Liminal inequality constraints and second-order optimality conditions

Anthony Horsley and Andrew J. Wrobel
Department of Economics
London School of Economics
Houghton Street
London WC2A 2AE
United Kingdom.
(e-mail: LSEecon123@mac.com)

1 December 2003
CDAM Research Report LSE-CDAM-2003-16

Abstract

Liminal constraints are active inequality constraints with zero Lagrange multipliers. This is the borderline case between inactive and binding inequality constraints. In correctly formulated second-order conditions, inactive constraints are ignored, and binding ones are treated like equality constraints. But liminal constraints can neither be ignored nor be treated like equalities; examples are given. The persistent assertion in economics texts that all active constraints can be treated like equalities is untrue, and gives a false "sufficient" second-order condition.

Keywords: Liminal constraints, second-order conditions.
1991 *Mathematics Subject Classification*: Primary 90C30.
1991 *Journal of Economic Literature Classification*: C61.

1 Introduction

In a scalar optimisation problem, oriented to maximisation over an intersection of sublevel sets, an inequality constraint has a nonnegative Lagrange multiplier, which is actually zero if the constraint is inactive, i.e., met as a *strict* inequality. But the multiplier may be zero also if the constraint is active. We propose calling such a constraint *liminal*, and to reserve the term "binding" for an inequality constraint with a *nonzero* multiplier.[1]

Liminal constraints are allowed in second-order conditions (SOCs), but—as we point out here—they must be distinguished both from the binding and from the inactive constraints. They also complicate sensitivity analysis, and have to be excluded by assumption to get the ordinary differentiability of the optimal solution and its multipliers with respect to the problem's parameters: see, e.g., [2, Theorems 2.4.4 and 3.2.2] or [6, Theorem 1]. With liminal constraints, i.e., without the strict complementarity assumption, the solution (and its multipliers) is usually nondifferentiable, although it is still *directionally* differentiable: see, e.g., [2, Theorem 2.4.5] or [6, Theorems 3 and 4]. Some topics, such as the Le Chatelier Principle, involve liminal constraints of necessity: see, e.g., [8].

Such constraints seem to have no established name apart from "just active" or "degenerate active", which are occasionally used. In both mathematics and economics, the terms "active", "effective", "tight" and "binding" are used as synonyms, which is misleading: one feels that "binding" should mean more than merely "active".

Terminology apart, mathematicians have long been aware of the difference between binding and liminal constraints, and of its relevance for correct formulations of SOCs—which we state for completeness in Section 2 by following, e.g., [3] or [4]. But this has been overlooked in several generations of economics texts. Many do not discuss SOCs with inequality constraints at all; those which do, persistently mishandle liminal constraints. For example, [1], [9] and [10] treat *all* the active constraints like equality constraints. The principle of doing so whenever possible is sound and useful, but it must be limited to those constraints which are binding in *our* sense (i.e., have nonzero multipliers). To treat every active constraint, binding or liminal, in the same way is incorrect because this weakens the SOCs by requiring the Hessian to be definite on a smaller set. As a result, the so-called "sufficient" SOC of [1, Theorems I.2.5 and II.3.4 (pp. 11 and 38)], [9, 19.8 and p. 468] and [10, 1.E.16 (ii)] is in fact *insufficient* (when there is a liminal constraint). Its use can even lead to a strict minimum point being misidentified as a strict maximum (or *vice versa*) when there is no binding inequality:[2] see Section 3. In that example, the stationary point does not even meet the standard Necessary SOC (NSOC) for a maximum—but it does meet that of [1, Theorems I.2.4 and II.3.3 (pp. 10 and

[1] Of course, when multipliers are nonunique, this classification of constraints will depend on the choice of multipliers.

[2] When there *is* a binding inequality, maximum and minimum points are distinguished from each other already by the multiplier signs in the FOCs: for a point to be stationary for both maximisation and minimisation, all the active inequality constraints must be liminal.

37)], [9, p. 468] or [10, 1.E.16 (i)]. This is because *their* necessary SOC, though indeed necessary, is unnecessarily weak. Its power to eliminate stationary points as candidates for an optimum is thus diminished. (This shortcoming is found also in mathematical literature, e.g., [7, 5.3.2]; it is obviously much less of a fault than the false "sufficient" SOC.)

So, liminal constraints must not be treated like binding ones in the SOCs. Nor can they be simply ignored like inactive constraints—except when there is just *one* liminal constraint. A single liminal constraint *can* be ignored because definiteness of a quadratic form on a half-space is equivalent to its definiteness on the whole space. But when there are two or more liminal constraints, to ignore even one of them is to strengthen the SOCs. This produces a different kind of error: the "necessary" SOC is then *not* in fact a necessary condition. Its use can lead to a point being wrongly rejected as a candidate for an optimum even though it does meet the standard Sufficient SOC (SSOC), and therefore actually *is* a strict optimum (Section 4). As for the sufficient SOC that ignores liminal constraints, though indeed sufficient, it is unnecessarily stringent as an optimality test.[3] Its proper place is with the directional-derivative results of solution-sensitivity analysis without strict complementarity, where it is known as the Strong SSOC: see, e.g., [2, Theorem 2.4.5] or [6, Theorems 3 and 4].

It is, of course, only the usefulness and lasting value of works such as [1], [9] and [10] that makes their errors worth correcting.

2 The standard second-order multiplier rules

The maximand, f, is assumed to be defined and twice continuously differentiable on an open set $D \subseteq \mathbb{R}^n$, as are the constraint functions. The equality constraints are $h_e(x) = 0$ for $e = 1, 2, \ldots, m$; and the inequality constraints are $g_i(x) \leq 0$ for $i = 1, \ldots, l$. So the constraint set is

$$C = \{x \in D : h(x) = 0,\ g(x) \leq 0\}. \tag{1}$$

In matrix multiplication, the n-tuple of decision variables $x = (x_1, \ldots, x_n)$ is regarded as a column; its transpose is a row $x^{\mathrm{T}} = [x_1, \ldots, x_n]$. The scalar product $p^{\mathrm{T}}x$ is also denoted by $p \cdot x$. The Jacobian matrix of the map h: $\mathbb{R}^n \to \mathbb{R}^m$ is the $m \times n$ matrix of partial derivatives $\mathrm{D}h(x) = \left[\frac{\partial h_r}{\partial x_s}\right]_{r=1}^{m}{}_{s=1}^{n}$, i.e., its r-th row is the gradient vector $\nabla h_r(x)^{\mathrm{T}}$. The Lagrangian is

$$L(\mu, \lambda, x) := f(x) - \mu \cdot h(x) - \lambda \cdot g(x)$$

for every $\mu \in \mathbb{R}^m$, $\lambda \in \mathbb{R}^l$ and $x \in D \subseteq \mathbb{R}^n$. The set of all active inequality constraints is

$$\mathrm{A}(x) := \{i : g_i(x) = 0\}$$

[3] This is so in, e.g., [7, 5.3.4]; it is much less of a fault than the false "necessary" SOC.

and the set of all binding constraints is

$$\mathrm{B}(\lambda) := \{i : \lambda_i > 0\}.$$

Under the Complementary Slackness Condition (5), in addition to (3)–(4), a binding constraint is active—i.e., $\mathrm{B}(\overline{\lambda}) \subseteq \mathrm{A}(\overline{x})$. The set of all liminal constraints is $\mathrm{A}(\overline{x}) \setminus \mathrm{B}(\overline{\lambda})$.

Theorem 1 (SOCs for programmes with inequality constraints) *Assume that $\overline{x}$ is a stationary point for maximisation, supported by multipliers $\overline{\mu} \in \mathbb{R}^m$ and $\overline{\lambda} \in \mathbb{R}^l$—i.e., that it meets the Kuhn-Tucker First-Order Conditions (FOCs)*

$$h(\overline{x}) = 0 \tag{2}$$

$$g(\overline{x}) \leq 0 \tag{3}$$

$$0 \leq \overline{\lambda} \tag{4}$$

$$0 = \overline{\lambda} \cdot g(\overline{x}) \tag{5}$$

$$0 = \nabla_x L(\overline{\mu}, \overline{\lambda}, \overline{x}) = \nabla f(\overline{x})^{\mathrm{T}} - \overline{\mu}^{\mathrm{T}} \mathrm{D}h(\overline{x}) - \overline{\lambda}^{\mathrm{T}} \mathrm{D}g(\overline{x}) \tag{6}$$

or, in expanded form, $\nabla f(\overline{x}) = \sum_{e=1}^{m} \overline{\mu}_e \nabla h_e(\overline{x}) + \sum_{i=1}^{l} \overline{\lambda}_i \nabla g_i(\overline{x})$.

In addition, assume that the vectors $(\nabla h_e(\overline{x}))_{e=1}^{m}$ and $(\nabla g_i(\overline{x}))_{i \in \mathrm{A}(\overline{x})}$ are linearly independent. Then:

1. *(Necessary SOC) If $\overline{x}$ is a local maximum point of f on the set C defined by (1), then*

$$\Delta x^{\mathrm{T}} \mathrm{D}^2_{xx} L(\overline{\mu}, \overline{\lambda}, \overline{x}) \, \Delta x \leq 0 \tag{7}$$

for every Δx such that

$$\mathrm{D}h(\overline{x}) \, \Delta x = 0, \text{ i.e., } \nabla h_e(\overline{x}) \cdot \Delta x = 0 \quad \text{for each } e \tag{8}$$

$$\mathrm{D}g_{\mathrm{B}(\overline{\lambda})}(\overline{x}) \, \Delta x = 0, \text{ i.e., } \nabla g_i(\overline{x}) \cdot \Delta x = 0 \quad \text{for each } i \text{ with } \overline{\lambda}_i > 0 \tag{9}$$

$$\mathrm{D}g_{\mathrm{A}(\overline{x}) \setminus \mathrm{B}(\overline{\lambda})}(\overline{x}) \, \Delta x \leq 0, \text{ i.e., } \nabla g_i(\overline{x}) \cdot \Delta x \leq 0 \quad \text{for } i \text{ with } \overline{\lambda}_i = 0 \text{ and } g_i(\overline{x}) = 0. \tag{10}$$

2. *(Sufficient SOC) Conversely, if*

$$\Delta x^{\mathrm{T}} \mathrm{D}^2_{xx} L(\overline{\mu}, \overline{\lambda}, \overline{x}) \, \Delta x < 0 \tag{11}$$

for every nonzero Δx meeting (8)–(10), then $\overline{x}$ is a strict local maximum point of f on the set C defined by (1). What is more, there exist numbers $\epsilon > 0$ and $\zeta > 0$ such that $f(x) \leq f(\overline{x}) - \zeta \|x - \overline{x}\|^2$ for every $x \in C$ with $\epsilon > \|x - \overline{x}\| := ((x - \overline{x}) \cdot (x - \overline{x}))^{1/2}$.

Proof. See, e.g., [3, Theorems 1.10.2 and 1.10.3] or [4, Theorems 7.4 and 7.5]. For a concise but complete exposition based on these sources, see [5]. ■

Comment: The assumption of linearly independent constraints can be weakened to that of regularity, in the sense of [3] or [4]. That is, it suffices to assume that $\overline{x}$ is regular as a point of the set

$$C_{\mathrm{b}}\left(\overline{\lambda}\right) := \left\{x : h(x) = 0,\ g_{\mathrm{B}(\overline{\lambda})}(x) = 0,\ g_{\mathrm{A}(\overline{x})\setminus\mathrm{B}(\overline{\lambda})} \leq 0\right\}$$

or, more precisely, that $\overline{x}$ is regular for the representation of $C_{\mathrm{b}}\left(\overline{\lambda}\right)$ by the functions $\left(h, g_{\mathrm{B}(\overline{\lambda})}\right)$ as equality constraints and $g_{\setminus\mathrm{B}(\overline{\lambda})}$ as inequality constraints. (Regularity depends on the particular functions representing the constraint set, and not just on the set itself.)

3 Counterexample to the "sufficient" SOC that treats a liminal constraint as binding

A one-variable example can be given.

Example 2 *Maximise $f(x) := x^2$ over x subject to $g(x) := x \leq 0$. The point $\overline{x} = 0$ is a global constrained strict* minimum *(and even the global unconstrained strict minimum). But it meets both the FOC and the false "sufficient" SOC for a local constrained (strict)* maximum *as stated in [1, Theorems I.2.5 and II.3.4 (pp. 11 and 38)], [9, 19.8 and p. 468] and [10, 1.E.16 (ii)]—i.e., with an equality instead of the correct inequality in (10).*

In this example, there is no local constrained maximum. For either maximisation or minimisation, the one stationary point is $\overline{x} = 0$, supported by the unique multiplier $\overline{\lambda} = 0$ (i.e., the pair $\left(\overline{\lambda}, \overline{x}\right) = (0, 0)$ meets (3)–(6), and also (2) vacuously). This is actually the minimum point (and even the global unconstrained minimum). It must fail the *correct* SSOC for a *maximum*, and it does: the Lagrangian's Hessian is $\mathrm{D}^2_{xx}L\left(\overline{\lambda}, \overline{x}\right) = f''(\overline{x}) - \overline{\lambda}g''(\overline{x}) = 2$, and its negative definiteness is to be tested on every increment $\Delta x \neq 0$ such that $0 \geq \mathrm{D}g(\overline{x})\,\Delta x = g'(0)\,\Delta x = \Delta x$, i.e., for $\Delta x < 0$—as per (10). Since $\Delta x \mathrm{D}^2_{xx}L \Delta x = 2(\Delta x)^2 > 0$, the point fails even the NSOC (for a maximum).

But the constraint is liminal; and if it is, incorrectly, treated as one that is binding—i.e., like an equality constraint—then the minimum point does meet the "sufficient" SOC for a *maximum* (local, constrained). This is because, after changing the inequality to an equality in (10), the Hessian's definiteness is to be tested only on any nonzero increment Δx such that $0 = \mathrm{D}g(\overline{x})\,\Delta x = g'(0)\,\Delta x = \Delta x$. Since no such Δx exists, there is nothing to check—and thus the minimum passes for a maximum!

4 Counterexample to the "necessary" SOC that ignores liminal constraints

A two-constraint, two-variable example is needed.

Example 3 *Maximise* $f(x_1, x_2) := x_1 x_2$ *over* (x_1, x_2) *subject to* $g_1(x_1, x_2) := x_1 \leq 0$ *and* $g_2(x_1, x_2) := -x_2 \leq 0$*. The point* $(\overline{x}_1, \overline{x}_2) = (0, 0)$ *is a global constrained maximum. But it does not meet the false "necessary" SOC (for a local constrained maximum)—i.e., it fails the test if (10) is deleted.*

In this example, the (global) constrained maximum points are those with $x_1 \leq 0$, $x_2 \geq 0$ and $x_1 x_2 = 0$. These are also the only stationary points for maximisation. There is no local constrained minimum. The only point stationary for minimisation is $(\overline{x}_1, \overline{x}_2)$ $= (0, 0)$, supported by the multipliers $\left(\overline{\lambda}_1, \overline{\lambda}_2\right) = (0, 0)$.[4] Being actually a *maximum* point, it must meet the *correct* NSOC for a maximum, and it does: the Lagrangian's Hessian form is

$$\Delta x^{\mathrm{T}} \mathrm{D}^2_{xx} L\left(\overline{\lambda}, \overline{x}\right) \Delta x = \Delta x^{\mathrm{T}} \mathrm{D}^2_{xx} f\left(\overline{x}\right) \Delta x = \begin{bmatrix} \Delta x_1 & \Delta x_2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} = 2\Delta x_1 \Delta x_2$$

and, as per (10), its negative semidefiniteness is to be tested on every increment Δx $= (\Delta x_1, \Delta x_2)$ such that

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \geq \begin{bmatrix} \nabla g_1(0,0)^{\mathrm{T}} \\ \nabla g_2(0,0)^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} = \begin{bmatrix} \Delta x_1 \\ -\Delta x_2 \end{bmatrix}$$

i.e., for $\Delta x_1 \leq 0$ and $\Delta x_2 \geq 0$. Since this implies that $0 \geq 2\Delta x_1 \Delta x_2 = \Delta x^{\mathrm{T}} \mathrm{D}^2_{xx} L \Delta x$, the point passes the test.

But *both* constraints are liminal, and if they are, incorrectly, ignored (treated like inactive constraints), then the "necessary" SOC for a constrained maximum becomes identical to the NSOC for an unconstrained (local) maximum. This test fails the point because the Lagrangian's Hessian form, $2\Delta x_1 \Delta x_2$, is indefinite on $\mathbb{R}^2$: without the constraints, $(0, 0)$ is a saddle point. (Ignoring just one of the two liminal constraints has the same effect, since definiteness on a half-space is equivalent to definiteness on the whole space.)

Comment: The example can be modified so that the point $(0, 0)$ is the strict global maximum and, furthermore, meets the standard SSOC (for a maximum): take the same f with $g_1(x_1, x_2) := x_1 + \epsilon x_2$ and $g_2(x_1, x_2) := -x_2 - \epsilon x_1$ for a small $\epsilon > 0$. Negativity of

[4] The multiplier system is unique because the constraint gradients are linearly independent: $\nabla g_1(0,0)$ $= (1, 0)$ and $\nabla g_2(0,0) = (0, -1)$.

$\Delta x^{\mathrm{T}} \mathrm{D}^2_{xx} L \Delta x = 2\Delta x_1 \Delta x_2$ is to be tested on every nonzero Δx with $0 \geq \nabla g_i(\overline{x}) \cdot \Delta x$ for $i = 1, 2$—i.e., for nonzero $(\Delta x_1, \Delta x_2)$ with both $\Delta x_1 + \epsilon \Delta x_2 \leq 0$ and $\epsilon \Delta x_1 + \Delta x_2 \geq 0$. Since this implies that

$$0 > (\Delta x_1 + \epsilon \Delta x_2)(\epsilon \Delta x_1 + \Delta x_2) - \epsilon (\Delta x_1)^2 - \epsilon (\Delta x_2)^2 = \left(1 + \epsilon^2\right) \Delta x_1 \Delta x_2$$

the point $(0, 0)$ meets the SSOC.

References

[1] El-Hodiri, M. A. (1971): *Constrained extrema: introduction to the differentiable case with economic applications*. (Lecture Notes in Operations Research and Mathematical Systems, vol. 56). Berlin-Heidelberg-New York: Springer.

[2] Fiacco, A. V. (1983): *Introduction to sensitivity and stability analysis in nonlinear programming*. New York: Academic Press.

[3] Hestenes, M. R. (1966): *Variational calculus and optimal control theory*. New York-London-Sydney: Wiley.

[4] Hestenes, M. R. (1975): *Optimization theory: the finite-dimensional case*. New York-London-Sydney: Wiley.

[5] Horsley, A., and A. J. Wrobel (2003): "First and second-order conditions in constrained optimisation", CDAM Research Report LSE-CDAM-2003-15.

[6] Jittorntrum, A. V. (1984): "Solution point differentiability without strict complementarity in nonlinear programming", *Mathematical Programming Study*, 21, 127–138.

[7] Ponstein, J. (1980): *Approaches to the theory of optimisation*. Cambridge-New York-Melbourne: Cambridge University Press.

[8] Silberberg, E. (1971): "The Le Chatelier Principle as a corollary to a generalized envelope theorem", *Journal of Economic Theory*, 3, 146–155.

[9] Simon, C., and L. Blume (1994): *Mathematics for economists*. New York-London: Norton.

[10] Takayama, A. (1985): *Mathematical economics*. Cambridge-London-New York: Cambridge University Press.