we'll show that (2) together with ($\neg$ pseudo-concavity) implies ($\neg$ quasi-concavity). Assume that there exists $\mathbf{x}, \mathbf{x}'$ such that $f(\mathbf{x}') > f(\mathbf{x})$ but $\bigtriangledown f(\mathbf{x}) \cdot (\mathbf{x}' - \mathbf{x}) \leq 0$.. Since $\mathbf{x}$ is not a global maximizer of f, (2) implies that $\bigtriangledown f(\mathbf{x}) \neq 0$. By continuity, we can pick $\epsilon > 0$ sufficiently small that for $\mathbf{y} = \mathbf{x}' - \epsilon \bigtriangledown f(\mathbf{x})$, $f(\mathbf{y}) > f(\mathbf{x})$. We'll show that a portion of the line segment joining $\mathbf{y}$ and $\mathbf{x}$ does not belong to the upper contour set of f corresponding to $f(\mathbf{x})$, proving that f is not quasi-concave. We have

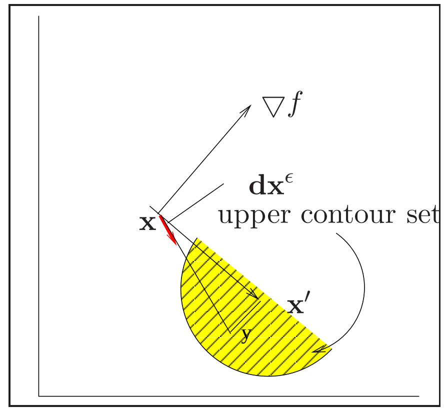

FIGURE 3. pseudo-concavity rules this out

$$\begin{aligned}
\bigtriangledown f(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}) &= \bigtriangledown f(\mathbf{x}) \cdot ((\mathbf{x}' - \epsilon \bigtriangledown f(\mathbf{x})) - \mathbf{x}) \\
&= \underbrace{\bigtriangledown f(\mathbf{x}) \cdot (\mathbf{x}' - \mathbf{x})}_{\leq 0 \text{ by assumption}} - \epsilon \underbrace{||\bigtriangledown f(\mathbf{x})||^2}_{>0} \quad < \quad 0
\end{aligned}$$

Let $\mathbf{dx}^\epsilon = \epsilon(\mathbf{y} - \mathbf{x})$. For all ϵ, $\bigtriangledown f(\mathbf{x}) \cdot \mathbf{dx}^\epsilon = \epsilon \bigtriangledown f(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}) < 0$. Now, by Taylor-Young's theorem, if $\mathbf{dx} \neq 0$ is sufficiently small, then the sign of $(f(\mathbf{x} + \mathbf{dx}^\epsilon) - f(\mathbf{x}))$ is the same as the sign of $\bigtriangledown f(\mathbf{x}) \cdot \mathbf{dx}^\epsilon$, i.e., $f(\mathbf{x} + \mathbf{dx}^\epsilon) < f(\mathbf{x})$. We have now established that a portion of the line segment joining $\mathbf{x}$ and $\mathbf{y}$ does not belong to the upper contour set of f corresponding to $f(\mathbf{x})$. □

The following modification, changing only the first strict inequality to a weak inequality, gives us a condition that implies *strict* quasi-concavity.

$$\forall \mathbf{x}, \mathbf{x}' \in X \text{ s.t. } f(\mathbf{x}') \geq f(\mathbf{x}), \bigtriangledown f(\mathbf{x}) \cdot (\mathbf{x}' - \mathbf{x}) > 0. \tag{1'}$$

Conclude that pseudo-concavity is a much weaker assumption than the non-vanishing gradient condition, and will give us just enough to ensure that the KT conditions are not only necessary but sufficient as well. In particular, pseudo-concavity admits the possibility that our solution to the NPP may be unconstrained.

6.2.2. *Second preliminary: quasi-concavity and the Hessian of f.* Recall from earlier that a condition sufficient to ensure that f is strictly (weakly) concave is to require that the Hessian of f be everywhere negative (semi) definite. Analogously, as we've seen in the Calculus section, a sufficient condition for f is strict (weak) quasi-concavity is a weaker "definiteness subject to constraint" property for f. The following result, which is a tiny bit stronger than the one we proved earlier, gives a sufficient condition for strict quasi concavity. Since we've already proved the earlier theorem, we won't bother to prove this variant.

Theorem (SQC): A sufficient condition for $f : \mathbb{R}^n \to \mathbb{R}$ to be strictly quasi-concave is that for all $\mathbf{x}$ and all $\mathbf{dx}$ such that $\bigtriangledown f(\mathbf{x})'\mathbf{dx} = 0$, $\mathbf{dx}'\mathrm{Hf}(\mathbf{x})\mathbf{dx} < 0$.

The following condition on the leading principal minors of the *bordered* hessian of f is equivalent to this "definiteness subject to constraint" property. *for all* $\mathbf{x}$, and all $k = 1, ..., n$, the sign of the k'th leading principal minor of the following bordered matrix must have the same sign as $(-1)^k$: $\begin{bmatrix} 0 & \bigtriangledown f(\mathbf{x})' \\ \bigtriangledown f(\mathbf{x}) & \mathrm{Hf}(\mathbf{x}) \end{bmatrix}$. where the k'th leading principal minor of this matrix is the determinant of the top-left $(k+1) \times (k+1)$ submatrix. We emphasize yet again that strict quasi-concavity is a *global* property, so that this leading principal minor property has to hold *for all* $\mathbf{x}$ in the domain of the function in order to guarantee strict quasi-concavity.

Note also that the above condition isn't *necessary* for strict quasi-concavity: the usual example, $f(x) = x^3$, establishes this: f is strictly quasi-concave, but at $\bar{\mathbf{x}} = 0$, and all $\mathbf{dx}$, $\bigtriangledown f(\bar{\mathbf{x}})\mathbf{dx} = 0$, while $\mathbf{dx}'\mathrm{Hf}(\bar{\mathbf{x}})\mathbf{dx} = 0$.

6.2.3. *Third preliminary: "Definiteness subject to constraint" and sufficiency.* A sufficient condition for strict concavity is that for all $\mathbf{x}$, $\mathbf{dx}'\mathrm{Hf}(\bar{\mathbf{x}})\mathbf{dx} < 0$, and all $\mathbf{dx} \neq 0$. For strict quasi-concavity, we only require this property of the Hessian holds for vectors that are *orthogonal to* $\bigtriangledown(\mathbf{x})$. Similarly, for g to be strictly quasi-convex, we only require that $\mathbf{dx}'\mathrm{Hg}(\bar{\mathbf{x}})\mathbf{dx} > 0$, and all $\mathbf{dx} \neq 0$ such that $\bigtriangledown g(\mathbf{x}) = 0$. These conditions are much weaker, infinitely weaker in fact, than the conditions for concavity and convexity. However, they are not quite as weak as they look: the condition on orthogonal vectors also has implications for $\mathbf{dx} \neq 0$'s that are *almost* orthogonal to $\bigtriangledown(\mathbf{x})$, and we need these implications in order to prove that the KT conditions are sufficient for a solution. Specifically, if $\bigtriangledown f(\bar{\mathbf{x}})\mathbf{dx} = 0$ implies $\mathbf{dx}'\mathrm{Hf}(\bar{\mathbf{x}})\mathbf{dx} < 0$, then by continuity,

$$\text{for } \mathbf{dx} \neq 0 \text{ such that } |\bigtriangledown f(\bar{\mathbf{x}})\mathbf{dx}| \text{ is sufficiently small, } \mathbf{dx}'\mathrm{Hf}(\bar{\mathbf{x}})\mathbf{dx} < 0 \tag{4a}$$

similarly, if $\bigtriangledown g(\bar{\mathbf{x}})\mathbf{dx} = 0$ implies $\mathbf{dx}'\mathrm{Hg}(\bar{\mathbf{x}})\mathbf{dx} > 0$, then

$$\text{for } \mathbf{dx} \neq 0 \text{ such that } |\bigtriangledown g(\bar{\mathbf{x}})\mathbf{dx}| \text{ is sufficiently small, } \mathbf{dx}'\mathrm{Hg}(\bar{\mathbf{x}})\mathbf{dx} > 0 \tag{4b}$$

Why is (4) so important? The problem we face, once again, is that in order to have a local maximum, you need an *ϵ-ball* around $\mathbf{x}$ such that $f(\cdot)$ is strictly less than $f(\mathbf{x})$ on the intersection of this ball with the constraint set. Now as we've discussed over and over again, you can't find this ball just by using first order conditions. You need your *second order conditions* to be cooperative in the region where the first order conditions fail you. If they are sufficiently uncooperative, i.e., if the signs of the quadratic terms in (4) are all reversed, then for any given ϵ-ball, there are going to be $\mathbf{dx}$'s that

- make an angle close to 90° with both $\bigtriangledown f(\mathbf{x})$ and $\bigtriangledown g(\mathbf{x})$, resulting in almost zero inner products $\bigtriangledown f(\mathbf{x})\mathbf{dx}$ and $\bigtriangledown g(\mathbf{x})\mathbf{dx}$, which are dominated by
- a positive second order term for f, resulting in a net increase in f, and
- a negative second order term for g, resulting in a net *decrease* in g, so you remain in the constraint set.
- in which case you don't have a maximum on the constraint set.

On the other hand, suppose that the point $\mathbf{x}$ in Fig. 4 satisfies the KKT conditions for the canonincal NPP with one constraint, i.e., $\max f(\mathbf{x})$ s.t. $g(\mathbf{x}) \leq b$. We will make the following assumptions:

(1) the KKT conditions are satisfied;

(2) the constraint is satisfied with equality;

(3) the two parts of (4) are satisfied for any vector $\mathbf{dx}$ in the double cone labeled C.

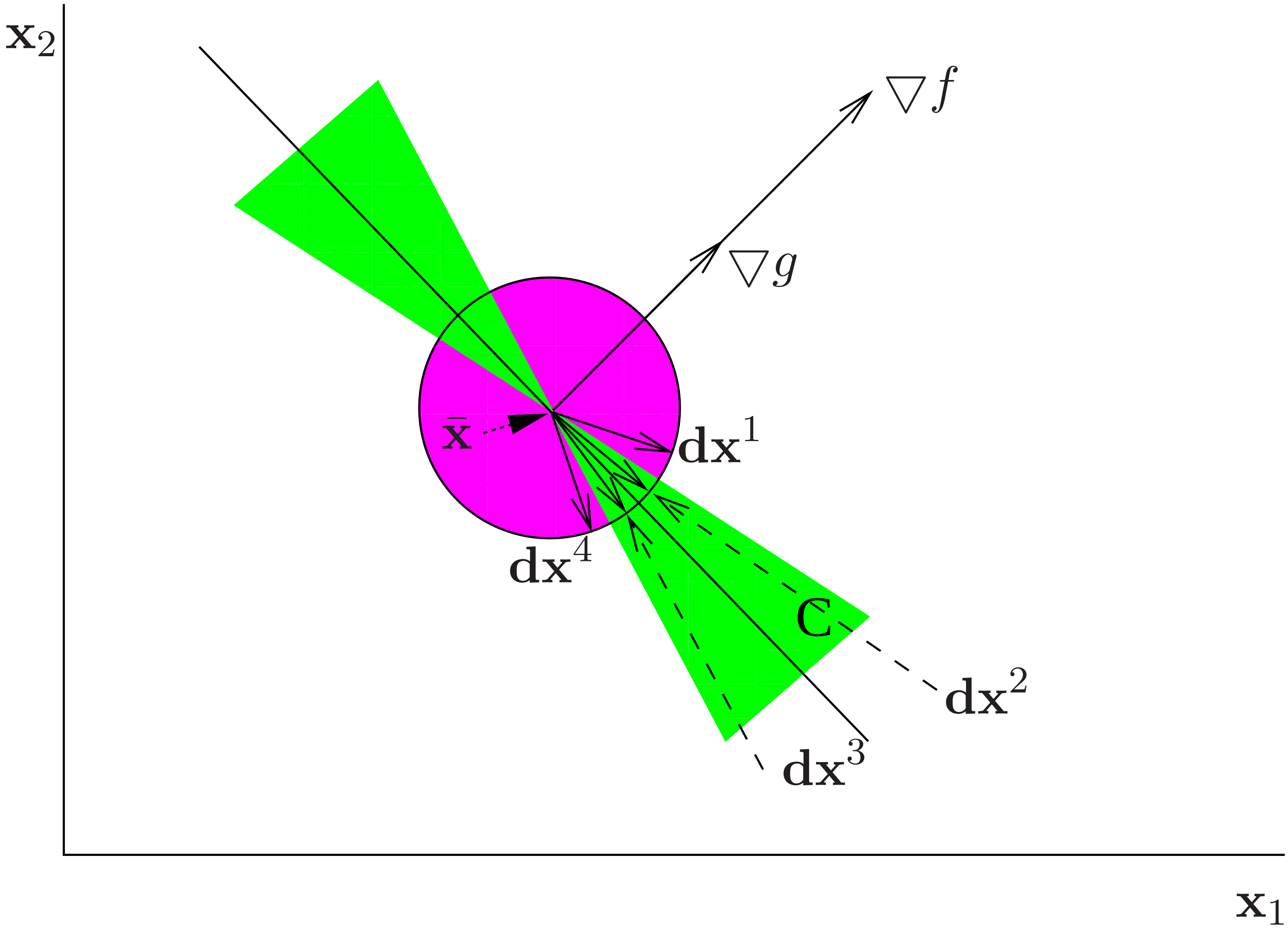

FIGURE 4. Second order conditions and sufficiency

(4) for any vector $\mathbf{dx}$ in the shaded circle but *not* in the set C, the first order terms in the Taylor expansions of f and g dominate, i.e., the signs of the first order terms $\bigtriangledown f(\mathbf{x})\mathbf{dx}$ and $\bigtriangledown g(\mathbf{x})\mathbf{dx}$ agree with, respectively, $f(\mathbf{x}+\mathbf{dx})-f(\mathbf{x})$ and $g(\mathbf{x}+\mathbf{dx})-g(\mathbf{x})$..

The following four part argument is a very informal sketch of the proof that $\mathbf{x}$ solves the constrained maximization problem.

(1) For a vector such as $\mathbf{dx}^1$ which makes an acute angle with $\bigtriangledown g(\mathbf{x})$, but does not belong to C, the positive first order term in the expansion of g dominates, so that $g(\mathbf{x}+\mathbf{dx}) > g(\mathbf{x})$, implying that $\mathbf{x}+\mathbf{dx}$ does not belong to the constraint set.

(2) For a vector such as $\mathbf{dx}^4$ which makes an obtuse angle with $\bigtriangledown f(\mathbf{x})$, but does not belong to C, the negative first order term in the expansion of f dominates, so that $f(\mathbf{x}+\mathbf{dx}) < f(\mathbf{x})$.

(3) For a vector such as $\mathbf{dx}^2 \in C$ which makes a near $90°$ acute angle with $\bigtriangledown g(\mathbf{x})$, the second order term $0.5\mathbf{dx}'\mathrm{Hg}(\bar{\mathbf{x}})\mathbf{dx} > 0$ reinforces rather than offsets the negligible first order term, ensuring that $\mathbf{x}+\mathbf{dx}$ does not belong to the constraint set.

(4) For a vector such as $\mathbf{dx}^3 \in C$ which makes a near $90°$ obtuse angle with $\bigtriangledown f(\mathbf{x})$, the second order term $0.5\mathbf{dx}'\mathrm{Hf}(\bar{\mathbf{x}})\mathbf{dx} < 0$ reinforces rather than offsets the negligible first order term, ensuring that $f(\mathbf{x}+\mathbf{dx}) < f(\mathbf{x})$.

How small is "sufficiently small" in (4)? The following example shows that the requirement for sufficiently small gets tougher and tougher, the less concave is f. Example: Consider the function $f(\mathbf{x},\mathbf{y}) = (\mathbf{xy})^\beta$, which is strictly quasi-concave but not concave for $\beta > 0.5$. We'll illustrate that regardless of the value of β, $\mathbf{dx}'\mathrm{Hf}^\lambda\mathbf{dx} < 0$, for any vector that is *almost* orthogonal to $\bigtriangledown f$, but that the criterion of "almost" gets tighter and tighter as β gets larger. That is, the higher is β (i.e., the less concave is f), the closer to orthogonal does $\mathbf{dx}$ have to be in order to ensure that $\mathbf{dx}'\mathrm{Hf}^\lambda\mathbf{dx}$ is negative.

We have $\bigtriangledown f(\mathbf{x}_1,\mathbf{x}_2) = \beta\left(\mathbf{x}_1^{\beta-1}\mathbf{x}_2^{\beta}, \mathbf{x}_1^{\beta}\mathbf{x}_2^{\beta-1}\right)$ and $\mathrm{Hf}(\mathbf{x}_1,\mathbf{x}_2) = \beta\begin{bmatrix}(\beta-1)\mathbf{x}_1^{\beta-2}\mathbf{x}_2^{\beta} & \beta\mathbf{x}_1^{\beta-1}\mathbf{x}_2^{\beta-1}\\ \beta\mathbf{x}_1^{\beta-1}\mathbf{x}_2^{\beta-1} & (\beta-1)\mathbf{x}_1^{\beta}\mathbf{x}_2^{\beta-2}\end{bmatrix}$.

Evaluated at $(\mathbf{x}_1,\mathbf{x}_2) = (1,1)$, we have $\mathrm{Hf}(1,1) = \beta^2\begin{bmatrix}\frac{\beta-1}{\beta} & 1\\ 1 & \frac{\beta-1}{\beta}\end{bmatrix} = \beta^2\begin{bmatrix}\lambda & 1\\ 1 & \lambda\end{bmatrix}$, where $\lambda = \frac{\beta-1}{\beta}$.

Note that $\lambda \to 1$ as $\beta \to \infty$.

Now choose a unit length vector $\mathbf{dx}$ and consider

$$\mathbf{dx}'\mathrm{Hf}^\lambda\mathbf{dx} = \lambda(dx_1^2+dx_2^2)+2dx_1dx_2 = (dx_1+dx_2)^2-(1-\lambda)(dx_1^2+dx_2^2) = (dx_1+dx_2)^2-(1-\lambda)$$

For $\mathbf{dx}$ such that $\mathbf{x}_1 = -\mathbf{x}_2$, $\mathbf{dx}'\mathrm{Hf}^\lambda\mathbf{dx} < 0$, for all $\lambda < 1$, verifying that f is strictly quasi-concave. However, the closer is λ to unity, the smaller is the set of unit vectors for which $\mathbf{dx}'\mathrm{Hf}^\lambda\mathbf{dx} < 0$.

6.3. **Sufficient Conditions for a solution to the NPP: the theorem**

The following theorem gives *sufficient* conditions for a solution (not necessarily unique) to the NPP.

Theorem (S): (*Sufficient* conditions for a solution to the NPP): If f is pseudo-concave and the g^j's are quasi-convex, then a sufficient condition for a solution to the NPP at $\bar{\mathbf{x}} \in \mathbb{R}^m_+$ is that there exists a vector $\bar{\boldsymbol{\lambda}} \in \mathbb{R}^m_+$ such that

$$\bigtriangledown f(\bar{\mathbf{x}})^T = \boldsymbol{\lambda}^T Jg(\bar{\mathbf{x}})$$

and $\bar{\boldsymbol{\lambda}}$ has the property that $\bar{\lambda}_j = 0$, for each j such that $g^j(\bar{\mathbf{x}}) < b_j$.

Note that Theorem (S) doesn't guarantee that a solution exists. Need compactness for this. Note also that the sufficient conditions are like the necessary conditions, except that you don't need the CQ but do need pseudo-concavity and quasi-convexity. (MWG's version of Theorem (S)—Theorem M.K.3—is just like mine except that they *do* include the constraint qualification. This addition is unnecessary (they're not wrong, they just have a meaningless additional condition). The C.Q. says that you can have a maximum without the non-negative cone condition holding. If you assume as in (S) that the nonnegative cone property holds, then, obviously, you don't need to worry that perhaps it mightn't hold!

Sketch of the proof of Theorem (S) for the case of one constraint: Suppose that the KT conditions are satisfied at $(\bar{\mathbf{x}}, \bar{\lambda})$, along with the other conditions of the sufficiency theorem, i.e., pseudo-concave objective and quasi-convex constraint functions.

First note that if $\bar{\lambda}$ is zero, then $\bigtriangledown f(\bar{\mathbf{x}})$ is zero also. But then we're done, because by (2), f must attain a *global* and hence constrained max at $\bar{\mathbf{x}}$. Now assume that $\bar{\lambda} > 0$. This in turn (complementary slackness) implies that $\bar{\mathbf{x}}$ is on the boundary of the constraint set, i.e., $g(\bar{\mathbf{x}}) = \mathbf{b}$. In this case, the Kuhn Tucker conditions say that $\bigtriangledown f(\bar{\mathbf{x}})$ must be pointing in the same direction as $\bigtriangledown g(\bar{\mathbf{x}})$.

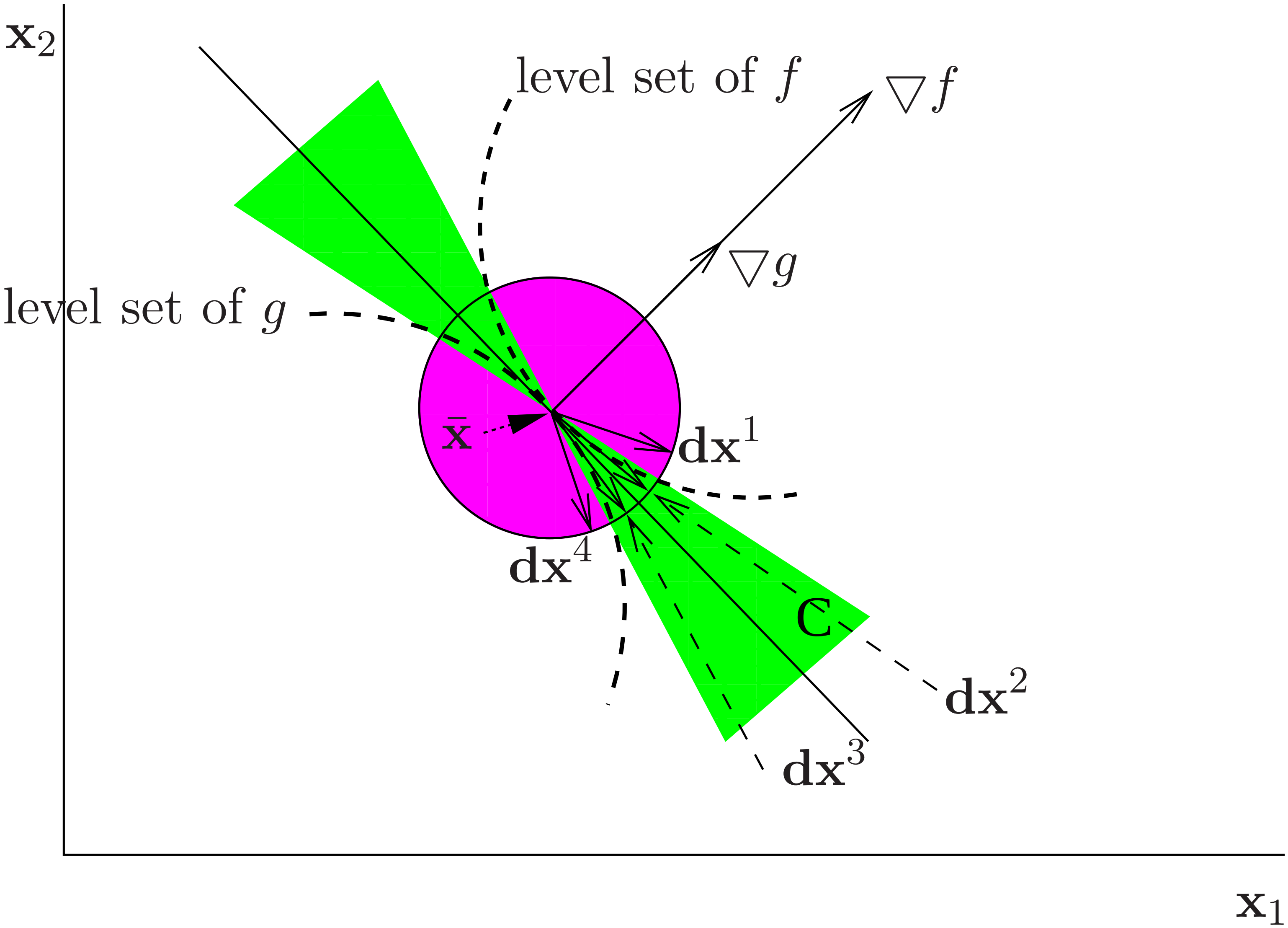

FIGURE 5. Sufficient conditions for a solution to the NPP

How does this guarantee us a max? We'll show that the KT conditions *plus* pseudo- and hence quasi-concavity (applied locally) guarantee a *local* max, and the quasi-concavity/quasi-convexity conditions (applied globally) do the rest.

- To establish a strictly local max, we have to show that there exists a ball about $\bar{\mathbf{x}}$ such that no matter where we move within that ball, we *either* decrease the value of f *or* move outside the constraint set. To find the right ball, we proceed as follows

(1) Since f and g are, respectively, strictly quasi-concave and strictly quasi-convex, we know that $\mathbf{dx}'\mathrm{Hf}\,\mathbf{dx} < 0$ and $\mathbf{dx}'\mathrm{Hg}\,\mathbf{dx} > 0$, for $\mathbf{dx}$'s that are orthogonal to the direction that the gradients of f and g both point in. By continuity, there exists an interval around 90 degrees such that for any vector $\mathbf{dx}$ that makes an angle in this interval with $\bigtriangledown f$ (i.e., lives in the cone-shaped object C in Fig. 5), $\mathbf{dx}'\mathrm{Hf}\,\mathbf{dx} < 0$, while $\mathbf{dx}'\mathrm{Hg}\,\mathbf{dx} > 0$.

(2) Next note that there exists $\epsilon^1 > 0$ such that for $\mathbf{dx}$'s in $B(\mathbf{x}, \epsilon^1)$ *but not in the cone-like set* C the first term in the Taylor expansion about $\mathbf{x}$ in the direction $\mathbf{dx}$ determines the sign of the entire expansion. *You should, hopefully, understand well, by now, why without excluding the cone we couldn't find such an* ϵ*.!*

(3) Finally, note that there exists $0 < \epsilon^2 \leq \epsilon^1$ such that for $\mathbf{dx}$'s in $B(\mathbf{x}, \epsilon^2)$ *and* in the cone-like set C, the first *two terms* in the Taylor expansion about $\mathbf{x}$ in the direction $\mathbf{dx}$ determines the sign of the entire expansion.

There are four classes of directions to consider:

(1) if we move from $\bar{\mathbf{x}}$ in a direction that makes a *big enough acute* angle with the gradient vector, $\bigtriangledown g(\bar{\mathbf{x}})$ (e.g., $\mathbf{dx}^1$), then by the *first order* version of Taylor's theorem, we *increase* the value of g, i.e., move outside the constraint set.

(2) if we move from $\bar{\mathbf{x}}$ in a direction that makes a *big enough obtuse* angle with the gradient vector, $\bigtriangledown f(\bar{\mathbf{x}})$ (e.g., $\mathbf{dx}^4$), then by the *first order* version of Taylor's theorem, enough we *reduce* the value of f,

(3) if we move from $\bar{\mathbf{x}}$ in a direction $\mathbf{dx}$ that makes an *barely acute* angle with the gradient vector, $\bigtriangledown g(\bar{\mathbf{x}})$, i.e., we move in a direction almost perpendicular to the gradient vector, (e.g., $\mathbf{dx}^2$), by the *second order* version of Taylor's theorem, enough we *increase* the value of g (since both of the first two terms in the expansion of g are positive).

(4) if we move from $\bar{\mathbf{x}}$ in a direction $\mathbf{dx}$ that makes an *barely obtuse* angle with the gradient vector, $\bigtriangledown f(\bar{\mathbf{x}})$, i.e., we move in a direction almost perpendicular to the gradient vector, (e.g., $\mathbf{dx}^3$), by the *second order* version of Taylor's theorem, enough we *reduce* the value of f (since both of the first two terms in the expansion of f are negative).

So we've shown that if we move a distance less than ϵ^2 in any possible direction away from $\bar{\mathbf{x}}$, either f goes down or g goes up. Conclude that we have a local max.

- Of course, these derivative arguments only guarantee a *local* maximum on the constraint set. That is, there exists a neighborhood of $\bar{\mathbf{x}}$ such that f is at least as large at $\bar{\mathbf{x}}$ as it is anywhere on the intersection of this neighborhood with the constraint set.
- Quasi-concavity and quasi-convexity are needed to ensure that $\bar{\mathbf{x}}$ is indeed a *solution* to the NPP, i.e., a global max on the constraint set. *Quasi-concavity* says that if there is a point $\mathbf{x}'$ *anywhere* in the constraint set at which f attained a strictly higher value than at $\bar{\mathbf{x}}$, then f will be strictly higher than at $\bar{\mathbf{x}}$ on the whole line segment joining $\mathbf{x}'$ to $\bar{\mathbf{x}}$. Quasi-convexity guarantees that the constraint set is convex, so that if $\mathbf{x}'$ is in the constraint set, then so is the whole line segment. But this means that if there exists a point such as $\mathbf{x}'$, then on *any* neighborhood of $\bar{\mathbf{x}}$ there will be points at which f is strictly higher than it is at $\bar{\mathbf{x}}$, contradicting the fact that we have a local max.

We'll make the above informal argument precise, for the special case of maximizing $f(\mathbf{x})$ such that $g(\mathbf{x}) \leq b$, where f is strictly quasi-concave with non-vanishing gradient and $g : \mathbb{R}^n \to \mathbb{R}$ is *linear*. Further more we'll make the assumption that $\mathbf{v}'\mathrm{Hf}(\bar{\mathbf{x}})\mathbf{v} < 0$ *for all* $\mathbf{x}$ and all $\mathbf{v}$ such that $\bigtriangledown f(\mathbf{x})\mathbf{v} = 0$. Recall that this condition is sufficient for quasi-concavity, but not necessary. We don't want to have to deal with the case where it isn't satisfied.

Suppose that $\bar{\mathbf{x}}$ satisfies the KKT conditions, i.e., $\bigtriangledown f(\bar{\mathbf{x}})$ and $\bigtriangledown g(\bar{\mathbf{x}})$ are collinear. We need to show that there exists $\epsilon > 0$ such that for all $\mathbf{dx} \in B(\mathbf{0}, \epsilon)$, $g(\bar{\mathbf{x}} + \mathbf{dx}) \leq b$ implies $f(\bar{\mathbf{x}} + \mathbf{dx}) - f(\bar{\mathbf{x}}) < 0$. Let S denote the (hollow) unit hypersphere, i.e., $S = \{\mathbf{v} \in \mathbb{R}^n : ||\mathbf{v}|| = 1\}$. Recall from the lecture on Taylor expansions that if f is thrice continuously differentiable, then for all $\mathbf{x}$, for all $\mathbf{v} \in S$, there exists $\lambda(\mathbf{v}) \in [0, 1]$ such that

$$f(\bar{\mathbf{x}} + \mathbf{v}) - f(\bar{\mathbf{x}}) \quad = \quad \bigtriangledown f(\bar{\mathbf{x}})\mathbf{v} \quad + \quad \mathbf{v}'\mathrm{Hf}(\bar{\mathbf{x}})\mathbf{v}/2 \quad + \quad \underbrace{\mathrm{Tf}_3(\bar{\mathbf{x}} + \lambda(\mathbf{v})\mathbf{v}, \mathbf{v})/6}_{\text{a cubic remainder term}}$$

so that for $\epsilon \in (0, 1]$ and $\mathbf{dx} = \epsilon\mathbf{v}$,

$$f(\bar{\mathbf{x}} + \mathbf{dx}) - f(\bar{\mathbf{x}}) \quad = \quad \epsilon^2\Big(\bigtriangledown f(\bar{\mathbf{x}})\mathbf{v}/\epsilon \quad + \quad \mathbf{v}'\mathrm{Hf}(\bar{\mathbf{x}})\mathbf{v}/2 \quad + \quad \epsilon\mathrm{Tf}_3(\bar{\mathbf{x}} + \lambda(\epsilon\mathbf{v})\epsilon\mathbf{v}, \mathbf{v})\Big)$$

Because the remainder term is continuous,[1] it is bounded on the compact set $S\times[0,1]$. $|\mathbf{v}'\mathrm{Hf}(\bar{\mathbf{x}})\mathbf{v}|$ is bounded also. Therefore, there exists $\omega \geq 1$ such that for all $\mathbf{v}\in S$ and all $\epsilon\in[0,1]$, $|\mathbf{v}'\mathrm{Hf}(\bar{\mathbf{x}})\mathbf{v}/2| < \omega$ and $|\mathrm{Tf}_3(\bar{\mathbf{x}}+\lambda(\epsilon\mathbf{v})\epsilon\mathbf{v},\mathbf{v})/6| < \omega$.[2] Because f is strictly quasi-concave, there exists $\delta > 0$ such that $\bigtriangledown f(\bar{\mathbf{x}})\mathbf{v} = 0$ implies $\mathbf{v}'\mathrm{Hf}(\bar{\mathbf{x}})\mathbf{v}/2 < -2\delta$. Because $\bigtriangledown f(\bar{\mathbf{x}})$ and $\mathbf{v}'\mathrm{Hf}(\bar{\mathbf{x}})\mathbf{v}$ are continuous w.r.t. $\mathbf{v}$, there exists $\gamma > 0$ such that for all $\mathbf{v}\in S$, $|\bigtriangledown f(\bar{\mathbf{x}})\mathbf{v}| < \gamma$ implies $\mathbf{v}'\mathrm{Hf}(\bar{\mathbf{x}})\mathbf{v}/2 < -\delta$. Let $\bar{\epsilon} = \min[\gamma,\delta,1]/2\omega$ and observe that for all $\mathbf{v}\in S$ and all $0<\epsilon<\bar{\epsilon}$, if $\mathbf{dx} = \epsilon\mathbf{v}$ then

$$|\bigtriangledown f(\bar{\mathbf{x}})\mathbf{v}| \geq \gamma \text{ implies: } \begin{cases} |\bigtriangledown f(\bar{\mathbf{x}})\mathbf{v}/\epsilon| & > \quad 2\omega, \\ |\mathbf{v}'\mathrm{Hf}(\bar{\mathbf{x}})\mathbf{v}/2| & < \quad \omega, \\ \epsilon|\mathrm{Tf}_3(\bar{\mathbf{x}}+\lambda(\epsilon\mathbf{v})\epsilon\mathbf{v},\mathbf{v})/6| & < \quad \omega \end{cases}$$

$$\text{while } |\bigtriangledown f(\bar{\mathbf{x}})\mathbf{v}| < \gamma \text{ implies: } \begin{cases} |\mathbf{v}'\mathrm{Hf}(\bar{\mathbf{x}})\mathbf{v}/2| & > \quad \delta, \\ \epsilon|\mathrm{Tf}_3(\bar{\mathbf{x}}+\lambda(\epsilon\mathbf{v})\epsilon\mathbf{v},\mathbf{v})/6| & < \quad \delta \end{cases}$$

Therefore, for all $\mathbf{v}\in S$, all $0<\epsilon<\bar{\epsilon}$,

$$\text{sign}\ \big(f(\bar{\mathbf{x}}+\mathbf{dx}) - f(\bar{\mathbf{x}})\big) \quad = \quad \begin{cases} \text{sign}\ \big(\bigtriangledown f(\bar{\mathbf{x}})\mathbf{v}\big) & \text{if } \bigtriangledown f(\bar{\mathbf{x}})\mathbf{v} \leq -\gamma \\ \text{sign}\ \big(\mathbf{v}'\mathrm{Hf}(\bar{\mathbf{x}})\mathbf{v}/2\big) & \text{if } -\gamma < \bigtriangledown f(\bar{\mathbf{x}})\mathbf{v} \leq 0 \end{cases}$$

It follows, therefore, that for all $\mathbf{v}\in S$ all $0<\epsilon<\bar{\epsilon}$, $\bigtriangledown f(\bar{\mathbf{x}})\mathbf{v} \leq 0$ implies $f(\bar{\mathbf{x}}+\mathbf{dx}) - f(\bar{\mathbf{x}}) < 0$. Moreover, since the KKT conditions are satisfied, $\bigtriangledown f(\bar{\mathbf{x}})\mathbf{v} > 0$ implies $\bigtriangledown g(\bar{\mathbf{x}})\mathbf{v} > 0$. Now by assumption, $\bigtriangledown f(\bar{\mathbf{x}}) \neq 0$ so that $g(\mathbf{x}) = b$. Since g is linear, $\bigtriangledown f(\bar{\mathbf{x}})\mathbf{v} > 0$ implies $g(\bar{\mathbf{x}}+\mathbf{dx}) > b$ (i.e., $\bar{\mathbf{x}}+\mathbf{dx}$ is outside of the constraint set so we don't have to worry about it). We have, therefore, shown that f attains a local maximum on the constraint set. Moreover, since the constraint set is convex and f is quasi-concave, a local maximum is a global maximum. It follows that $\bar{\mathbf{x}}$ is a solution to the NPP $\square$.

[1] To see that $\mathrm{Tf}_3(\bar{\mathbf{x}}+\lambda(\mathbf{dx})\mathbf{dx},\mathbf{dx})$ is continuous w.r.t. $\mathbf{dx}$, observe that

$$\mathrm{Tf}_3(\bar{\mathbf{x}}+\lambda(\mathbf{dx})\mathbf{dx},\mathbf{dx}) \quad = \quad f(\bar{\mathbf{x}}+\mathbf{dx}) - \big(f(\bar{\mathbf{x}}) + \bigtriangledown f(\bar{\mathbf{x}})\mathbf{dx} + \mathbf{dx}'\mathrm{Hf}(\bar{\mathbf{x}})\mathbf{dx}/2\big)$$

Since $f(\cdot)$ is continuous and the difference between continuous functions is continuous, the left-hand side is continuous.

[2] The function $h(\mathbf{v}) = \mathbf{v}'\mathrm{Hf}(\bar{\mathbf{x}})\mathbf{v}$ is continuous w.r.t $\mathbf{v}$; since S is compact, it follows from Weierstrass's theorem that $h(\cdot)$ attains a maximum at, say $\underline{\mathbf{v}}$. Moreover, since $h(\cdot)$ is negative on S, $h(\underline{\mathbf{v}}) < -\delta/2 < 0$, for some $\delta > 0$. Now consider the function $\mathrm{Tf}_3 : B(\mathbf{x},1)\times S \to \mathbb{R}$, where $\mathrm{Tf}_3(\mathbf{x}',\mathbf{v})$ is the third order Taylor term centered at $\mathbf{x}'$ in the direction $\mathbf{v}$. Since f is thrice continuously differentiable, $\mathrm{Tf}_3(\cdot,\cdot)$ is a continuous function. Since $B(\mathbf{x},1)\times S$ is compact, $\mathrm{Tf}_3(\cdot,\cdot)$ is bounded. Hence there exists $\omega > 0$ such that for all $\mathbf{v}\in S$ and all $\epsilon\in[0,1]$, $|\mathrm{Tf}_3(\bar{\mathbf{x}}+\lambda(\epsilon\mathbf{v})\epsilon\mathbf{v},\mathbf{v})/6| < \omega$.

6.4. **Second Order conditions Without Quasi-Concavity**

This is a hard section that I'm not going to teach this year. Rather than include it in the notes, I've suppressed the entire section to save paper. If anybody is interested I'll print off the section.