Notes for Econ. 7001

by Gabriel A. Lozada

The equation numbers and page numbers refer to Knut Sydsæter and Peter J. Hammond's textbook *Mathematics for Economic Analysis* (ISBN 0-13-583600-X, 1995).

1. Convexity, Quadratic Forms, and Minors

Let $\mathbf{A}$ denote a matrix. It does not have to be square. A "**minor** of $\mathbf{A}$ of order r" is obtained by deleting all but r rows and r columns of $\mathbf{A}$, then taking the determinant of the resulting $r \times r$ matrix. [p. 466; this page has a good example.] One can show that a matrix of dimension $m \times n$ has $\big(m!/[r!\,(m-r)!]\big) \cdot \big(n!/[r!\,(n-r)!]\big)$ minors of order r for $r = 1, \ldots, n$.

Now let $\mathbf{A}$ denote a square matrix. A "**principal minor** of $\mathbf{A}$ of order r" is obtained by deleting all but r rows and the *corresponding* r columns of $\mathbf{A}$, then taking the determinant of the resulting $r \times r$ matrix. [p. 536] (For example, if you keep the first, third, and fourth rows, then you have to keep the first, third, and fourth columns.) A principal minor of $\mathbf{A}$ of order r is denoted by Δ_r of $\mathbf{A}$. [p. 638] One can show that a square matrix of dimension $n \times n$ has $n!/[r!\,(n-r)!]$ principal minors of order r for $r = 1, \ldots, n$.

Again let $\mathbf{A}$ denote a square matrix. A "**leading principal minor** of $\mathbf{A}$ of order r" is obtained by deleting all but the *first* r rows and the *first* r columns of $\mathbf{A}$, then taking the determinant of the resulting $r \times r$ matrix. [important] [p. 534] A leading principal minor of $\mathbf{A}$ of order r is denoted by D_r of $\mathbf{A}$. A square matrix of dimension $n \times n$ has only 1 leading principal minor of order r for $r = 1, \ldots, n$.

Example. Suppose $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix}$. This matrix is not symmetric. Usually one is interested in the minors only of symmetric matrices, but there is nothing wrong with finding the minors of this non-symmetric matrix.

- The leading principal minor of order 1 of $\mathbf{A}$ is $D_1 = |1|$.

  There are four principal minors of order 1 of $\mathbf{A}$; they are the Δ_1's: $|1| = D_1$, $|6|$, $|11|$, and $|16|$.

  There are sixteen minors of $\mathbf{A}$ of order 1.

- The leading principal minor of order 2 of $\mathbf{A}$ is $D_2 = \begin{vmatrix} 1 & 2 \\ 5 & 6 \end{vmatrix}$.

  There are six principal minors of order 2 of $\mathbf{A}$; they are the Δ_2's: $\begin{vmatrix} 1 & 2 \\ 5 & 6 \end{vmatrix} = D_2$ (from rows and columns 1 and 2), $\begin{vmatrix} 1 & 3 \\ 9 & 11 \end{vmatrix}$ (from rows and columns 1 and 3), $\begin{vmatrix} 1 & 4 \\ 13 & 16 \end{vmatrix}$ (from rows and columns 1 and 4), $\begin{vmatrix} 6 & 7 \\ 10 & 11 \end{vmatrix}$ (from rows and columns 2 and 3), $\begin{vmatrix} 6 & 8 \\ 14 & 16 \end{vmatrix}$ (from rows and columns 2 and 4), and $\begin{vmatrix} 11 & 12 \\ 15 & 16 \end{vmatrix}$ (from rows and columns 3 and 4).

  There are thirty-six minors of $\mathbf{A}$ of order 2.

- The leading principal minor of order 3 of $\mathbf{A}$ is $D_3 = \begin{vmatrix} 1 & 2 & 3 \\ 5 & 6 & 7 \\ 9 & 10 & 11 \end{vmatrix}$.

  There are four principal minors of order 3 of $\mathbf{A}$; they are the Δ_3's: $\begin{vmatrix} 1 & 2 & 3 \\ 5 & 6 & 7 \\ 9 & 10 & 11 \end{vmatrix} = D_3$ (from rows and columns 1, 2, and 3), $\begin{vmatrix} 1 & 2 & 4 \\ 5 & 6 & 8 \\ 13 & 14 & 16 \end{vmatrix}$ (from rows and columns 1, 2 and 4), $\begin{vmatrix} 1 & 3 & 4 \\ 9 & 11 & 12 \\ 13 & 15 & 16 \end{vmatrix}$ (from rows and columns 1, 3 and 4), and $\begin{vmatrix} 6 & 7 & 8 \\ 10 & 11 & 12 \\ 14 & 15 & 16 \end{vmatrix}$ (from rows and columns 2, 3 and 4).

  There are sixteen minors of $\mathbf{A}$ of order 3.

- The leading principal minor of order 4 of $\mathbf{A}$ is $D_4 = |\mathbf{A}|$.

  There is only one principal minor of order 4 of $\mathbf{A}$; it is Δ_4 and it is equal to $|\mathbf{A}|$.

  There is only one minor of order 4 of $\mathbf{A}$; it is $|\mathbf{A}|$.

[End of Example]

Let f be a C^2 function mapping R^n into R^1. Denote the Hessian matrix of $f(\mathbf{x})$ by $\nabla^2 f(\mathbf{x})$; this matrix has dimension $n \times n$. Let "D_r of $\nabla^2 f(\mathbf{x})$" denote the rth-order leading principal minor of the Hessian of f. Let "Δ_r of $\nabla^2 f(\mathbf{x})$" denote all the rth-order principal minors of the Hessian of f.

Proposition 1. *One has*

$$D_r \text{ of } \nabla^2 f(\mathbf{x}) > 0 \text{ for } r = 1, \ldots, n \text{ and for all } \mathbf{x} \in S \tag{1}$$

$$\iff \nabla^2 f(\mathbf{x}) \text{ is positive definite for all } \mathbf{x} \in S \tag{2}$$

$$\implies f(\mathbf{x}) \text{ is strictly convex on } S\,. \tag{3}$$

Also,

$$\begin{aligned} &\text{\textit{All the} } \Delta_r \text{ \textit{of} } \nabla^2 f(\mathbf{x}) \geq 0 \text{ \textit{for} } r = 1, \ldots, n \text{ \textit{and for all} } \mathbf{x} \in S && (4)\\ &\iff \nabla^2 f(\mathbf{x}) \text{ \textit{is positive semidefinite for all} } \mathbf{x} \in S && (5)\\ &\iff f(\mathbf{x}) \text{ \textit{is convex on} } S\,. && (6) \end{aligned}$$

If $\nabla^2 f(\mathbf{x})$ is replaced by an arbitrary symmetric matrix, it is still true that (1) $\iff$ (2) and (4) $\iff$ (5).

Proof. (1) $\iff$ (2): Th. 15.3 (a), p. 535. (2) $\implies$ (3): 17.24, p. 637 or Th. 17.14 (b), p. 641.

(4) $\iff$ (5): p. 536. (5) $\iff$ (6): 17.26, p. 638, or Th. 17.13 (b), p. 640. ∎

The typical procedure is to check (1) first. If (1) doesn't apply because one of the signs was strictly negative, then the contrapositive of "(4) iff (6)" tells you that the function is not convex. (This is because each $D_i \in \Delta_i$.)

[Proposition $1'$: Similarly,

$$\begin{aligned} &D_r \text{ of } \nabla^2 f(\mathbf{x}) \text{ alternate in sign beginning with } < 0 \text{ for } r = 1, \ldots, n\\ &\qquad \text{and } \forall\, \mathbf{x} \in S && (1')\\ &\iff \nabla^2 f(\mathbf{x}) \text{ is negative definite for all } \mathbf{x} \in S && (2')\\ &\implies f(\mathbf{x}) \text{ is strictly concave on } S\,. && (3') \end{aligned}$$

Also,

$$\begin{aligned} &\text{All the } \Delta_r \text{ of } \nabla^2 f(\mathbf{x}) \text{ alternate in sign beginning with } \leq 0 \text{ for } r = 1, \ldots, n\\ &\qquad \text{and } \forall\, \mathbf{x} \in S && (4')\\ &\iff \nabla^2 f(\mathbf{x}) \text{ is negative semidefinite for all } \mathbf{x} \in S && (5')\\ &\iff f(\mathbf{x}) \text{ is concave on } S\,. && (6') \end{aligned}$$

]

Section 17.10 of your textbook has an excellent treatment of quasiconvexity and quasiconcavity. You should learn it. However, you should replace Theorem 17.17 on p. 647 with the remarks which come between here and the end of this section.

Simon and Blume (p. 526) give the following definition (they use different notation and a different but equivalent statement): Let S be an open convex subset of R^n. A C^1 function $f : S \to R$ is **quasiconvex** at $\mathbf{x}^* \in S$ if and only if

$$\nabla f(\mathbf{x}^*) \cdot (\mathbf{z} - \mathbf{x}^*) > 0 \quad \text{implies} \quad f(\mathbf{z}) > f(\mathbf{x}^*) \quad \text{for all } \mathbf{z} \in S. \tag{7}$$

One says that f is "quasiconvex on S" if (8) holds for all $\mathbf{x}^* \in S$.

[Aside: A quasiconcave function is defined by reversing both inequalities in (7).]

Simon and Blume (p. 527) give the following definition: Let S be an open convex subset of R^n. A C^1 function $f : S \to R$ is **pseudoconvex** at $\mathbf{x}^* \in S$ if

$$\nabla f(\mathbf{x}^*) \cdot (\mathbf{z} - \mathbf{x}^*) \geq 0 \quad \text{implies} \quad f(\mathbf{z}) \geq f(\mathbf{x}^*) \quad \text{for all } \mathbf{z} \in S. \tag{8}$$

One says that f is "pseudoconvex on S" if (8) holds for all $\mathbf{x}^* \in S$.

[Aside: A pseudoconcave function is defined by reversing both inequalities in (8).][1]

[1]Simon and Blume actually give $\leq$ instead of $\geq$ as the second relation in (8); however, their definition of pseudoconcavity completely agrees with mine, and I believe their definition of pseudoconvexity is wrong. To see why, combine Simon and Blume's Theorems 21.2 and 21.3:

$$f \text{ concave} \iff f(\mathbf{z}) - f(\mathbf{x}^*) \leq \nabla f(\mathbf{x}^*) \cdot (\mathbf{z} - \mathbf{x}^*)\,.$$

So as they note in their Corollary 21.4:

$$f \text{ concave} \iff f(\mathbf{z}) - f(\mathbf{x}^*) \leq \nabla f(\mathbf{x}^*) \cdot (\mathbf{z} - \mathbf{x}^*) \leq 0 \Rightarrow f(\mathbf{z}) \leq f(\mathbf{x}^*)\,.$$

Then Simon and Blume define a function f to be pseudoconcave if, as I wrote above:

$$\nabla f(\mathbf{x}^*) \cdot (\mathbf{z} - \mathbf{x}^*) \leq 0 \Rightarrow f(\mathbf{z}) \leq f(\mathbf{x}^*)\,.$$

By analogy and by Simon and Blume's Theorem 21.2,

$$f \text{ convex} \iff f(\mathbf{z}) - f(\mathbf{x}^*) \geq \nabla f(\mathbf{x}^*) \cdot (\mathbf{z} - \mathbf{x}^*)\,;$$

then in the spirit of their Corollary 21.4:

$$f \text{ convex} \iff f(\mathbf{z}) - f(\mathbf{x}^*) \geq \nabla f(\mathbf{x}^*) \cdot (\mathbf{z} - \mathbf{x}^*) \geq 0 \Rightarrow f(\mathbf{z}) \geq f(\mathbf{x}^*)\,.$$

The analogous definition of f being pseudoconvex is therefore

$$\nabla f(\mathbf{x}^*) \cdot (\mathbf{z} - \mathbf{x}^*) \geq 0 \Rightarrow f(\mathbf{z}) \geq f(\mathbf{x}^*)$$

which is (8) above. My definition is also the same as the one in S. Schaible, "Second-order Characterizations of Pseudoconvex Quadratic Functions," Journal of Optimization Theory and Applications, vol. 21 no. 1, Jan. 1977, pp. 15–26. Also, note that since everyone defines pseudoconcavity by $\nabla f(\mathbf{x}^*) \cdot (\mathbf{z} - \mathbf{x}^*) \leq 0 \Rightarrow f(\mathbf{z}) \leq f(\mathbf{x}^*)$, and since that implies $\nabla(-f(\mathbf{x}^*)) \cdot (\mathbf{z} - \mathbf{x}^*) \geq 0 \Rightarrow -f(\mathbf{z}) \geq -f(\mathbf{x}^*)$, my definition ensures that if f is pseudoconcave then $-f$ is pseudoconvex, as is appropriate.

Pseudoconvexity is not very interesting in and of itself (though see point 4 of Section 4 below); its main use lies in its similarity with quasiconvex functions. By Theorem 21.17 of Simon and Blume (p. 528), one has:

Proposition 2. [Pseudoconvexity and Quasiconvexity.] *Let S be a convex subset of R^n and let $f : S \to R$ be a C^1 function. Then:*

1. *f is pseudoconvex on $S \implies f$ is quasiconvex on S.*
2. *if S is open and if $\nabla f(\mathbf{x}) \neq \mathbf{0}$ for all $\mathbf{x} \in S$, then:*
   f is pseudoconvex on $S \iff f$ is quasiconvex on S.

[Proposition 2′: Similarly,

1. f is pseudoconcave on $S \implies f$ is quasiconcave on S.
2. if S is open and if $\nabla f(\mathbf{x}) \neq \mathbf{0}$ for all $\mathbf{x} \in S$, then:
   f is pseudoconcave on $S \iff f$ is quasiconcave on S.

]

The neat thing about pseudoconvex functions is that there is an easy second derivative test for such functions, a test that is also usually the easiest check that a given function is *quasiconvex* [I am partially quoting from Simon and Blume p. 528]. Here is that test [Simon and Blume Th. 21.19 p. 530]:

Proposition 3. [Test of Pseudoconvexity.] *Let f be a C^2 function defined in an open, convex set S in R^n. Define the "bordered Hessian" determinants $\delta_r(\mathbf{x})$, $r = 1, \ldots, n$ by*

$$\delta_r(\mathbf{x}) = \begin{vmatrix} 0 & f'_1(\mathbf{x}) & \cdots & f'_r(\mathbf{x}) \\ f'_1(\mathbf{x}) & f''_{11}(\mathbf{x}) & \cdots & f''_{1r}(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ f'_r(\mathbf{x}) & f''_{r1}(\mathbf{x}) & \cdots & f''_{rr}(\mathbf{x}) \end{vmatrix}.$$

A sufficient condition for f to be pseudoconvex is that $\delta_r(\mathbf{x}) < 0$ for $r = 2, \ldots, n$, and all $\mathbf{x} \in S$.

[Proposition 3′: Similarly, a sufficient condition for f to be pseudoconcave is that $\delta_r(\mathbf{x})$ alternate in sign beginning with > 0 for $r = 2, \ldots, n$, and all $\mathbf{x} \in S$.]

In other words, if you want to check quasiconvexity, you usually check pseudoconvexity, using Proposition 3, then appeal to Proposition 2.

2. First-Order Conditions

This is Theorem 18.6 from p. 437 of Simon and Blume:

Proposition 4. *Suppose that f, $h_1, \ldots, h_j$, and $g_1, \ldots, g_k$ are C^1 functions of n variables. Suppose that $\mathbf{x}^* \in R^n$ is a local minimum of $f(\mathbf{x})$ on the constraint set defined by the j equalities and k inequalities*

$$h_1(\mathbf{x}) = c_1, \quad \ldots, \quad h_j(\mathbf{x}) = c_j\,, \tag{9}$$

$$g_1(\mathbf{x}) \geq b_1, \quad \ldots, \quad g_k(\mathbf{x}) \geq b_k\,. \tag{10}$$

Without loss of generality, we can assume that the first k_0 inequality constraints are binding ("active") at $\mathbf{x}^*$ and that the other $k - k_0$ inequality constraints are not binding.

Suppose that the following "nondegenerate constraint qualification" is satisfied: the rank at $\mathbf{x}^*$ of the Jacobian matrix of equality constraints and the binding *inequality constraints*

$$\begin{pmatrix} \dfrac{\partial h_1(\mathbf{x}^*)}{\partial x_1} & \cdots & \dfrac{\partial h_1(\mathbf{x}^*)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial h_j(\mathbf{x}^*)}{\partial x_1} & \cdots & \dfrac{\partial h_j(\mathbf{x}^*)}{\partial x_n} \\ \dfrac{\partial g_1(\mathbf{x}^*)}{\partial x_1} & \cdots & \dfrac{\partial g_1(\mathbf{x}^*)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial g_{k_0}(\mathbf{x}^*)}{\partial x_1} & \cdots & \dfrac{\partial g_{k_0}(\mathbf{x}^*)}{\partial x_n} \end{pmatrix}$$

is $k_0 + j$ —as large as it can be.

Form the Lagrangian

$$\mathscr{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum_{i=1}^{j} \lambda_i \left[c_i - h_i(\mathbf{x})\right] + \sum_{i=1}^{k} \mu_i \left[b_i - g_i(\mathbf{x})\right].$$

Then there exist multipliers $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$ such that: [important: $\mathscr{L}$ & conditions 1–3 below]

1. *$\partial \mathscr{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)/\partial \lambda_i = 0$ for all $i = 1, \ldots, j$. This is equivalent to: $h_i(\mathbf{x}^*) = c_i$ for all $i = 1, \ldots, j$.*

2. $\partial\mathscr{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)/\partial x_i = 0$ *for all* $i = 1, \ldots, n$.

   (Or, if in addition one requires that $x_i^* \geq 0$*, this condition becomes:* $x_i^* \geq 0$*,* $\partial\mathscr{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)/\partial x_i \geq 0$*, and* $x_i^*\big[\partial\mathscr{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)/\partial x_i\big] = 0$*.) [p. 691–3]*

3. $\mu_i^* \geq 0$, $g_i(\mathbf{x}^*) - b_i \geq 0$, *and* $\mu_i^*\,[g_i(\mathbf{x}^*) - b_i] = 0$ *for all* $i = 1, \ldots, k$.

These three conditions are often called the Kuhn-Tucker conditions. The last condition is sometimes called the "complementary slackness condition." Be sure to read the "Note" on p. 702. (Also, there are other constraint qualifications you can use instead of the "nondegenerate constraint qualification" used above. The other constraint qualifications are given in Th. 19.12, p. 476 of Simon and Blume.)

[Proposition 4′: For a maximum, change (10) to:

$$g_1(\mathbf{x}) \leq b_1, \quad \ldots, \quad g_k(\mathbf{x}) \leq b_k. \tag{10′}$$

Then the Lagrangian is formed in the same way. Condition 1 is unchanged. The first sentence of Condition 2 is unchanged. In the second, parenthesized sentence of Condition 2, the sign of $\partial\mathscr{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)/\partial x_i$ should be changed from ≥ 0 to ≤ 0. Condition 3 becomes: $\mu_i^* \geq 0$, $g_i(\mathbf{x}^*) - b_i \leq 0$, and $\mu_i^*\,[g_i(\mathbf{x}^*) - b_i] = 0$ for all $i = 1, \ldots, k$.]

By the way, it is easy to prove that if $g_1, \ldots, g_k$ are all quasiconcave, then the set of $\mathbf{x}$'s which satisfy (10) is a convex set. Here is the proof. For $i = 1, 2, \ldots, k$, define $S_i = \{\mathbf{x}\colon g_i(\mathbf{x}) \geq b_i\}$. This is an upper level set of g_i. It is convex if g_i is a quasiconcave function. If $g_1, \ldots, g_k$ are all quasiconcave, then $S_1, \ldots, S_k$ are all convex, and their intersection $\bigcap_{i=1}^k S_i$ is also convex [[17.13] p. 620]. This intersection is the set of all $\mathbf{x}$ which satisfy the constraints (10). This completes the proof. (See the conditions on $g_1, \ldots, g_k$ in Section 4 points 3, 4, and 5.)

[Aside: Similarly, if $g_1, \ldots, g_k$ are all quasiconvex, then $\{\mathbf{x}\colon (10') \text{ is satisfied}\}$ is a convex set. See Section 4 points 3′, 4′, and 5′.]

3. Second-Order Conditions: Local

Let $\mathscr{L}$ be the Lagrangian of the optimization problem. In Section 2, I named the Lagrange multipliers "λ" if they were associated with one of the j equality constraints and "μ" if they were associated with one of the k

inequality constraints. In this section: (a) ignore all the nonbinding inequality constraints at $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$; and (b) rename the Lagrange multipliers of the binding inequality constraints $\lambda_{j+1}, \lambda_{j+2}, \ldots, \lambda_m$, where:

> m is the number of equality constraints plus the number of binding inequality constraints.

It is allowed to have $m = 0$; if $m = 0$ then there are no Lagrange multipliers. Denote the m binding Lagrange multipliers collectively by $\boldsymbol{\lambda}$. Let there be n variables with respect to which the optimization is occurring; denote these variables collectively by $\mathbf{x}$.

First, a heuristic explanation from Simon and Blume (p. 399). Using a multidimensional version of Taylor's Theorem, one can show that

$$f(\mathbf{x}^* + \mathbf{z}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*) \cdot \mathbf{z} + \tfrac{1}{2}\mathbf{z}^T \left[\nabla^2 f(\mathbf{x}^*)\right]\mathbf{z} + R(\mathbf{z}) \tag{11}$$

where a 'T' superscript denotes the transpose and where $R(\mathbf{z})$ is a remainder term. I will write

$$f(\mathbf{x}^* + \mathbf{z}) \approx f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*) \cdot \mathbf{z} + \tfrac{1}{2}\mathbf{z}^T \left[\nabla^2 f(\mathbf{x}^*)\right]\mathbf{z}\,. \tag{12}$$

The first order condition gives $\nabla f(\mathbf{x}) = 0$ at $\mathbf{x}^*$. Therefore

$$f(\mathbf{x}^* + \mathbf{z}) - f(\mathbf{x}^*) \approx \tfrac{1}{2}\mathbf{z}^T \left[\nabla^2 f(\mathbf{x}^*)\right]\mathbf{z}\,. \tag{13}$$

If $f(\mathbf{x}^*)$ is to really be a local minimum, then the left-hand side of (13) should be greater than or equal to zero, which is equivalent to $\nabla^2 f(\mathbf{x}^*)$ being positive semidefinite. Furthermore, if $f(\mathbf{x}^*)$ is to really be a strict local minimum, then the left-hand side of (13) should be greater than zero, which is equivalent to $\nabla^2 f(\mathbf{x}^*)$ being positive definite.

[Aside: Similarly, if $f(\mathbf{x}^*)$ is to really be a local maximum, then the left-hand side of (13) should be less than or equal to zero, which is equivalent to $\nabla^2 f(\mathbf{x}^*)$ being negative semidefinite. If $f(\mathbf{x}^*)$ is to really be a strict local maximum, then the left-hand side of (13) should be less than zero, which is equivalent to $\nabla^2 f(\mathbf{x}^*)$ being negative definite.]

Let $\nabla^2 \mathscr{L}$ be the following particular Hessian of the Lagrangian: first differentiate $\mathscr{L}$ with respect to all the Lagrange multipliers, then differentiate it with respect to the original variables $\mathbf{x}$.

$$\nabla^2\mathscr{L} = \begin{pmatrix} \mathscr{L}''_{\lambda_1\lambda_1} & \mathscr{L}''_{\lambda_1\lambda_2} & \cdots & \mathscr{L}''_{\lambda_1\lambda_m} & | & \mathscr{L}''_{\lambda_1x_1} & \mathscr{L}''_{\lambda_1x_2} & \cdots & \mathscr{L}''_{\lambda_1x_n} \\ \mathscr{L}''_{\lambda_2\lambda_1} & \mathscr{L}''_{\lambda_2\lambda_2} & \cdots & \mathscr{L}''_{\lambda_2\lambda_m} & | & \mathscr{L}''_{\lambda_2x_1} & \mathscr{L}''_{\lambda_2x_2} & \cdots & \mathscr{L}''_{\lambda_2x_n} \\ \vdots & \vdots & \ddots & \vdots & | & \vdots & \vdots & \ddots & \vdots \\ \mathscr{L}''_{\lambda_m\lambda_1} & \mathscr{L}''_{\lambda_m\lambda_2} & \cdots & \mathscr{L}''_{\lambda_m\lambda_m} & | & \mathscr{L}''_{\lambda_mx_1} & \mathscr{L}''_{\lambda_mx_2} & \cdots & \mathscr{L}''_{\lambda_mx_n} \\ \text{—} & \text{—} & \text{—} & \text{—} & \text{—} & \text{—} & \text{—} & \text{—} & \text{—} \\ \mathscr{L}''_{x_1\lambda_1} & \mathscr{L}''_{x_1\lambda_2} & \cdots & \mathscr{L}''_{x_1\lambda_m} & | & \mathscr{L}''_{x_1x_1} & \mathscr{L}''_{x_1x_2} & \cdots & \mathscr{L}''_{x_1x_n} \\ \mathscr{L}''_{x_2\lambda_1} & \mathscr{L}''_{x_2\lambda_2} & \cdots & \mathscr{L}''_{x_2\lambda_m} & | & \mathscr{L}''_{x_2x_1} & \mathscr{L}''_{x_2x_2} & \cdots & \mathscr{L}''_{x_2x_n} \\ \vdots & \vdots & \ddots & \vdots & | & \vdots & \vdots & \ddots & \vdots \\ \mathscr{L}''_{x_n\lambda_1} & \mathscr{L}''_{x_n\lambda_2} & \cdots & \mathscr{L}''_{x_n\lambda_m} & | & \mathscr{L}''_{x_nx_1} & \mathscr{L}''_{x_nx_2} & \cdots & \mathscr{L}''_{x_nx_n} \end{pmatrix}$$

If you do this right, $\nabla^2\mathscr{L}$ should have an $m \times m$ zero matrix in its upper left-hand corner:

$$\nabla^2\mathscr{L} = \begin{pmatrix} \mathscr{L}''_{\boldsymbol{\lambda\lambda}} & \mathscr{L}''_{\boldsymbol{\lambda x}} \\ \mathscr{L}''_{\boldsymbol{x\lambda}} & \mathscr{L}''_{\boldsymbol{xx}} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathscr{L}''_{\boldsymbol{\lambda x}} \\ \mathscr{L}''^{T}_{\boldsymbol{\lambda x}} & \mathscr{L}''_{\boldsymbol{xx}} \end{pmatrix}$$

where $\mathscr{L}''_{\boldsymbol{\lambda x}}$ is an $m \times n$ matrix and where a 'T' superscript denotes the transpose.

One has the following result:

Proposition 5a. *A sufficient condition for the point* $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ *identified in Proposition 4 to be a strict local minimum is that* $(-1)^m$ *has the same sign as all of the following when they are evaluated at* $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$*:* D_{2m+1} *of* $\nabla^2\mathscr{L}$, D_{2m+2} *of* $\nabla^2\mathscr{L}$, $\ldots$, D_{m+n} *of* $\nabla^2\mathscr{L}$. [important]

[Proposition 5a′: Similarly, one will have a strict local maximum if, when they are evaluated at $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$, the following alternate in sign beginning with the sign of $(-1)^{m+1}$: D_{2m+1} of $\nabla^2\mathscr{L}$, D_{2m+2} of $\nabla^2\mathscr{L}$, $\ldots$, D_{m+n} of $\nabla^2\mathscr{L}$.]

If $m = 0$, this is equivalent to the condition that $\nabla^2\mathscr{L}$ (which in such a case equals $\nabla^2 f(\mathbf{x})$) be *positive definite*, which occurs iff $f(\mathbf{x})$ is strictly convex. (See the discussion concerning (13).)

If $m > 0$, suppose as above that the binding constraints are written $h_1(\mathbf{x}) = c_1$, $h_2(\mathbf{x}) = c_2$, $\ldots$, $h_m(\mathbf{x}) = c_m$. Abbreviate this by $\mathbf{h}(\mathbf{x}) = \mathbf{c}$. As Simon and Blume write (p. 459), "the natural linear constraint set for this problem is the hyperplane which is tangent to the constraint set $\{\mathbf{x} \in R^n \colon \mathbf{h}(\mathbf{x}) = \mathbf{c}\}$ at the point $\mathbf{x}^*$. . . . the tangent space to $\{\mathbf{h}(\mathbf{x}) = \mathbf{c}\}$ at the point $\mathbf{x}^*$ is the set of vectors $\mathbf{z}$ such that $\nabla \mathbf{h}(\mathbf{x}^*)\, \mathbf{z} = \mathbf{0}$ [i.e., $\nabla h_i(\mathbf{x}^*) \cdot \mathbf{z} = 0$ for all $i = 1, 2, \ldots, n$] (considered as vectors with their tails at $\mathbf{x}^*$)." It can be shown that the above second-order sufficient condition for a minimum (Proposition 5a) is equivalent to saying that, for any non-zero vector $\mathbf{z}$ which is tangent to the constraint set—that is, for any non-zero vector $\mathbf{z}$ which obeys $\nabla \mathbf{h}(\mathbf{x}^*)\, \mathbf{z} = \mathbf{0}$ — (14) holds. Formally: [not very important]

Proposition 5b. *A sufficient condition for the point* $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ *identified in Proposition 4 to be a strict local minimum is that for any non-zero vector* $\mathbf{z}$ *which obeys* $\nabla \mathbf{h}(\mathbf{x}^*)\, \mathbf{z} = \mathbf{0}$, *one has*

$$\mathbf{z}^T \left[\nabla^2_{\boldsymbol{x}} \mathscr{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*)\right] \mathbf{z} > 0\,. \tag{14}$$

In a sense, then, $\nabla^2 \mathscr{L}$ is *positive definite* with respect to vectors which satisfy a linearized version of the constraints (or which "satisfy the constraints to first order"). [See also p. 530.]

[Proposition 5b′: A sufficient condition for a strict local maximum is that the sign in (14) be reversed. When this holds, there is a sense in which $\nabla^2 \mathscr{L}$ is negative definite with respect to vectors which satisfy a linearized version of the constraints.]

There is a second-order necessary condition for a minimum, also. This condition is that the strict inequality $>$ in (14) should be replaced by a weak inequality $\geq$. Formally:

Proposition 6b. *A necessary condition for the point* $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ *identified in Proposition 4 to be a local minimum is that for any non-zero vector* $\mathbf{z}$ *which obeys* $\nabla \mathbf{h}(\mathbf{x}^*)\, \mathbf{z} = \mathbf{0}$, *one has*

$$\mathbf{z}^T \left[\nabla^2_{\boldsymbol{x}} \mathscr{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*)\right] \mathbf{z} \geq 0\,. \tag{15}$$

In other words, $\nabla^2 \mathscr{L}$ should be *positive semidefinite* with respect to vectors which satisfy a linearized version of the constraints (or which "satisfy the constraints to first order").

[Proposition 6b′. A necessary condition for a local maximum is that the sign in (15) be reversed. When this holds, there is a sense in which $\nabla^2 \mathscr{L}$ is negative semidefinite with respect to vectors which satisfy a linearized version of the constraints.]

To check the condition given in Proposition 6b, using (4) and (5) for inspiration, the most obvious analog of Proposition 5a would be:

> Incorrect Guess: A necessary condition for the point $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ identified in Proposition 4 to be a local minimum is that "$(-1)^m$ or zero" have the same sign as all of the following when they are evaluated at $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$: Δ_{2m+1} of $\nabla^2\mathscr{L}$, Δ_{2m+2} of $\nabla^2\mathscr{L}$, $\ldots$, Δ_{m+n} of $\nabla^2\mathscr{L}$.

This guess is incorrect because it is needlessly strict. It is actually not necessary to check every one of Δ_{2m+1} of $\nabla^2\mathscr{L}$, Δ_{2m+2} of $\nabla^2\mathscr{L}$, $\ldots$, Δ_{m+n} of $\nabla^2\mathscr{L}$; you only have to check some of them. The following proposition describes which of the Δ's have to be checked.

Proposition 6a. *Define "$\widehat{\Delta}_i$ of $\nabla^2\mathscr{L}$" to be the subset of "Δ_i of $\nabla^2\mathscr{L}$" formed by only considering those "Δ_i of $\nabla^2\mathscr{L}$" which retain (parts of) the first m rows and first m columns of $\nabla^2\mathscr{L}$. (If $m = 0$, there is no difference between the Δ's and the $\widehat{\Delta}$'s.)*

Then a necessary condition for the point $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ identified in Proposition 4 to be a local minimum is that "$(-1)^m$ or zero" have the same sign as all of the following when they are evaluated at $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$: $\widehat{\Delta}_{2m+1}$ of $\nabla^2\mathscr{L}$, $\widehat{\Delta}_{2m+2}$ of $\nabla^2\mathscr{L}$, $\ldots$, $\widehat{\Delta}_{m+n}$ of $\nabla^2\mathscr{L}$.

[See Theorem M.D.3 (ii) on p. 938 of Mas-Colell, Whinston, and Green, and see p. 468 of Simon and Blume. For the case of $m = 0$, Proposition 6a is the same result as given in Th. 17.12d p. 639 of your textbook.]

The typical procedure is to check Proposition 5a first. If Proposition 5a doesn't apply because one of the signs was strictly the same as $(-1)^{m+1}$, then Proposition 6a tells you that $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ is not a local minimum point. (This is because each $D_i \in \widehat{\Delta}_i$.)

[Proposition 6a$'$: The version of Proposition 6a for a local maximum requires that the following, if they are evaluated at $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$, alternate in sign beginning with the sign of "$(-1)^{m+1}$ or zero" (then having the sign of "$(-1)^{m+2}$ or zero" and so forth): $\widehat{\Delta}_{2m+1}$ of $\nabla^2\mathscr{L}$, $\widehat{\Delta}_{2m+2}$ of $\nabla^2\mathscr{L}$, $\ldots$, $\widehat{\Delta}_{m+n}$ of $\nabla^2\mathscr{L}$.]

While it is nice that Proposition 6a tells you to check only the $\widehat{\Delta}_i$'s instead of all the Δ_i's, there still are very many $\widehat{\Delta}_i$'s in a typical problem. Here is an example.

Example. Suppose $n = 6$ and $m = 2$. Then $\nabla^2\mathscr{L}$ is 8×8. Proposition 5a requires checking D_5, D_6, D_7, and D_8. Proposition 6a requires checking

$\widehat{\Delta}_5$, $\widehat{\Delta}_6$, $\widehat{\Delta}_7$, and $\widehat{\Delta}_8$. For all these $\widehat{\Delta}$'s, you keep rows and columns 1 and 2; this is what makes the $\widehat{\Delta}_i$'s a subset of the Δ_i's. For example, each $\widehat{\Delta}_5$ will keep rows and columns 1 and 2, then choose three out of the remaining six rows and columns. The number of combinations of x things taken y at a time is $x!/\big(y!\,(x-y)!\big)$. Hence the number of ways of choosing 3 out of the remaining 6 rows and columns is $6!/\big(3!\,(6-3)!\big) = 20$ ways, so there are 20 $\widehat{\Delta}_5$'s (compared with 56 Δ_5's and one D_5). Similarly, there are $6!/\big(4!\,(6-4)!\big) = 15$ $\widehat{\Delta}_6$'s (compared with 28 Δ_6's and one D_6), and $6!/\big(5!\,(6-5)!\big) = 6$ $\widehat{\Delta}_7$'s (compared with 8 Δ_7's and one D_7). There is only one $\widehat{\Delta}_8$; it equals Δ_8 and D_8. If you are curious, in this example the twenty $\widehat{\Delta}_5$'s contain the following rows and columns: 1 2 3 4 5 ($= D_5$), 1 2 3 4 6, 1 2 3 4 7, 1 2 3 4 8, 1 2 3 5 6, 1 2 3 5 7, 1 2 3 5 8, 1 2 3 6 7, 1 2 3 6 8, 1 2 3 7 8, 1 2 4 5 6, 1 2 4 5 7, 1 2 4 5 8, 1 2 4 6 7, 1 2 4 6 8, 1 2 4 7 8, 1 2 5 6 7, 1 2 5 6 8, 1 2 5 7 8, and 1 2 6 7 8. The fifteen $\widehat{\Delta}_6$'s contain the following rows and columns: 1 2 3 4 5 6 ($= D_6$), 1 2 3 4 5 7, 1 2 3 4 5 8, 1 2 3 4 6 7, 1 2 3 4 6 8, 1 2 3 4 7 8, 1 2 3 5 6 7, 1 2 3 5 6 8, 1 2 3 5 7 8, 1 2 3 6 7 8, 1 2 4 5 6 7, 1 2 4 5 6 8, 1 2 4 5 7 8, 1 2 4 6 7 8, and 1 2 5 6 7 8. The six $\widehat{\Delta}_7$'s contain the following rows and columns: 1 2 3 4 5 6 7 ($= D_7$), 1 2 3 4 5 6 8, 1 2 3 4 5 7 8, 1 2 3 4 6 7 8, 1 2 3 5 6 7 8, and 1 2 4 5 6 7 8.

4. Second-Order Conditions: Global

Let $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ be a point identified in Proposition 4. Let j be the number of equality constraints and k be the number of inequality constraints.

1. If $j = k = 0$ (an **unconstrained** problem) and $f(\mathbf{x})$ is convex, then $\mathbf{x}^*$ is a **global minimum** point of f in S. (The converse also holds.) [p. 631] Furthermore, if $j = k = 0$ and $f(\mathbf{x})$ is *strictly* convex, then $\mathbf{x}^*$ is the **unique** global minimum point of f in S. (The converse also holds.)

2. If $k = 0$ (only **equality** constraints) and $\mathscr{L}(\mathbf{x})$ is convex in $\mathbf{x}$, then $\mathbf{x}^*$ is a **global constrained minimum** point of f. [p. 667] Furthermore, if $k = 0$ and $\mathscr{L}(\mathbf{x})$ is *strictly* convex in $\mathbf{x}$, then $\mathbf{x}^*$ is the **unique** global constrained minimum point of f.

3. If $j = 0$ (only **inequality** constraints), if $g_1, \ldots, g_k$ are all concave, and if $f(\mathbf{x})$ is convex, then $\mathbf{x}^*$ is a **global constrained minimum** point of f. [p. 699] Furthermore, if $j = 0$ and $f(\mathbf{x})$ is *strictly* convex, then $\mathbf{x}^*$ is the **unique** global constrained minimum point of f.

4. Suppose either: (a) $j = 0$ (only inequality constraints); or (b) $j > 0$ but all the equality constraints are linear. Then if the binding g_1, $\ldots$, g_k are all quasiconcave, and if $f(\mathbf{x})$ is pseudoconvex, then $\mathbf{x}^*$ is a **global constrained minimum** point of f. [p. 700; Simon and Blume Th. 21.16 p. 528 and Th. 21.22 p. 532; and Mas-Colell, Whinston, and Green Th. M.K.3 pp. 961–2]

5. If the constraint set defined in Proposition 4 is a convex set—which occurs, for example, when $j = 0$ and g_1, $\ldots$, g_k are all quasiconcave—and if $f(\mathbf{x})$ is strictly quasiconvex, then $\mathbf{x}^*$ is the **unique global constrained minimum** point of f. If $f(\mathbf{x})$ is quasiconvex but not strictly so, then the set of solutions $\mathbf{x}^*$ is a convex set. [Mas-Colell, Whinston, and Green Th. M.K.4 p. 962. See also Simon and Blume Th. 21.23 p. 533.]

Note on (5): Strict quasiconcavity—and by analogy, strict quasiconvexity—is defined on p. 646 of your text. It means that the contour lines cannot have straight segments.

[Aside: Similarly, using the problem defined in Proposition 4′ and (10′):

1′. If $j = k = 0$ (an **unconstrained** problem) and $f(\mathbf{x})$ is concave, then $\mathbf{x}^*$ is a **global maximum** point of f in S. (The converse also holds.) Furthermore, if $j = k = 0$ and $f(\mathbf{x})$ is *strictly* concave, then $\mathbf{x}^*$ is the **unique** global maximum point of f in S. (The converse also holds.)

2′. If $k = 0$ (only **equality** constraints) and $\mathscr{L}(\mathbf{x})$ is concave in $\mathbf{x}$, then $\mathbf{x}^*$ is a **global constrained maximum** point of f. Furthermore, if $k = 0$ and $\mathscr{L}(\mathbf{x})$ is *strictly* concave in $\mathbf{x}$, then $\mathbf{x}^*$ is the **unique** global constrained maximum point of f.

3′. If $j = 0$ (only **inequality** constraints), if g_1, $\ldots$, g_k are all convex, and if $f(\mathbf{x})$ is concave, then $\mathbf{x}^*$ is a **global constrained maximum** point of f. Furthermore, if $j = 0$ and $f(\mathbf{x})$ is *strictly* concave, then $\mathbf{x}^*$ is the **unique** global constrained maximum point of f.

4′. Suppose either: (a) $j = 0$ (only inequality constraints); or (b) $j > 0$ but all the equality constraints are linear. Then if g_1, $\ldots$, g_k are all quasiconvex, and if $f(\mathbf{x})$ is pseudoconcave, then $\mathbf{x}^*$ is a **global constrained maximum** point of f.

5′. If the constraint set defined in Proposition 4′ is a convex set—which occurs, for example, when $j = 0$ and g_1, $\ldots$, g_k are all quasiconvex—and if $f(\mathbf{x})$ is strictly quasiconcave, then $\mathbf{x}^*$ is the **unique global constrained maximum** point of f. If $f(\mathbf{x})$ is quasiconcave but not strictly so, then the set of solutions $\mathbf{x}^*$ is a convex set.

]