On Second Order Conditions for Equality Constrained Extremum Problems

David M. Mandy*
Department of Economics
University of Missouri
118 Professional Building
Columbia, MO 65203 USA
mandyd@missouri.edu
573-882-1763
May 21, 2013

Abstract
We prove a relationship between the bordered Hessian in an equality constrained extremum problem and the Hessian of the equivalent lower-dimension unconstrained problem. This relationship can be used to derive principal minor conditions for the former from the relatively simple and accessible conditions for the latter.

It is surprisingly difficult to find a complete and accessible proof of conditions on principal minors of the bordered Hessian matrix for extremum problems with an arbitrary number \( n \) of choice variables and \( m < n \) equality side conditions. Calculus texts rarely venture beyond two choice variables and one constraint, while advanced calculus texts sometimes present the general problem but do not carry the proofs through to principal minor conditions (for example, Apostol [1, §13.7]).

*Comments on an earlier draft by Saku Aura, William Brock, Oksana Loginova, Peter Mueser, Gregory Piepmeyer and Xinghe Wang are gratefully acknowledged. As usual, I retain full responsibility for content.
Specialty texts such as Hestenes [5] do not focus on such practicalities. A careful, accessible and relatively complete modern treatment is in Simon and Blume [7], where definiteness of the matrix subject to side conditions is proven to be part of the sufficient conditions for an extremum (pp. 841-844), but the principal minor conditions equivalent to this definiteness are stated without proof (p. 389). One must typically supplement this type of discussion with proofs from Debreu [2] or Gantmacher [3, pp. 306-307] of the relationships between principal minors and (semi) definiteness subject to side conditions.

This is a bothersome situation due to the central role of optimality conditions in economics. Especially noticeable is a lack of accessible and complete proofs of necessary principal minor conditions for equality constrained problems because these conditions often form important refutable hypotheses of optimality-based theories.

The unconstrained case is usually presented in more complete form. For example, a complete proof of sufficient principal minor conditions can be obtained by combining Sections 16.4 and 30.4 in [7]. Hence one way to provide an accessible and complete proof of principal minor conditions when there are equality constraints is to apply the conditions for the unconstrained case to the lower-dimension version of the constrained problem obtained by substituting the constraint for a subset of the choice variables. This approach suffers, however, from the relative disutility of the resulting conditions. Generally, conditions on the bordered Hessian are easier to apply and more immediately usable, as originally emphasized by Hancock [4, Chapter 6] and Samuelson [6, Appendix A].

This paper derives the traditional principal minor conditions on the bordered Hessian directly from the corresponding conditions on the lower-dimension unconstrained problem. It builds naturally on textbook presentations of the uncon-
strained case without using any concepts except calculus and matrix algebra familiar to graduate students in economics. In particular, the approach completely eliminates consideration of quadratic forms subject to side conditions. Such consideration is unavoidable with extant proofs but is tedious and repetitive once principal minor conditions have been established for (semi) definite quadratic forms without side conditions. The new proof integrates constrained and unconstrained statements of principal minor conditions, both necessary and sufficient.

The traditional exclusive emphasis on the bordered Hessian stems from Hancock, who criticized the lower-dimension approach as “unsymmetric” in the treatment of choice variables [4, p. 103], and from Samuelson [6, p. 363], who apparently relied on Hancock in this regard.\footnote{Samuelson declares: “However, there is a great loss of symmetry in such a procedure since not all our variables are treated alike. Fortunately, by the use of an artifice which can be rigorously justified, it is possible to derive a more symmetrical set of conditions.” The “artifice” referred to here is the Lagrangian function.} The classical Hancock/Samuelson presentation remains the tradition today even though it is tedious and somewhat incomplete especially regarding necessary conditions. In contrast, the approach herein obtains the standard and more useful principal minor conditions on the bordered Hessian more directly from the “unsymmetric” approach. We note in passing that the traditional approach is equally “unsymmetric” when properly interpreted.

1 Notation

Denote by $A'$ the transpose of the matrix $A$. If $A$ is square of dimension $k$, $A^{(J)}$ denotes the principal submatrix consisting of rows and columns $J \subset \{1, \ldots, k\}$.

Assume $X \subset \mathbb{R}^n$ is an open set via the usual Euclidean metric; let $f : X \to \mathbb{R}^1$
be the objective and \( h: X \to \mathbb{R}^m \) for \( m < n \) be the equality constraint. The values of a vector-valued function like \( h \) are interpreted as column vectors in matrix equations. \( Dh(x) \) is the \( m \times n \) Jacobian matrix of \( h \) evaluated at \( x \in X \) [1, p. 351]. To avoid notational clutter, the arguments of a Jacobian are omitted when the point of evaluation is clear from the context. The Lagrangian function \( L: X \times \mathbb{R}^m \to \mathbb{R}^1 \) is defined by \( L(x, \lambda) = f(x) - \lambda^T h(x) \).

Let \( x = (y, z) \) be a partition of the choice vector \( x \) into the first \( n - m \) and last \( m \) components so the constraint can be substituted for \( z \) to form the lower-dimension unconstrained problem. When differentiating with respect to part of a partition, notation like \( Dh_y(y, z) \) denotes the \( m \times (n - m) \) Jacobian of \( h \) with respect to \( y \) evaluated at \( (y, z) \).

For a real-valued function like \( f \), the \( n \times n \) matrix \( D^2 f(x) \equiv D(Df)'(x) \) is the Hessian matrix of \( f \) evaluated at \( x \in X \). As above, \( D^2 f_y(y, z) \) denotes the \( (n - m) \times (n - m) \) Hessian of \( f \) with respect to \( y \) evaluated at \( (y, z) \). Recall that the Hessian is symmetric at \( x \) when it both exists in a neighborhood of \( x \) and each entry is continuous at \( x \) [1, Theorem 12.13].

Constraint qualification is assumed throughout; without loss of generality assume the variables have been ordered so that \( Dh_z(y, z) \) is nonsingular at a critical point \( \hat{x} = (\hat{y}, \hat{z}) \in X \). If \( h \) is continuously differentiable in an open ball about \( \hat{x} \), the implicit function theorem [1, Theorem 13.7] ensures existence of an open ball \( B_{\epsilon}(\hat{y}) \subset \mathbb{R}^{n-m} \) and a unique continuously differentiable function \( \phi: B_{\epsilon}(\hat{y}) \to \mathbb{R}^m \) such that \((\hat{y}, \phi(\hat{y})) = \hat{x} \) and

\[
h(y, \phi(y)) = 0 \text{ for } y \in B_{\epsilon}(\hat{y}). \tag{1}\]

Therefore \( \hat{y} \) is a local extremum of the composite function \( \tilde{f}(y) = f(y, \phi(y)) \) over \( y \in B_{\epsilon}(\hat{y}) \) if and only if \( \hat{x} \) is a local extremum of \( f \) subject to \( h = 0 \). \( \tilde{f} \) is the lower-
2 Relationship between Hessian and Bordered Hessian

The following theorem provides the algebraic relationship between the Hessian of \( \tilde{f} \) and the Hessian of \( L \) at a stationary point of \( L \).

**Theorem 1.** For \( \hat{x} = (\hat{y}, \hat{z}) \in X \), assume:

1. \( h(\hat{x}) = 0 \),
2. \( D^2 f \) and \( D^2 h \) exist in an open ball about \( \hat{x} \) and are continuous at \( \hat{x} \), and
3. \( D_z h(\hat{y}, \hat{z}) \) is nonsingular.

If there exists \( \hat{\lambda} \in \mathbb{R}^m \) such that \( D_z L(\hat{y}, \hat{z}, \hat{\lambda}) = 0 \) then

\[
D^2 \tilde{f} = D^2_y L - (D_y (D_{(z,\lambda)} L)') (D^2_{(z,\lambda)} L)^{-1} D_y (D_{(z,\lambda)} L)' \text{ at } (\hat{x}, \hat{\lambda}). \tag{2}
\]

**Proof.** From the definition of \( L \),

\[
D^2_{(z,\lambda)} L = \begin{bmatrix}
D^2_z L & -(D_z h)' \\
-D_z h & 0_m
\end{bmatrix}. \tag{3}
\]

Using nonsingularity of \( D_z h(\hat{x}) \), it is straightforward to verify

\[
(D^2_{(z,\lambda)} L(\hat{y}, \hat{z}, \hat{\lambda}))^{-1} = - \begin{bmatrix}
0_m & (D_z h)^{-1} \\
((D_z h)')^{-1} & ((D_z h)')^{-1} D^2_z L (D_z h)^{-1}
\end{bmatrix}.
\]

Substitute this and \( D_y (D_{(z,\lambda)} L)' = [D_z (D_y L)' - (D_y h)']' \) into (2), and use symme-
try of \( D^2_x L \) (which is implied by continuity at \( \hat{x} \)), to obtain a restatement of (2) as

\[
D^2 \tilde{f} = D^2_y L + [D_z (D_y L)']' - (D_y h)'
\begin{bmatrix}
0 & (D_z h)^{-1} \\
((D_z h)')^{-1} & ((D_z h)')^{-1} D^2_z L (D_z h)^{-1}
\end{bmatrix}
\left[
\begin{array}{c}
(D_z (D_y L)')' \\
-D_y h
\end{array}
\right]
\]

\[
= D^2_y L - (D_y h)'((D_z h)')^{-1} D_y (D_z L)' - D_z (D_y L)'(D_z h)^{-1} D_y h
\]

\[
+ (D_y h)'((D_z h)')^{-1} D^2_z L (D_z h)^{-1} D_y h. \tag{4}
\]

Now differentiate (1) with respect to \( y \) using the matrix form of the Chain Rule [1, Section 12.10] to obtain

\[
D_y h(y, \phi(y)) + D_z h(y, \phi(y)) D \phi(y) = 0 \text{ for } y \in B_\epsilon(\hat{y}). \tag{5}
\]

Perform the same differentiation on the definition of \( \tilde{f} \), premultiply the former by \( \hat{\lambda}' \), and subtract the product from the latter to obtain

\[
D \tilde{f}(y) = D_y L(y, \phi(y), \hat{\lambda}) + D_z L(y, \phi(y), \hat{\lambda}) D \phi(y) \text{ for } y \in B_\epsilon(\hat{y}). \tag{6}
\]

As this holds in an open ball about \( \hat{y} \), we may differentiate again to obtain

\[
D^2 \tilde{f} = [D^2_y L + D_y ((D \phi)'(D_z L)')] + [D_z (D_y L)'+ D_z ((D \phi)'(D_z L)')] D \phi \text{ for } y \in B_\epsilon(\hat{y}).
\]

\( D \phi \) is an \( m \times m \) matrix; expanding the first term where \( D \phi \) appears in individual rows yields

\[
D_y ((D \phi)'(D_z L)') = D_y \left[
\begin{bmatrix}
(D_{y_1} \phi)'(D_z L)' \\
\vdots \\
(D_{y_{n-m}} \phi)'(D_z L)'
\end{bmatrix}
\right] = \left[
\begin{array}{c}
D_y ((D_{y_1} \phi)'(D_z L)') \\
\vdots \\
D_y ((D_{y_{n-m}} \phi)'(D_z L)')
\end{array}
\right].
\]
Apply the product rule to obtain
\[
D_y((D_y\phi)'(D_zL)') = (D_zL)D_y(D_y\phi) + (D_y\phi)'D_y(D_zL)' \quad \text{for } i = 1, \ldots, n - m.
\]

By assumption \(D_zL(\hat{y}, \phi(\hat{y}), \hat{\lambda}) = 0\), so evaluating at \(\hat{y}\) yields
\[
D_y((D_y\phi)'(D_zL)') = (D_y\phi)'D_y(D_zL)' \quad \text{for } i = 1, \ldots, n - m.
\]

Therefore
\[
D_y((D\phi)'(D_zL)') = \begin{bmatrix}
(D_y\phi)'D_y(D_zL)'
\vdots
(D_y\phi_{n-m}'D_y(D_zL)'
\end{bmatrix} = (D\phi)'D_y(D_zL)'.
\]

This same derivation shows \(D_z((D\phi)'(D_zL)') = (D\phi)'D_z(D_zL)' \quad \text{at } (\hat{y}, \phi(\hat{y}), \hat{\lambda}).\) So
\[
D^2\bar{f} = [D_y^2L + (D\phi)'D_y(D_zL)'] + [D_z(D_yL)' + (D\phi)'D_z^2L]D\phi.
\]

Substitute \(D\phi = -(D_zh)^{-1}D_yh\) from (5) to obtain (4). \(\square\)

**Corollary 1.** Assume the conditions of Theorem 1. Then
\[
|D^2L^{(J+)})| = (-1)^m |D_zh|^2 |D^2\bar{f}^{(J)}| \quad \text{for every } J \subset \{1, \ldots, n - m\},
\]

where \(J^+ = J \cup \{n - m + 1, \ldots, n + m\}\) and all derivatives are evaluated at \((\hat{y}, \phi(\hat{y}), \hat{\lambda})\).

That is, at a stationary point of the Lagrangian function, each border-preserving principal minor of the bordered Hessian of order greater than 2m is proportional to the corresponding principal minor of the Hessian for the lower-dimension unconstrained problem, with proportion \((-1)^m|D_zh|^2\).

**Proof.** Partition the Hessian of \(L\) as
\[
D^2L(y, z, \lambda) = \begin{bmatrix}
D_y^2L & (D_y(D(z,\lambda)L)')'
D_y(D(z,\lambda)L)' & D_{(z,\lambda)}^2L
\end{bmatrix}.
\]
Letting \((D_y(D_{(z,\lambda)} L))^{(J_c)}\) denote the submatrix formed from \(D_y(D_{(z,\lambda)} L)'\) by retaining the set \(J\) of columns (and retaining all rows), we have

\[
(D^2 L)^{(J_+)} = \begin{bmatrix}
(D_y L)^{(J)} & ((D_y(D_{(z,\lambda)} L))^{(J_c)})'

(D_y(D_{(z,\lambda)} L))^{(J_c)} & D^2_{(z,\lambda)} L
\end{bmatrix}.
\]

Applying the standard formula for determinants of partitioned matrices [3, pp. 45-46] yields

\[
\left| (D^2 L)^{(J_+)} \right| = \left| D^2_{(z,\lambda)} L \right| \times \left| (D^2 L)^{(J)} - ((D_y(D_{(z,\lambda)} L))^{(J_c)})'(D^2_{(z,\lambda)} L)^{-1}(D_y(D_{(z,\lambda)} L))^{(J_c)} \right|.
\]

From Theorem 1:

\[
(D^2 \tilde{f})^{(J)} = (D_y L)^{(J)} - ((D_y(D_{(z,\lambda)} L))^{(J_c)})'(D^2_{(z,\lambda)} L)^{-1}(D_y(D_{(z,\lambda)} L))^{(J_c)}.
\]

Therefore

\[
\left| (D^2 L)^{(J_+)} \right| = \left| D^2_{(z,\lambda)} L \right| \left| (D^2 \tilde{f})^{(J)} \right|.
\]

From (3), and again using standard properties of determinants, we have

\[
\left| D^2_{(z,\lambda)} L \right| = (-1)^m \begin{vmatrix}
-(D_z h)' & D^2_z L
0_m & -D_z h
\end{vmatrix} = (-1)^m \left| D_z h \right|^2.
\]

Substituting yields the result.

The term \(\left| D_z h \right|^2\) is positive in Corollary 1 because \(D_z h\) is assumed nonsingular. So Corollary 1 says the relevant border-preserving principal minors of the bordered Hessian differ in sign from the corresponding principal minors of the Hessian in the unconstrained problem by \((-1)^m\). 
3 Application to Conditions for Equality Constrained Extrema

As noted above, a complete proof of the standard sufficient conditions for an unconstrained extremum of $\tilde{f}$ can be found in [7, §16.4 & 30.4]:

**Theorem 2.** Assume $D^2 \tilde{f}$ exists in an open ball about $\hat{y}$ and is continuous at $\hat{y}$, and $D\tilde{f}(\hat{y}) = 0$. If every leading principal minor of $D^2 \tilde{f}(\hat{y})$ is positive (of order $k$ has sign $(-1)^k$) then $\hat{y}$ is a local minimum (maximum) of $\tilde{f}$.

Utilizing Corollary 1, we thus immediately obtain a complete proof of the traditional sufficient second order conditions for a constrained extremum directly from the unconstrained case.

**Corollary 2.** For $\hat{x} = (\hat{y}, \hat{z}) \in X$, assume:

1. $D^2 f$ and $D^2 h$ exist in an open ball about $\hat{x}$ and are continuous at $\hat{x}$,
2. $D_z h(\hat{y}, \hat{z})$ is nonsingular, and
3. There exists $\hat{\lambda} \in \mathbb{R}^m$ such that $D(x,\lambda)L(\hat{x}, \hat{\lambda}) = 0$.

If every leading border-preserving principal minor of $D^2 L(\hat{x}, \hat{\lambda})$ of order $k = 2m + 1, \ldots, n + m$ has sign $(-1)^m$ (sign $(-1)^{k-m}$) then $\hat{x}$ is a local minimum (maximum) of $f$ subject to $h = 0$.

**Proof.** Item 3 implies the $h(\hat{x}) = 0$ and $D_z L(\hat{x}, \hat{\lambda}) = 0$ requirements of Theorem 1 and, from (6), the $D\tilde{f}(\hat{y}) = 0$ requirement of Theorem 2. From Corollary 1, the principal minor conditions stated here imply the principal minor conditions of Theorem 2. Hence $\hat{y}$ is a local minimum (maximum) of $\tilde{f}$, implying $\hat{x}$ is a local minimum (maximum) of $f$ subject to $h = 0$. \[\square\]
The same connection works for necessary conditions, which are more important in some applications due to their role as refutable hypotheses in theories of optimizing behavior. Standard necessary conditions for an unconstrained extremum of $f$ are:

**Theorem 3.** Assume $D^2 \tilde{f}$ exists in an open ball about $\hat{y}$ and is continuous at $\hat{y}$. If $\hat{y}$ is a local minimum (maximum) of $\tilde{f}$ then $D \tilde{f}(\hat{y}) = 0$ and every principal minor of $D^2 \tilde{f}(\hat{y})$ is non-negative (of order $k$ has sign $(-1)^k$ or is zero).

There are many complete and accessible proofs of the standard necessary first order condition for a constrained extremum of $f$ (for example, [1, Theorem 13.12]):

**Theorem 4** (Lagrange Multiplier Rule). Assume $f$ and $h$ are continuously differentiable in an open ball about $\hat{x} \in X$ and $Dh(\hat{x})$ has rank $m$. If $\hat{x}$ is a local extremum of $f$ subject to $h = 0$ then there exists $\hat{\lambda} \in \mathbb{R}^m$ such that $D_x L(\hat{x}, \hat{\lambda}) = 0$.

Thus Corollary 1 immediately yields a complete proof of the traditional necessary second order conditions for a constrained extremum directly from the unconstrained case.

**Corollary 3.** For $\hat{x} = (\hat{y}, \hat{z}) \in X$, assume:

1. $D^2 f$ and $D^2 h$ exist in an open ball about $\hat{x}$ and are continuous at $\hat{x}$, and
2. $D_z h(\hat{y}, \hat{z})$ is nonsingular.

If $\hat{x}$ is a local minimum (maximum) of $f$ subject to $h = 0$ then there exists $\hat{\lambda} \in \mathbb{R}^m$ such that $D_{(x,\lambda)} L(\hat{x}, \hat{\lambda}) = 0$ and every border-preserving principal minor of $D^2 L(\hat{x}, \hat{\lambda})$ of order $k = 2m + 1, \ldots, n + m$ has sign $(-1)^m$ or is zero (sign $(-1)^{k-m}$ or is zero).

**Proof.** $\hat{y}$ is a local minimum (maximum) of $\tilde{f}$, so the principal minor conditions of Theorem 3 hold. The Lagrange Multiplier Rule yields $\hat{\lambda}$ such that $D_x L(\hat{x}, \hat{\lambda}) = 0$, and...
implying the $DL_z(\hat{x}, \hat{\lambda}) = 0$ requirement of Theorem 1. From Corollary 1, the principal minor conditions stated here hold.

Although Corollaries 2 and 3 are the standard useful conditions on the bordered Hessian, they are proven here in a nonstandard and simpler manner. No consideration of constrained quadratic forms is needed. Note that the choice variables $y$ and $z$ are treated asymmetrically in these standard conditions. Rows and columns involving $D^2L$ are never deleted in the traditional Hancock/Samuelson approach, precisely because only $n - m$ of the choice variables are truly free. Some arbitrarily chosen subset of the choice variables (but with respect to which the Jacobian of $h$ is nonsingular), equal in number to the number of constraints, is treated differently from the other choice variables no matter which approach is used. The loss of symmetry noted by Samuelson and Hancock is neither more nor less prominent in either approach.

References


