Sufficiency conditions: Key points

- Quasi-concavity and semi-definiteness on a subspace: $\mathbf{dx}' Hf(\mathbf{x})\mathbf{dx} \leq 0$, for all $\mathbf{dx}$ s.t. $\bigtriangledown f(\mathbf{x})\mathbf{dx} = 0$.
- The principal minor representation of strict quasi-concavity: $\forall \mathbf{x}$, and all $k = 1, \ldots, n$, the sign of the k'th leading principal minor of the bordered matrix $\begin{bmatrix} 0 & \bigtriangledown f(\mathbf{x})' \\ \bigtriangledown f(\mathbf{x}) & \mathrm{Hf}(\mathbf{x}) \end{bmatrix}$ must have $\mathrm{sgn}((-1)^k)$, where the k'th leading principal minor of this matrix is the det of the top-left $(k+1) \times (k+1)$ submatrix.
- Sufficiency: suppose definiteness properties hold *and* $\bigtriangledown f(\bar{\mathbf{x}}) \neq 0$. If $\bar{\mathbf{x}}$ satisfies KKT, it solves the NPP
  - the role of second order conditions in the sufficiency argument
  - The $\bigtriangledown f(\bar{\mathbf{x}})$ caveat is highly unsatisfactory
- Our sufficiency conditions—quasi $\begin{array}{l} \text{concavity of } f \\ \text{convexity of } g \end{array}$— are way strong
  - e.g., max f with elliptical lower contour sets on g with circular ones
  - just need f to be “less quasi-convex” than g

Key Points (continued)

- Understanding the problem of the vanishing gradient
- Defn of pseudo-concavity: f is pseudo-concave if
$$\forall \mathbf{x}, \mathbf{x}' \in X, \text{ if } f(\mathbf{x}') > f(\mathbf{x}) \text{ then } \bigtriangledown f(\mathbf{x}) \cdot (\mathbf{x}' - \mathbf{x}) > 0.$$
- Pseudo-concavity and its relationship to quasi-concavity:
  Theorem: Assume f is $\mathbb{C}^2$. Then f is pseudo-concave iff
  1. f is quasi-concave
  2. $\bigtriangledown f(\cdot) = 0$ at $\mathbf{x}$ implies $f(\cdot)$ attains a global max at $\mathbf{x}$.
- NASC for a solution to the NPP (inequality constraints):
  If f is pseudo-concave and the g^j's are quasi-convex, then a necessary and 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 $g^j(\bar{\mathbf{x}}) < b_j$.

How do you know if a Hessian is quasi-con?

Definition: The bordered Hessian:

- top row is $[0, \bigtriangledown f(\mathbf{x})]$;
- left column below top row is $\bigtriangledown f(\mathbf{x})'$
- the rest is the Hessian

Definition: Leading k'th principal minor of a bordered Hessian:

- determinant of the top-left $k+1 \times k+1$ submatrix of the bordered matrix

For f to be strictly quasi-concave

$$\text{sgn}\left(k\text{'th leading principal minor of } \begin{bmatrix} 0 & \bigtriangledown f(\mathbf{x})' \\ \bigtriangledown f(\mathbf{x}) & \text{Hf}(\mathbf{x}) \end{bmatrix} \right) = \text{sgn}((-1)^k).$$

For f to be strictly quasi-convex

$$\text{sgn}\left(k\text{'th leading principal minor of } \begin{bmatrix} 0 & \bigtriangledown f(\mathbf{x})' \\ \bigtriangledown f(\mathbf{x}) & \text{Hf}(\mathbf{x}) \end{bmatrix} \right) = -1.$$

Two examples follow below

Bordered Hessian check for quasi-concavity: example

- f is sqc: $f(\mathbf{x}) = x_1 x_2$; $\quad \triangledown f(\mathbf{x}) = (x_2, x_1)$; $\quad Hf(\mathbf{x}) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

  Evaluate bordered Hessian at $\mathbf{x} = (1,1)$.

$$\mathrm{BHf}(\mathbf{x})|_{(1,1)} = \begin{bmatrix} 0 & \triangledown f(1,1)' \\ \triangledown f(1,1) & \mathrm{Hf}(1,1) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

  - First principal minor is det $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} < 0$;
  - Second principal minor is det $\mathrm{BHf}(\mathbf{x})|_{(1,1)} = 2$.
  - Passes test for quasi-concavity

Bordered Hessian check for quasi-convexity: example

- f is convex: $f(\mathbf{x}) = 0.5(x_1^2 + x_2^2)$; $\quad \bigtriangledown f(\mathbf{x}) = (x_1, x_2)$; $\quad Hf(\mathbf{x}) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

  Evaluate bordered Hessian at $\mathbf{x} = (1, 1)$.

$$\mathrm{BHf}(\mathbf{x})\big|_{(1,1)} = \begin{bmatrix} 0 & \bigtriangledown f(1,1)' \\ \bigtriangledown f(1,1) & \mathrm{Hf}(1,1) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

  - First principal minor is det $\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} < 0$;
  - Second principal minor is det $\mathrm{BHf}(\mathbf{x})\big|_{(1,1)} = -2$.
  - Passes test for quasi-convexity

Definiteness plus $\bigtriangledown f(\bar{\mathbf{x}}) \neq 0$ imply sufficiency

Theorem: Suppose $\forall \mathbf{x}$, $\begin{matrix} Hf(\mathbf{x}) \\ Hg(\mathbf{x}) \end{matrix}$ is $\begin{matrix} \text{neg} \\ \text{pos} \end{matrix}$ definite on subspace orthog to $\bigtriangledown f(\mathbf{x})$.

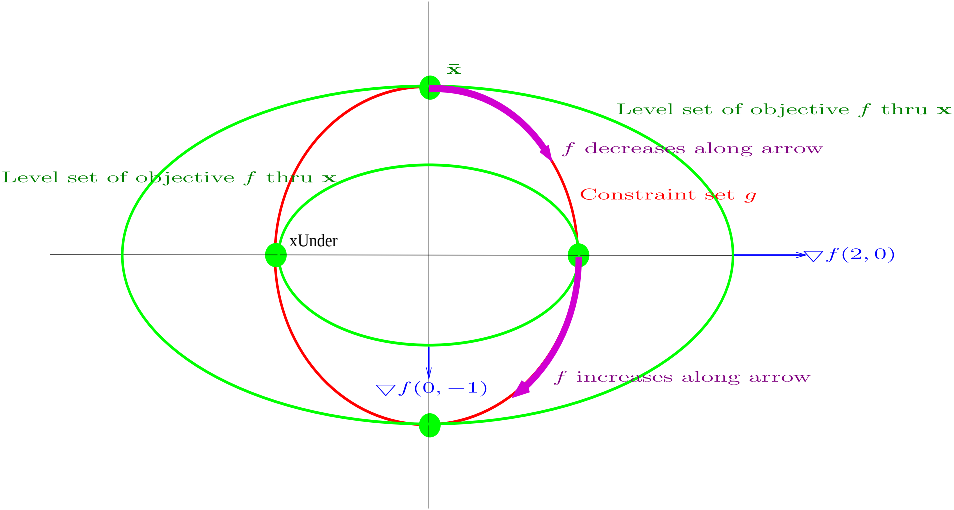
If $\bar{\mathbf{x}}$ satisfies NPP and $\bigtriangledown f(\bar{\mathbf{x}}) \neq 0$, then $\bar{\mathbf{x}}$ solves the NPP

Proof outline

1. Prove that $\bar{\mathbf{x}}$ is a *local* max on constraint set:
   - **Need to show**: $\exists \bar{\varepsilon} > 0$ s.t. $\forall \mathbf{dx}$ with $||\mathbf{dx}|| < \bar{\varepsilon}$, $f(\bar{\mathbf{x}} + \mathbf{dx}) > f(\bar{\mathbf{x}}) \Longrightarrow g^j(\bar{\mathbf{x}} + \mathbf{dx}) > b_j$ for some j
   - See argument below
2. Definiteness conditions imply
   - upper contour sets of f are convex
   - constraint set is convex.
3. Use convexity properties to prove that $\bar{\mathbf{x}}$ is a *global* max on constraint set:
   - suppose for arbitrary $\mathbf{dx}$, $f(\bar{\mathbf{x}} + \mathbf{dx}) > f(\bar{\mathbf{x}})$ & $g^j(\bar{\mathbf{x}} + \mathbf{dx}) \leq b_j$, $\forall j$.
   - lower contour sets convex $\Longrightarrow g^j(\bar{\mathbf{x}} + \delta\mathbf{dx}) \leq b_j$, $\forall j$ & $\forall \delta < 1$.
   - since $\{\mathbf{x} : f(\mathbf{x}) > f(\bar{\mathbf{x}})\}$ is convex, $f(\bar{\mathbf{x}} + \delta\mathbf{dx}) > f(\bar{\mathbf{x}})$, $\forall \delta > 0$.
   - $\exists$ sufficiently small $\delta > 0$ s.t. $||\delta\mathbf{dx}|| < \bar{\varepsilon}$
   - contradiction of point 1

Sufficient conditions without quasi-concavity: example

Maximize a (quasi-convex) ellipse on a circle:



Sufficient conditions without Quasi-concavity: theory

Maximize an quasi-convex function f s.t. $g(\mathbf{x}) = b$

- $f(\cdot)$ is *locally maximized* at $\bar{\mathbf{x}}$ on constraint
  - if level set of $f(\cdot)$ thru $\bar{\mathbf{x}}$ is *less curved* than level set of $g(\cdot)$ thru $\bar{\mathbf{x}}$.
- $f(\cdot)$ is *locally minimized* at $\underline{\mathbf{x}}$ on constraint
  - if level set of $f(\cdot)$ thru $\underline{\mathbf{x}}$ is *more curved* than level set of $g(\cdot)$ thru $\underline{\mathbf{x}}$.

Test: check 2nd-order Taylor expansion of *Lagrangian* for **dx** s.t. $\mathbf{dx} \perp \bigtriangledown f(\mathbf{x})$.

$$\begin{aligned}
L(\lambda, \mathbf{x}) &= f(\mathbf{x}) + \lambda(b - g(\mathbf{x})) \\
L(\lambda, \mathbf{x} + \mathbf{dx}) - L(\lambda, \mathbf{dx}) &\approx \bigtriangledown L(\lambda, \mathbf{x})[0, \mathbf{dx}] \quad + \quad 0.5[0, \mathbf{dx}]' HL(\lambda, \mathbf{dx})[0, \mathbf{dx}] \\
&= 0.5[0, \mathbf{dx}]' HL(\lambda, \mathbf{dx})[0, \mathbf{dx}]
\end{aligned}$$

1. $f(\cdot)$ *locally max-ed* at $\bar{\mathbf{x}}$ if $HL(\bar{\mathbf{x}})$ is *neg definite* for $\mathbf{dx} \perp \bigtriangledown f(\mathbf{x})$.
2. $f(\cdot)$ *locally min-ed* at $\underline{\mathbf{x}}$ if $HL(\underline{\mathbf{x}})$ is *pos definite* for $\mathbf{dx} \perp \bigtriangledown f(\mathbf{x})$.
3. quasi- $\begin{matrix}\text{concavity} \\ \text{convexity}\end{matrix}$ of $\begin{matrix} f \\ g \end{matrix}$ is sufficient for (1) but not necessary

Check definiteness of bordered Hessian from *Lagrangian*

To get bordered Hessian we need, first construct the Hessian of Lagrangian:

$$\begin{aligned} \triangledown L(\lambda, \mathbf{x}) &= \begin{bmatrix} \frac{\partial L}{\partial \lambda} & \frac{\partial L}{\partial x_1} & , \cdots , & \frac{\partial L}{\partial x_n} \end{bmatrix} \\ &= \begin{bmatrix} b - g(\mathbf{x}), & \triangledown f(\mathbf{x}) - \lambda \triangledown g(\mathbf{x}) \end{bmatrix} \end{aligned}$$

when $f : \mathbb{R}^2 \to \mathbb{R}$

$$\mathrm{HL}(\lambda, \bar{\mathbf{x}}) = \begin{bmatrix} 0 & g_1(\bar{\mathbf{x}}) & g_2(\bar{\mathbf{x}}) \\ g_1(\bar{\mathbf{x}}) & f_{11}(\bar{\mathbf{x}}) - \bar{\lambda} g_{11}(\bar{\mathbf{x}}) & f_{12}(\bar{\mathbf{x}}) - \bar{\lambda} g_{12}(\bar{\mathbf{x}}) \\ g_2(\bar{\mathbf{x}}) & f_{21}(\bar{\mathbf{x}}) - \bar{\lambda} g_{21}(\bar{\mathbf{x}}) & f_{22}(\bar{\mathbf{x}}) - \bar{\lambda} g_{22}(\bar{\mathbf{x}}) \end{bmatrix}$$

$HL(\lambda, \mathbf{x})$ is a bordered Hessian of $Hf(\mathbf{x}) - \lambda Hg(\mathbf{x})$. We need to check:

$$\mathbf{dx}' \left(\mathrm{Hf}(\bar{\mathbf{x}}) - \bar{\lambda} \mathrm{Hg}(\bar{\mathbf{x}}) \right) \mathbf{dx} \lesseqgtr 0, \text{ for all } \mathbf{dx} \text{ such that } \triangledown g(\bar{\mathbf{x}}) \mathbf{dx} = 0$$

- compute minors of Hessian $\mathrm{Hf}(\bar{\mathbf{x}}) - \bar{\lambda}\mathrm{Hg}(\bar{\mathbf{x}})$ bordered by $\triangledown g(\mathbf{x})$.
  - neg def iff k'th principal minor of bordered Hessian has sign $(-1)^k$.
  - pos def iff k'th principal minor of bordered Hessian is negative
  - note that *first* principal minor is *necessarily* negative
    - when $f : \mathbb{R}^2 \to \mathbb{R}$, only have to check second minor

Sufficient conditions for a solution: one inequality constraint

Assume conditions of the theorem are satisfied at $\bar{\mathbf{x}}$. In particular assume

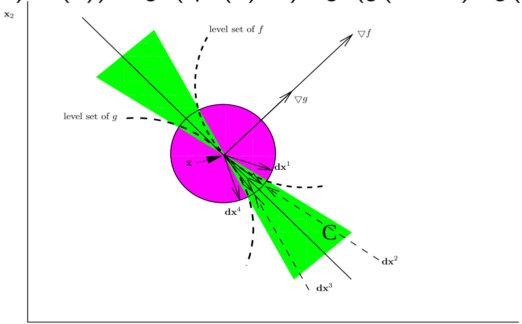
- $\mathbf{dx}' Hf(\mathbf{x})\mathbf{dx} < 0$, for all $\mathbf{dx}$ s.t. $\bigtriangledown f(\mathbf{x})\mathbf{dx} = 0$.
- $\mathbf{dx}' Hg(\mathbf{x})\mathbf{dx} > 0$, for all $\mathbf{dx}$ s.t. $\bigtriangledown f(\mathbf{x})\mathbf{dx} = 0$.

By continuity, $\exists$ an open cone C and $\delta > 0$ s.t. $\forall y \in B(\mathbf{x}, \delta)$

- $\mathbf{dx}' Hf(\mathbf{y})\mathbf{dx} < 0$, for all $\mathbf{dx} \in C$
- $\mathbf{dx}' Hg(\mathbf{y})\mathbf{dx} > 0$, for all $\mathbf{dx} \in C$

For all $\mathbf{dx} \in \mathbb{R} \setminus C$, $\exists \varepsilon(\mathbf{dx}) > 0$ s.t. if $||\mathbf{dx}|| < \varepsilon$, then

$\text{sgn}(f(\mathbf{x}+\mathbf{dx}) - f(\mathbf{x})) = \text{sgn}(\bigtriangledown f(\mathbf{x})\mathbf{dx})$; $\text{sgn}(g(\mathbf{x}-\mathbf{dx}) - g(\mathbf{x})) = \text{sgn}(\bigtriangledown g(\mathbf{x})\mathbf{dx})$

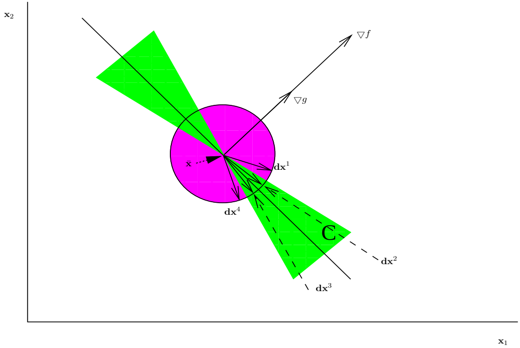


Sufficient conditions for a solution (cont)

Let $\bar{\varepsilon} = \min[\delta, \min\{\varepsilon(\mathbf{dx}) : \mathbf{dx} \in \mathbb{R} \backslash C\}]$ be radius of purple circle

- $\mathrm{sgn}(g(\mathbf{x}+\mathbf{dx}^1) - g(\mathbf{x})) = \mathrm{sgn}(\bigtriangledown g(\mathbf{x})\mathbf{dx}^1) > 0$
- $\mathrm{sgn}(f(\mathbf{x}+\mathbf{dx}^4) - f(\mathbf{x})) = \mathrm{sgn}(\bigtriangledown f(\mathbf{x})\mathbf{dx}^4) < 0$
- $g(\mathbf{x}+\mathbf{dx}^2) - g(\mathbf{x}) = \bigtriangledown g(\mathbf{x})\mathbf{dx}^2 + 0.5\mathbf{dx}^{2'} Hg(\mathbf{y})\mathbf{dx}^2 > 0,$
- $f(\mathbf{x}+\mathbf{dx}^3) - f(\mathbf{x}) = \bigtriangledown f(\mathbf{x})\mathbf{dx}^3 + 0.5\mathbf{dx}^{3'} Hf(\mathbf{y})\mathbf{dx}^3 < 0,$

Conclude: $\forall \mathbf{dx} \in B(\mathbf{x}, \bar{\varepsilon})$, $f(\mathbf{x}+\mathbf{dx}) > f(\mathbf{x}) \Longrightarrow g(\mathbf{x}+\mathbf{dx}) > g(\mathbf{x})$.



The problem of the vanishing gradient

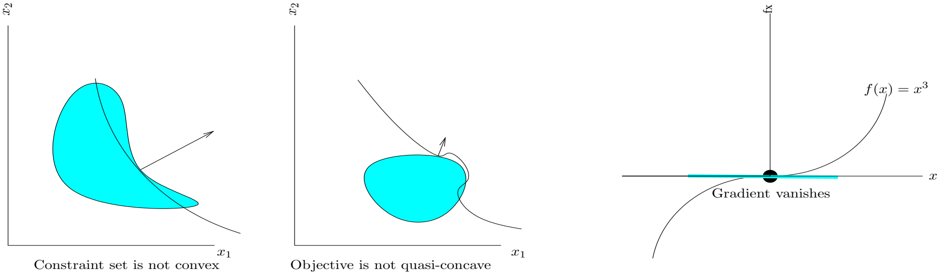
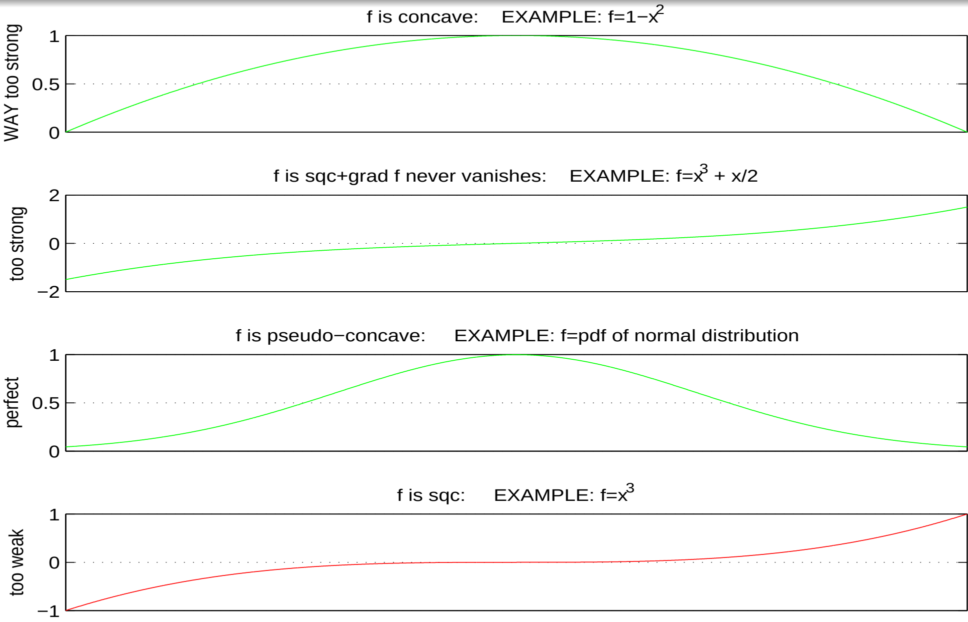


FIGURE 1. Three examples where KKT conditions are not sufficient for a soln

Weakest condition on f s.t. soln to KKT $\Longrightarrow$ soln to NPP



Definition: f is pseudo-concave if

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

Pseudo-concavity defn simply assumes away the vanishing gradient problem.

Note: if f is pseudo-concave, $\bigtriangledown f(\mathbf{x}) = 0 \Longrightarrow \mathbf{x}$ is a global max on X

- suppose $\mathbf{x}$ were not a global max on X, i.e., $\exists \mathbf{x}' \in X,$ s.t. $f(\mathbf{x}') > f(\mathbf{x})$
- since f is pseudo-concave, $\bigtriangledown f(\mathbf{x}) \cdot (\mathbf{x}' - \mathbf{x}) > 0.$
- therefore $\bigtriangledown f(\mathbf{x})$ can't be zero

Pseudo-concavity and Quasi-concavity

Theorem: Assume f is $\mathbb{C}^2$. Then f is pseudo-concave iff

1. f is quasi-concave
2. $\bigtriangledown f(\cdot) = 0$ at $\mathbf{x}$ implies $f(\cdot)$ attains a global max at $\mathbf{x}$.

Proof: $\Longrightarrow$ Assume f is not quasi-concave

- there's a non-convex upper contour set
- $\exists \mathbf{x}', \mathbf{x}'' \in X$ & $\mathbf{x} = \lambda \mathbf{x}' + (1-\lambda)\mathbf{x}''$, s.t. $f(\mathbf{x}'') \geq f(\mathbf{x}') > f(\mathbf{x})$.
- assume w.l.o.g. that $\mathbf{x}$ minimizes $f(\cdot)$ on $[\mathbf{x}', \mathbf{x}'']$.
- by KKT, $\bigtriangledown f(\mathbf{x}) \cdot (\mathbf{x}' - \mathbf{x}) = 0$ but $f(\mathbf{x}') > f(\mathbf{x})$, so f not pseudo-concave

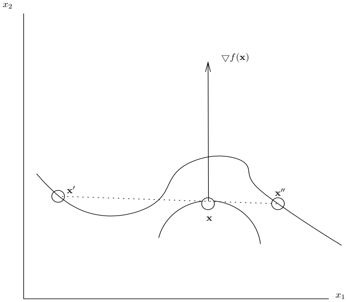


FIGURE 1. ¬-pseudo-concavity implies ¬-quasi-concavity

Pseudo-concavity and Quasi-concavity: necessity

Proof: $\Leftarrow$ Assume f is not pseudo-concave

- $\exists \mathbf{x}, \mathbf{x}' \in X$ s.t. $f(\mathbf{x}') > f(\mathbf{x})$ and $\bigtriangledown f(\mathbf{x}) \cdot (\mathbf{x}' - \mathbf{x}) \leq 0$.
- by continuity, $\exists \mathbf{x}, \mathbf{y} \in X$, s.t. $f(\mathbf{y}) > f(\mathbf{x})$ and $\bigtriangledown f(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}) < 0$.
- let $\mathbf{dx} = \varepsilon(\mathbf{y} - \mathbf{x})$. If ε small enough, $f(\mathbf{x} + \mathbf{dx}) < f(\mathbf{x})$;
- f is not quasi-concave

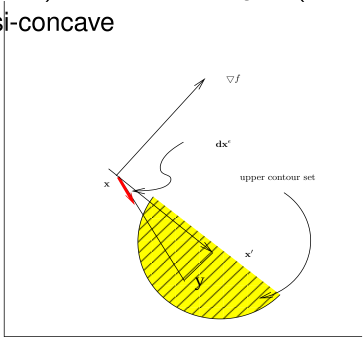


FIGURE 1. $\neg$-quasi-concavity implies $\neg$-pseudo-concavity

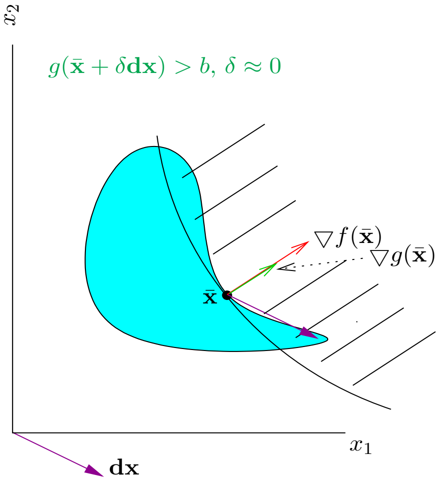
Sufficient conditions for a soln to NPP

Theorem: Assume: f, g^j's are $\mathbb{C}^3$, f is pseudo concave & g^j's are quasi-convex. If the KKT are satisfied at $\bar{\mathbf{x}}$, then $\bar{\mathbf{x}}$ solves the NPP

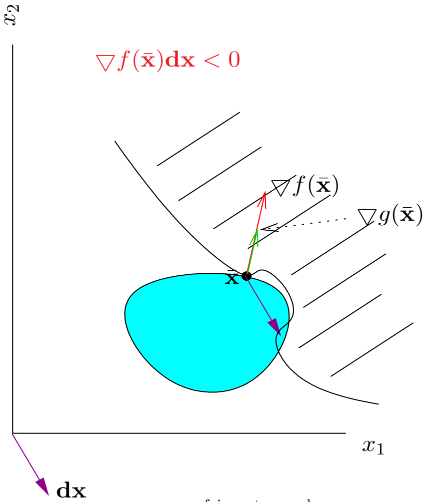
Proof:

1. suppose $\bar{\mathbf{x}}$ does not solve the NPP, i.e., for some $\mathbf{dx}$;
   - $f(\bar{\mathbf{x}}+\mathbf{dx}) > f(\bar{\mathbf{x}})$ & $g^j(\bar{\mathbf{x}}+\mathbf{dx}) \leq b_j, \forall j$.
   - since f is pseudo-concave, $\bigtriangledown f(\bar{\mathbf{x}})\mathbf{dx} > 0$.
   - we'll show $\mathbf{x}$ fails the KKT:
     - i.e., consider *any* $\boldsymbol{\lambda} \in \mathbb{R}^m_+$ s.t. $g^j(\bar{\mathbf{x}}) < b^j \implies \lambda^j = 0$;
     - we'll show below that $\bigtriangledown f(\bar{\mathbf{x}}) \neq \boldsymbol{\lambda}^T Jg(\bar{\mathbf{x}})$
2. lower contour sets convex $\implies g^j(\bar{\mathbf{x}}+\delta\mathbf{dx}) \leq b_j$, $\forall j$ & $\forall \delta < 1$
3. therefore, $\forall j$ s.t. $g^j(\bar{\mathbf{x}}) = b^j$, $\forall \delta < 1$, $g^j(\bar{\mathbf{x}}+\delta\mathbf{dx}) - g^j(\mathbf{dx}) \leq 0$.
   - it then follows from Local Taylor that for each such j, $\bigtriangledown g^j(\bar{\mathbf{x}})\mathbf{dx} \leq 0$,
     - to see this, suppose to the contrary that for one such j, $\bigtriangledown g^j(\bar{\mathbf{x}})\mathbf{dx} > 0$;
       - ◇ then by Local Taylor: for $\delta \approx 0$, $g^j(\bar{\mathbf{x}}+\delta\mathbf{dx}) > g^j(\mathbf{dx}) = b_j$, contradicting (2);
   - note: we *can't* conclude that for j s.t. $g^j(\bar{\mathbf{x}}) < b^j$, $\bigtriangledown g^j(\bar{\mathbf{x}})\mathbf{dx} \leq 0$.
     - but this won't invalidate (4) since by assumption, if $g^j(\bar{\mathbf{x}}) < b^j$ then $\lambda^j = 0$.
4. Part (3) now implies $(\boldsymbol{\lambda}^T Jg(\bar{\mathbf{x}})) \cdot \mathbf{dx} = \sum_{j=1}^m \lambda^j \bigtriangledown g^j(\bar{\mathbf{x}}) \cdot \mathbf{dx} \leq 0$
5. But since $\bigtriangledown f(\bar{\mathbf{x}})\mathbf{dx} > 0$, $\bigtriangledown f(\bar{\mathbf{x}}) \neq \boldsymbol{\lambda}^T Jg(\bar{\mathbf{x}})$, i.e., $\bar{\mathbf{x}}$ fails the KKT.

KKT conditions satisfied but sufficiency conditions aren't



lower contour set is not convex



f is not pseudo-concave