Optimization and Lagrange Multipliers:

Non-C^1 Constraints and "Minimal" Constraint Qualifications

by

Leonid Hurwicz and Marcel K. Richter

Discussion Paper No. 280, March 1995

Center for Economic Research
Department of Economics
University of Minnesota
Minneapolis, MN 55455

Optimization and Lagrange Multipliers:

Non-C^1 Constraints

and

"Minimal" Constraint Qualifications

by

Leonid Hurwicz

and

Marcel K. Richter

University of Minnesota

ABSTRACT

When do Lagrange multipliers exist at constrained maxima? In this paper we establish:

a) Existence of multipliers, replacing C^1 smoothness of equality constraint functions by differentiability (for Jacobian constraint qualifications) or, for both equalities and inequalities, by the existence of partial derivatives (for path-type constraint qualifications). This unifies the treatment of equality and inequality constraints.

b) A notion of "minimal" *Jacobian* constraint qualifications. We give new Jacobian qualifications and prove they are minimal over certain classes of constraint functions.

c) A *path-type* constraint qualification, weaker than previous constraint qualifications, that is necessary and sufficient for existence of multipliers. (It only assumes existence of partial derivatives.)

A survey of earlier results, beginning with Lagrange's own multipliers for equality constraints is contained in the last section. Among others, it notes contributions and formulations by Weierstrass; Bolza; Bliss; Carathéodory; Karush; Kuhn and Tucker; Arrow, Hurwicz, and Uzawa; Mangasarian and Fromovitz; and Gould and Tolle.

Optimization and Lagrange Multipliers:

Non-C^1 Constraints

and

"Minimal" Constraint Qualifications*

by

Leonid Hurwicz

and

Marcel K. Richter

University of Minnesota

I Introduction

Constrained optimization is central to economics, and Lagrange multipliers are a basic tool in solving such problems, both in theory and in practice. In this paper we extend the applicability of Lagrange multipliers to a wider class of problems, by reducing smoothness hypotheses (for classical Lagrange inequality constraints as well modern inequality constraints), and by reducing constraint qualifications to minimal levels.

We focus on constrained maximization using calculus, and in particular on first order necessary conditions. While there have been important contributions that go beyond the use of calculus tools (subdifferentials, Clarke cones, etc.), calculus tools are important, especially for giving explicit results that are helpful in many economic applications.(1) First order necessary conditions are particularly valuable in problems where convexity conditions are not satisfied, as in economic models involving production sets with increasing returns to scale, and in characterizing the role of marginal cost pricing in such economies.

---

* We are indebted to Professor Kam-Chau Wong, Chinese University of Hong Kong, for valuable comments on an earlier version.

(1) For some of the economic applications of Lagrangean techniques see Hicks [23], Samuelson [41], Takayama [44, pp. 129–168], and many others.

Two centuries years ago, Lagrange (with Euler as precursor) introduced "indeterminate" multipliers, placing the necessary consequences of constrained maximization in a general framework [**19**].(2) His result for equality constraints is known in most analysis texts today as the Lagrange Multiplier Theorem. Half a century ago, Karush [**26**] obtained an analogue for inequality constraints, and he was followed independently by Kuhn and Tucker's celebrated paper [**30**] a decade later.

To extend those, and several more recent Lagrange multiplier results, to a wider class of problems, we employ:

a) reduced smoothness requirements on constraints and maximands;

b) weaker constraint qualifications;

c) notions of minimal constraint qualifications.

Under (a), we weaken the differentiability hypotheses for proving existence of Lagrange multipliers. In the classical equality-constrained context, for example, we require only continuity and differentiability of the constraints, and only at the maximizer, instead of the usual *continuous* differentiability. Under (b), we provide new constraint qualifications that are weaker, yet still guarantee existence of Lagrange multipliers. These cover problems with inequality constraints and problems with mixed constraints. Under (c), we first introduce notions of minimal constraint qualifications of two types. The first type is defined by the Jacobian matrix at the maximand, and the second is defined by more general properties of paths lying in, or related to the constraint set. Using these notions, we prove that the Jacobian conditions introduced in (b) are minimal Jacobian conditions for "Lagrange regularity." We also prove that the path conditions introduced in (b) are necessary and sufficient for existence of Lagrange multipliers.

There are two mathematical bases on which our results depend. The first is a strong form of the Theorem of the Alternative from linear algebra. (This is closely related to the tools, such as Farkas' Lemma or Motzkin's Transposition Theorem, which others have used for solving linear equalities and inequalities.) Persistent exploitation of the algebraic result allows us to present simple proofs, to clarify the separate roles of algebra and analysis, and to make transparent the essential role of constraint qualifications.

The second base is a new implicit function theorem [**24**] with very weak differentiability hypotheses. The classical approach ([**6**], [**7**], [**11**]) to Lagrange mul-

(2) See QUATRIEME SECTION, paragraphs 1–8, pages 44–49 of *Méchanique Analitique*. A similar development is in [**33**], SECTION QUATRIÈME, Sections 2–8, pages 77–83, which is an 1888 (fourth) edition of [**19**] under the name *Mécanique Analytique*; "Méthode des multiplicateurs" occurs as the heading here, but not in the first edition.

See also [**32**], SECONDE PARTIE, Chapter XI, paragraph, pp. 291–292.

tipliers with equality constraints used the classical Implicit Function Theorem (assuming C^1 functions), and for that reason it was necessary, in the Lagrangean theorems, to make sure that the equality constraints were C^1. (For inequality constraints, there was a breakthrough in Kuhn and Tucker [**30**], which already replaced C^1 by simple differentiability.)

For mixed equality-inequality optimizations problems, which are typical of economics, the C^1 assumption has always been retained for equality constraints while in [**35**] it was relaxed to differentiability at the maximizing point for the inequality constraints. Here we will formulate more uniform conditions, where the C^1 hypothesis is dropped for the equality constraints as well as for the inequality constraints. We have been able to accomplish this by proving a generalization of the Implicit Function Theorem that reduces the C^1 hypothesis to continuity and differentiability.

Because we focus here on first order (necessary) differential conditions for maximization, our theorems do not introduce the type of convexity or concavity conditions used in [**4**] or [**35**]. However, one could easily adjoin convexity conditions to our hypotheses; that would yield something more general than the results in those contributions.

We do not shy away from redundancies or explanations that would be obvious to a seasoned mathematician, but may be helpful to a student.

Since we are weaving together many strands, the following outline may be helpful.

II. Notation and Terminology: page 5

Also includes an index of the major definitions.

40

III. Constrained Maximization: page 8

Defines the Lagrange constrained maximization problem and the role of Lagrange multipliers, leading to definitions of the basic types of Lagrange regularity. Proves the algebraic Fundamental Lemma, on which later proofs will be based, and which explains the need for constraint qualifications.

IV. The Jacobian Criterion: page 15

Defines Jacobian constraint qualifications, their sufficiency for Lagrange regularity, and the notion of minimal sufficiency. States the Jacobian Criterion for mixed, inequality, and equality problems.

V. The Jacobian Criterion is Sufficient: page 19

VI. The Jacobian Criterion Is Minimal: page 30

VII. The Tangency-Path Criterion: page 34 Explains the need to go beyond Jacobian constraint qualifications. States the Tangency-Path Criterion for mixed, inequality, and equality problems, and proves it is necessary as well as sufficient for Lagrange regularity.

VIII. Comparison of Jacobian and Tangency-Path Conditions: page 50 Compares the verifiability and computability aspects of Jacobian and Tangency-Path Conditions.

IX. Appendix: page 52
The two mathematical results on which the main results are based: a Theorem of the Alternative and the Non-C^1 Implicit Function Theorem.

X. Historical Comments and Comparisons: page 54.

II Notation and Terminology

We denote the set of natural numbers by $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$, and the set of real numbers by $\mathbb{R}$. For $u \in \mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^m$ we write $u = (x, y)$, where $x \in \mathbb{R}^k$ and $y \in \mathbb{R}^m$. When $F : \mathbb{R}^k \times \mathbb{R}^m \to \mathbb{R}^p$ and when all its partial derivatives $\frac{\partial F}{\partial u^i}(\bar{u})$ exist at $\bar{u}$, then $F'(u)$ denotes the $p \times n$ Jacobian matrix:

$$\begin{bmatrix} \frac{\partial F^1(u)}{\partial u_1} & \cdots & \frac{\partial F^1(u)}{\partial u_n} \\ \vdots & & \vdots \\ \frac{\partial F^p(u)}{\partial u_1} & \cdots & \frac{\partial F^p(u)}{\partial u_n} \end{bmatrix}. \tag{1}$$

When the function F not only has partial derivatives at $\bar{u}$, but has the stronger[3] property of possessing a Fréchet derivative at $\bar{u}$, then we denote it by $F_u(\bar{u})$; of course in this case the linear transformation $F_u(\bar{u})$ is represented with respect to the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^p$ by the matrix $F'(\bar{u})$,[4] and so, for any $z \in \mathbb{R}^n$:

When F is Fréchet differentiable at $\bar{u}$, then $F'(\bar{u})z$ is a matrix representation of the vector $F_u(\bar{u})z$. (2)

Similarly, when $F(\,\cdot\,, \bar{y})$ possesses a Fréchet derivative at $\bar{x}$, it is denoted by $F_x(\bar{x}, \bar{y})$; and when $F(\bar{x}, \,\cdot\,)$ possesses a Fréchet derivative at $\bar{y}$, it is denoted by $F_y(\bar{x}, \bar{y})$. For brevity, differentiability will always mean Fréchet differentiability unless otherwise noted.

A function $F : X \to Y$ is said to be *locally continuous* at a point $x \in X$ if F is continuous on some neighborhood of x.

If A and B are linear subspaces of $\mathbb{R}^n$, and if every element u of $\mathbb{R}^n$ can be written uniquely as a sum of elements in A and B: $u = x + y$, where $x \in A$ and $y \in B$, then we say that $\mathbb{R}^n$ is the *direct sum* of A and B, and we write:

$$\mathbb{R}^n = A \oplus B. \tag{3}$$

Because all norms in a finite dimensional linear vector space lead to the same notions of convergence and differentiability, it will not matter which norm we use. As convenient, we will use three different norms on our basic spaces: for

---

(3) Cf. [**40**, p. 240, Exercise 14].

(4) Cf. [**40**, Theorem 9.17, p. 215].

any $v \in \mathbb{R}^l$,

$$\text{the } \textit{Euclidean norm}\text{: } \|v\| = \sqrt{v_1^2 + \cdots + v_l^2} \tag{4a}$$

$$\text{the } \textit{maximum norm}\text{: } \|v\| = \max\{|v_1|, \ldots, |v_l|\} \tag{4a}$$

$$\text{the } \textit{sum norm}\text{: for any normed subspaces } A \text{ and } B, \tag{4c}$$

$$\text{if } \mathbb{R}^n = A \oplus B, \text{ and } v = a + b \text{ with } a \in A \text{ and } b \in B,$$

$$\text{then } \|v\| = \|a\| + \|b\|.$$

For any $v \in \mathbb{R}^l$ and any real γ, we denote the closed γ-ball about v by

$$B_\gamma(v) = \{v + w \in \mathbb{R}^l : \|w\| \leqq \gamma\}. \tag{5}$$

For x and y in $\mathbb{R}^l$:

$$\begin{aligned} x \geqq y \quad & \text{means } x_i \geqq y_i \text{ for all } i = 1, \ldots, n \\ x \geq y \quad & \text{means } x \geqq y \text{ and } x \neq y \\ x > y \quad & \text{means } x_i > y_i \text{ for all } i = 1, \ldots, n. \end{aligned} \tag{6}$$

For an open set $U \subseteqq \mathbb{R}^n$ and for $f = (f^1, \ldots, f^q) : U \to \mathbb{R}^q$, the notation $f \geqq 0$ means $f^i(x_1, \ldots, x^n) \geqq 0$ for each $i = 1, \ldots, q$ and for each $i = 1, \ldots, q$.

For any subset S of $\mathbb{R}^l$,

$$\begin{aligned} \mathrm{ch}(S) &= \text{the convex hull of } S \\ &= \text{the intersection of all convex sets } T \supseteqq S \\ &= \{t_0 x_0 + t_1 x_1 + \cdots + t_m x_m : m \in \mathbb{N} \ \& \ x_0, x_1, \ldots, x_m \in S \\ &\qquad \& \ t_0, t_1, \ldots, t_m \geqq 0 \ \& \ t_0 + t_1 + \cdots + t_m = 1\} \end{aligned} \tag{7a}$$

$$\mathrm{cl}(S) = \text{the topological closure of } S \tag{7b}$$

$$\begin{aligned} \mathrm{cone}(S) &= \text{the conical closure of } S \\ &= \{tx : x \in S \ \& \ \text{real } t \geqq 0\} \end{aligned} \tag{7c}$$

$$\begin{aligned} \mathrm{wedge}(S) &= \text{the wedge generated by } S \\ &= \text{the convex cone generated by } S \\ &= \mathrm{cone}(\mathrm{ch}(S)) \end{aligned} \tag{7d}$$

$$\begin{aligned} \mathrm{span}(S) &= \text{the linear subspace generated by } S \\ &= \{t_0 + t_1 x_1 + \cdots + t_m x_m : m \in \mathbb{N} \\ &\qquad \& \ x_1, \ldots, x_m \in S \ \& \ t_0, t_1, \ldots, t_m \in \mathbb{R}\}. \end{aligned} \tag{7e}$$

For any subsets S and T of $\mathbb{R}^l$:

$$\begin{aligned} S + T &= \{x + y : x \in S \ \& \ y \in T\} \\ S + \emptyset &= S = \emptyset + S. \end{aligned} \tag{8a}$$

Nonnegative orthants are denoted by:

$$\mathbb{R}^l_+ = \{x \in \mathbb{R}^l : x \geqq 0\}. \tag{9}$$

When interpreted for matrix multiplication, elements $x \in \mathbb{R}^l$ will be treated as column vectors. For any such x, we denote the transpose by x^T, which we treat as a row vector for matrix multiplication.

We collect page references here for some notions to be defined later:

Lagrange regularity	page 10
Jacobian constraint qualifications	page 15
Sufficient	page 15
Minimal	page 15
Jacobian Criterion	page 17
Tangency-Path Criterion	page 37.

III Constrained Maximization

We are concerned with maximizing a real valued function f on an open set $U \subseteqq \mathbb{R}^n$, subject to conditions of the form $g^1 \geqq 0, \ldots, g^m \geqq 0$ and $h^1 = 0, \ldots, h^k = 0$, where the g^i and h^i are also functions from U to $\mathbb{R}^n$. Either g or h, or both, may be absent. Let $g = (g^1, \ldots, g^m)$ and $h = (h^1, \ldots, h^k)$, and define the *constraint set* by

$$C(g,h) = \{x \in u : \forall i_{\,i=1,\ldots,m}\ g^i(x) \geqq 0 \ \ \& \ \ \forall j_{\,j=1,\ldots,k}\ h^j(x) = 0\}; \tag{10}$$

i.e., when both g and h are present:

$$C(g,h) = \{x \in U : g(x) \geqq 0 \ \ \& \ \ h(x) = 0\}. \tag{11}$$

A point $\bar{u} \in U$ is said to *maximize f on $C(g,h)$* if:

$$\begin{aligned} &\bar{u} \in C(g,h) \\ &\forall u_{\,u \in C(g,h)}\ f(\bar{u}) \geqq f(u). \end{aligned} \tag{12}$$

When g is absent, this is a problem of classical mathematics; and when h is absent, it is well known in economics as a "Kuhn-Tucker" problem. We write $C(g)$ when h is absent, and $C(h)$ when g is absent, trusting the context will avoid ambiguity. When both g and h are absent, i.e., when there are no constraints, then $C(g,h) = \mathbb{R}^n$.

III. A Lagrange Regularity

Suppose that $\bar{u} \in U$ maximizes f on $C(g,h)$ and that f, g, and h have partial derivatives at $\bar{u}$. We are interested in the existence of "Lagrange multipliers" $\lambda \in \mathbb{R}^m_+$ and $\mu \in \mathbb{R}^k$, satisfying:

$$f'(\bar{u}) + \lambda^T g'(\bar{u}) + \mu^T h'(\bar{u}) = 0. \tag{13}$$

In addition, we will also want λ to satisfy a nonnegativity condition. But simple examples show that — quite apart from any additional requirements, there may exist no λ and μ satisfying (13). Consider inequalities defined by:[5]

$$\begin{aligned} g^1(u_1,u_2) &= u_2 - u_1^2 \geqq 0 \\ g^2(u_1,u_2) &= -u_2 - u_1^2 \geqq 0, \end{aligned} \tag{14}$$

---

[5] This is analogous to Slater's example in [42].

or equalities defined by:

$$\begin{aligned} h^1(u_1, u_2) &= u_2 - u_1^2 = 0 \\ h^2(u_1, u_2) &= -u_2 - u_1^2 = 0. \end{aligned} \tag{15}$$

In each case the constraint set is the singleton $\{0\}$, and so any function $f(u_1, u_2)$ has a constrained maximum there. But the function $f(u_1, u_2) = u_1$ does not admit any λ or μ satisfying (13) when the constraints are (14) or (15), since $\frac{\partial f}{\partial u_1}(0,0) = 1$.

Special assumptions are therefore needed to guarantee the existence of Lagrange multipliers λ and μ. To state these conditions, as well as to sharpen the question, we first explain when an inequality constraint is "binding."

Binding Constraints. When $\bar{u}$ maximizes f subject to $g = (g^1, \ldots, g^m) \geqq 0$ and $h = (h^1, \ldots, h^k) = 0$, some of the $g^j(\bar{u})$ may be positive, while others may equal 0. If $g^j(\bar{u}) > 0$ then small movements from $\bar{u}$ will, by continuity, preserve the constraint property $g^j(\bar{u}) \geqq 0$; the same is true if g^j is constant throughout some neighborhood of $\bar{u}$. Since the condition (13) is a local condition, we can then restrict attention to a small enough neighborhood on which $g^j \geqq 0$ will hold throughout — and so we can effectively ignore such a constraint. To distinguish the constraints that cannot be so ignored from those that can, we say that a constraint g^i is *binding at* $\bar{u}$ if $g^j(\bar{u}) \geqq 0$ for all $j = 1, \ldots, m$, and if $\bar{u}$ belongs to the boundary of the "individual" constraint set $\{x \in \mathbb{R}^n : g^i(x) \geqq 0\}$. Then we say that g^i is a *binding constraint at* $\bar{u}$. For example, if g^i is continuous and if either $g^i(\bar{u}) > 0$ or g^i is constant over a neighborhood of $\bar{u}$, then g^i is not binding at $\bar{u}$.[6]

With U an open subset of $\mathbb{R}^n$, with $\bar{u} \in U$, and with $g : U \to \mathbb{R}^m$, we partition $M = \{1, \ldots, m\}$ to distinguish the g^i that are binding at $\bar{u}$ from the others:[7]

$$\begin{aligned} I &= \{i \in M : \bar{u} \text{ is in the boundary of } \{u \in U : g^i(u) \geqq 0\}\} \\ J &= M \setminus I \end{aligned} \tag{16}$$

(either I, J, or both may be empty). When I or J is nonempty, we write:

$$\begin{aligned} g_I &: U \to \mathbb{R}^p \\ g_J &: U \to \mathbb{R}^{m-p} \\ g &= (g_I, g_J) : U \to \mathbb{R}^p \times \mathbb{R}^{m-p} = \mathbb{R}^m. \end{aligned} \tag{17}$$

---

(6) Note that if $g^1 = g^2$, then *each* constraint may be binding at $\bar{u}$.

(7) One could introduce an analogous distinction between binding and nonbinding equality constraints, but we shall not do so.

Regularity. Now we seek conditions under which the fact that $\bar{u} \in U$ maximizes f on $C(g,h)$ implies the existence of a $\lambda \in \mathbb{R}^m$ and a $\mu \in \mathbb{R}^k$ such that:

$$f'(\bar{u}) + \lambda^T g'(\bar{u}) + \mu^T h'(\bar{u}) = 0, \tag{18a}$$

i.e.,

$$\begin{aligned} \frac{\partial f}{\partial u_i}(\bar{u}) + \lambda_1 \frac{\partial g^1}{\partial u_i}(\bar{u}) + \ldots + \lambda_m \frac{\partial g^m}{\partial u_i}(\bar{u}) \\ + \mu_1 \frac{\partial h^1}{\partial u_i}(\bar{u}) + \ldots + \mu_k \frac{\partial h^k}{\partial u_i}(\bar{u}) = 0 \qquad (i = 1, \ldots, n), \end{aligned} \tag{18b}$$

and such that:[8]

$$\lambda \geqq 0 \tag{20a}$$

and

$$\lambda_i = 0 \text{ if } g^i \text{ is not binding at } \bar{u} \qquad (\text{i.e., } i \in J). \tag{20b}$$

Since (18) and (20) are linear in the λ_i and the μ_j, it follows that, when the values of the derivatives of g^i and h^j are known, then the existence of λ_i and μ_j satisfying (18) and (20) can be determined algorithmically by "elimination of quantifiers" (e.g. by Fourier elimination).[9] But the question is often asked in a different form.

Rather than seeking solvability of (18) and (20) for λ and μ given particular g^i, h^i, and f, one often asks whether given functions g^i and h^j allow solvability for *all* f in some family $\mathcal{F}$ that are maximized on $C(g,h)$ at $\bar{u}$. (We shall always assume that each $f \in \mathcal{F}$ has partial derivatives at $\bar{u}$.) So we say that (g,h) is *Lagrange mixed-regular for* $(\bar{u}, \mathcal{F})$ if: [10]

i) all the g^i and h^j all have partial derivatives at $\bar{u}$;

ii) for every function $f \in \mathcal{F}$, if $\bar{u}$ maximizes f on $C(g,h)$ then (18) and (20) hold for some (λ, μ).

We have a corresponding definition when h is absent; we say that g is *Lagrange inequality-regular* for $(\bar{u}, \mathcal{F})$ if:

---

[8] It is tempting to suppose that (20) implies $\lambda_i > 0$ whenever g^i is binding at $\bar{u}$. Simple examples show that is not true. Consider this example in $\mathbb{R}^1$:

$$\begin{aligned} f(x) &= -x^2 \\ g(x) &= x \\ \bar{u} &= 0. \end{aligned} \tag{19}$$

Here the constraint g is binding at the maximizer $\bar{u} = 0$, but $\lambda = 0$ is required by (18).

[9] Cf. [**29**]; [**43**,chapter 1].

[10] The terminology is modified from [**4**]. We really should call this *Lagrange regularity for* $(U, \bar{u}, \mathcal{F})$; for brevity, we omit the reference to U.

i) all the g^i have partial derivatives at $\bar{u}$;

ii) for every function $f \in \mathcal{F}$, if $\bar{u}$ maximizes f on $C(g)$ then:

$$f'(\bar{u}) + \lambda^T g'(\bar{u}) = 0 \tag{21}$$

and (20) hold for some λ.

And when g is absent we say that h is *Lagrange equality-regular* for $(\bar{u}, \mathcal{F})$ if:

i) all the h^j have partial derivatives at $\bar{u}$;

ii) for every function $f \in \mathcal{F}$, if $\bar{u}$ maximizes f on $C(h)$ then:

$$f'(\bar{u}) + \mu^T h'(\bar{u}) = 0 \tag{22}$$

for some μ.

We will often refer simply to Lagrange regularity, counting on the context to make the type clear. When both g and h are absent, then we say that we have *Lagrange regularity for* $(\bar{u}, \mathcal{F})$ if, for every $f \in \mathcal{F}$, if $\bar{u}$ maximizes f on U then $f'(\bar{u}) = 0$.

And now, seeking λ and μ that satisfy (18) and (20) for *all* such f, the answer is less trivial. We begin by restating the problem in a simpler form with a natural

Reduction. Let $g : U \to \mathbb{R}^m$, $h : U \to \mathbb{R}^k$, and $f : U \to \mathbb{R}$ have partial derivatives at $\bar{u}$.[11] Suppose that the binding constraints are $g_I = (g^1, \ldots, g^p)$. Then a solution $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^k$ to (18) and (20) exists if and only if there exists a $\tilde{\lambda} \in \mathbb{R}^p$ and a $\mu \in \mathbb{R}^k$ such that:

$$f'(\bar{u}) + \tilde{\lambda}^T g_I'(\bar{u}) + \mu^T h'(\bar{u}) = 0, \tag{23a}$$

i.e.,

$$\begin{aligned} \frac{\partial f}{\partial u_i}(\bar{u}) + \tilde{\lambda}_1 \frac{\partial g^1}{\partial u_i}(\bar{u}) + \ldots + \tilde{\lambda}_p \frac{\partial g^p}{\partial u_i}(\bar{u}) \\ + \mu_1 \frac{\partial h^1}{\partial u_i}(\bar{u}) + \ldots + \mu_k \frac{\partial h^k}{\partial u_i}(\bar{u}) = 0 \qquad (i = 1, \ldots, n), \end{aligned} \tag{23b}$$

and such that:

$$\tilde{\lambda} \geqq 0. \tag{24}$$

---

[11] Actually, the $g^{p+1}, \ldots, g^m$ need not have partial derivatives, since they don't appear in (23), and they effectively don't appear in (18) since the corresponding λ_i vanish according to (20b).

Remark. In defining Lagrange regularity, we assumed that partial derivatives of the g^i, h^j, and f exist at $\bar{u}$, even when not binding, in order to give meaning to (18). It is possible to avoid assuming that the nonbinding g^i possess partial derivatives at $\bar{u}$, by using (23) and (24) in place of (18) and (20). While this would be more general, it would be less convenient to apply in situations where the structure of the constraint set is not known a priori.

III. B The Fundamental Lemma. Constraint Qualifications

Although our basic maximization question is one of analysis, it is useful to view it algebraically. For basically what it asks is whether the vector $-f'(\bar{u})$ is a nonnegative linear combination of the vectors $g^{i\prime}(\bar{u})$ plus a linear combination of the $h^i(\bar{u})$ — i.e., whether it lies in the sum of the wedge (the convex cone) generated by the $g^{i\prime}(\bar{u})$ vectors plus the span of the $h^{i\prime}(\bar{u})$ vectors. Thus separating out algebraic from analytical aspects of the problem will be quite helpful.

Our answers all rest on the following simple consequence of the Theorem of the Alternative.

Fundamental Lemma. Let U be an open subset of $\mathbb{R}^n$, and let $\bar{u} \in U$.

A) Suppose $f : U \to \mathbb{R}$, $g : U \to \mathbb{R}^p$, and $h : U \to \mathbb{R}^k$ are functions on U that have partial derivatives at $\bar{u}$, with $g(\bar{u}) = 0$, and suppose there does *not* exist $(\lambda, \mu) \in \mathbb{R}^p \times \mathbb{R}^k$ satisfying the following two conditions:

$$\text{i) } f'(\bar{u}) + \lambda^T g'(\bar{u}) + \mu^T h'(\bar{u}) = 0, \tag{25a}$$

i.e., such that:

$$\begin{aligned} \frac{\partial f}{\partial u_i}(\bar{u}) + \lambda_1 \frac{\partial g^1}{\partial u_i}(\bar{u}) + \ldots + \lambda_p \frac{\partial g^p}{\partial u_i}(\bar{u}) \\ + \mu_1 \frac{\partial h^1}{\partial u_i}(\bar{u}) + \ldots + \mu_k \frac{\partial h^k}{\partial u_i}(\bar{u}) = 0 \qquad (i = 1, \ldots, n), \end{aligned} \tag{25b}$$

and:

$$\text{ii) } \lambda \geqq 0. \tag{26a}$$

Then there exists a $\tilde{z} \in \mathbb{R}^n$ such that:

$$g'(\bar{u})\tilde{z} \geqq 0 \tag{27a}$$

$$h'(\bar{u})\tilde{z} = 0 \tag{27b}$$

$$f'(\bar{u}) \cdot \tilde{z} > 0. \tag{27c}$$

B) When g is absent, part (A) holds if we delete references to, and terms and relations involving λ or g.

C) When h is absent, part (A) holds if we delete references to, and terms and relations involving μ or h.

Proof. This is almost an immediate application of the Theorem of the Alternative.[12]

Part A: By hypothesis there do not exist λ_i and μ_j satisfying:

$$\begin{aligned}
&\lambda_1 \frac{\partial g^1}{\partial u_1}(\bar{u}) + \cdots + \lambda_p \frac{\partial g^p}{\partial u_1}(\bar{u}) \\
&\qquad + \mu_1 \frac{\partial h^1}{\partial u_1}(\bar{u}) + \cdots + \mu_k \frac{\partial h^k}{\partial u_1}(\bar{u}) = -\frac{\partial f}{\partial u_1}(\bar{u}) \\
&\quad \vdots \qquad\qquad\qquad \vdots \\
&\lambda_1 \frac{\partial g^1}{\partial u_n}(\bar{u}) + \cdots + \lambda_p \frac{\partial g^p}{\partial u_n}(\bar{u}) \\
&\qquad + \mu_1 \frac{\partial h^1}{\partial u_n}(\bar{u}) + \cdots + \mu_k \frac{\partial h^k}{\partial u_n}(\bar{u}) = -\frac{\partial f}{\partial u_n}(\bar{u})
\end{aligned} \tag{28a}$$

$$\begin{aligned}
&\lambda_1 1 + \lambda_2 0 + \cdots + \lambda_p 0 \geqq 0 \\
&\qquad\qquad \ddots \qquad\qquad \vdots \\
&\lambda_1 0 + \cdots + \lambda_{p-1} 0 + \lambda_p 1 \geqq 0.
\end{aligned} \tag{28b}$$

So by part (II) of the Theorem of the Alternative there exist $z \in \mathbb{R}^n$ and $v \in \mathbb{R}^p$ with $v \geqq 0$, such that:

$$z_1 \frac{\partial g^j}{\partial u_1}(\bar{u}) + \cdots + z_n \frac{\partial g^j}{\partial u_p}(\bar{u}) + v_j = 0 \qquad (j = 1, \ldots, n) \tag{29a}$$

$$z_1 \frac{\partial h^j}{\partial u_1}(\bar{u}) + \cdots + z_n \frac{\partial h^j}{\partial u_n}(\bar{u}) = 0 \qquad (j = 1, \ldots, k) \tag{29b}$$

$$-\left(z_1 \frac{\partial f}{\partial u_1}(\bar{u}) + \cdots + z_n \frac{\partial f}{\partial u_n}(\bar{u})\right) > 0. \tag{29c}$$

Defining $\tilde{z} = -z$, and taking account of the fact that $v \geqq 0$, we have (27).

Parts (B, C): The proof is analogous to that for part (A) above, using the appropriate section of part (II) of the Theorem of the Alternative. ∎

---

[12] See the Appendix.

We note that the lemma is a purely algebraic statement about derivatives. Though it does not involve maximization, it is basic to maximization theory. The following intuition shows why.

Suppose that f, g, and h are differentiable at $\bar{u}$, which maximizes f on the constraint set $C(g,h)$. If (g,h) is not Lagrange regular, then the lemma implies that (27) holds. If there were a path $u(\cdot)$ with values $u(t)$ lying in $C(g,h)$, starting at $\bar{u}$ and with derivative $u'(0) = \tilde{z}$, then[13] we could rewrite (27) as:

$$\begin{aligned} \frac{d}{dt} g(u(t))|_{t=0} &= g'(\bar{u})\tilde{z} && (30a) \\ &\geqq 0 \qquad \text{(by (27}a\text{))} \\ \frac{d}{dt} h'(u(t))|_{t=0} &= h'(\bar{u}) \cdot \tilde{z} && (30b) \\ &= 0 \qquad \text{(by (27}b\text{))} \\ \frac{d}{dt} f(u(t))|_{t=0} &= f'(\bar{u}) \cdot \tilde{z} && (30c) \\ &> 0 \qquad \text{(by (27}c\text{))}. \end{aligned}$$

So small movements along the path would keep us in the constraint set but increase the value of f. And that would contradict the assumption that $\bar{u}$ maximizes f on the constraint set $C(g,h)$.

Thus the algebra highlights the role of analysis. As we know from (14) and (15), some restrictions on the constraint functions g and h are required by Lagrange regularity. We now see that one such restriction is the existence of paths $u(\cdot)$ with $u'(0) = \tilde{z}$ for any $\tilde{z}$ satisfying (27). (In the non-Lagrange-regular examples (14) and (15), such paths did not exist.) For historical reasons, such hypotheses are typically called *constraint qualifications*.[14]

Two general types of constraint qualifications have been developed.[15] The first is a direct attack on the problem. It simply asserts that any $\tilde{z}$ satisfying $(27a, b)$ is indeed the derivative of some path lying in $C(g,h)$. We will call these *path conditions*.

The second type is computationally oriented. These constraint qualifications

---

[13] If g, h, and f are differentiable at $\bar{u}$.

[14] Cf. [31].

[15] Slater [42] proposed a constraint qualification that does not fall into the two types we discuss. It was intended, however, for application to the Saddle Point Equivalence Theorem, assuming concavity of the constraint functions g^i, rather than for a proof of Lagrange regularity. Because of the convexity of the constraint set defined by such g^i, Slater's condition implies a path-type constraint qualification of the type we discuss below. (See the historical comments in Section X below.)

are expressed as algebraic properties of the Jacobian matrix

$$\begin{bmatrix} g'(\bar{u}) \\ h'(\bar{u}) \end{bmatrix}. \tag{31}$$

They may involve its nonsingularity, the span of its rows, or other algebraic aspects. We call these *Jacobian conditions.*

In simple examples, checking whether a path condition holds can be rather easy. That is the case with (14) and (15). In more complicated examples, it may be much harder. By contrast, the Jacobian conditions that we discuss are decidable in an algorithmic fashion, as we will see later.

IV The Jacobian Criterion

Our Jacobian condition will guarantee that every $\tilde{z}$ satisfying $(27a, b)$ (or the corresponding variant when g or h is absent) is the derivative of some path lying in $C(g, h)$ (or $C(g)$ or $C(h)$). In Theorem 1 we will prove the condition is sufficient for Lagrange regularity. And then in Theorem 2 we will prove it is as weak as possible in the class of sufficient Jacobian conditions. In order to make these notions precise, we use the following terminology.

Sufficient and minimal Jacobian constraint qualifications. Let $\bar{u} \in U \subseteqq \mathbb{R}^n$, let $\mathcal{G}$ be a set of functions $g : U \to \mathbb{R}^m$, let $\mathcal{H}$ be a set of functions $h : U \to \mathbb{R}^k$, and let $\mathcal{F}$ be a set of functions $f : U \to \mathbb{R}$. Assume that all functions in $\mathcal{G}$, $\mathcal{H}$, and $\mathcal{F}$ have partial derivatives at $\bar{u}$.

By a *Jacobian mixed-constraint qualification for* $(\bar{u}, \mathcal{G}, \mathcal{H})$ we mean a property $\mathbb{Q}$ of some members of the set $\{\begin{pmatrix} g'(\bar{u}) \\ h'(\bar{u}) \end{pmatrix} : g \in \mathcal{G} \ \& \ h \in \mathcal{H}\}$ of $(m+k) \times n$ (real) matrices. To each property $\mathbb{Q}$ corresponds the set Q of matrices with that property.

We say that a Jacobian mixed-constraint qualification $\mathbb{Q}$ for $(\bar{u}, \mathcal{G}, \mathcal{H})$ is $(\bar{u}, \mathcal{G}, \mathcal{H}, \mathcal{F})$*-sufficient for Lagrange mixed-regularity* if: for all $g \in \mathcal{G}$ and $h \in \mathcal{H}$,

$$\begin{pmatrix} g'(\bar{u}) \\ h'(\bar{u}) \end{pmatrix} \in Q \ \Rightarrow \ (g, h) \text{ is Lagrange mixed-regular for } (\bar{u}, \mathcal{F}). \tag{32}$$

We say that a Jacobian constraint qualification $\mathbb{Q}$ for $(\bar{u}, \mathcal{G}, \mathcal{H})$ is *minimally* $(\bar{u}, \mathcal{G}, \mathcal{H}, \mathcal{F})$*-sufficient for Lagrange mixed-regularity* if:

i) $\mathbb{Q}$ is $(\bar{u}, \mathcal{G}, \mathcal{H}, \mathcal{F})$-sufficient for Lagrange mixed-regularity in the sense of (32);

ii) no weaker Jacobian property (i.e., no proper superset of Q) is $(\bar{u}, \mathcal{G}, \mathcal{H}, \mathcal{F})$-sufficient for Lagrange mixed-regularity: if $\begin{pmatrix} g'(\bar{u}) \\ h'(\bar{u}) \end{pmatrix} \notin Q$ then there is some $\hat{g} \in \mathcal{G}$ and some $\hat{h} \in \mathcal{H}$ with $\hat{g}'(\bar{u}) = g'(\bar{u})$ and $\hat{h}'(\bar{u}) = h'(\bar{u})$ for which $(\hat{g}, \hat{h})$ is not Lagrange mixed-regular for $(\bar{u}, \mathcal{F})$.

We may abbreviate "minimally sufficient" to "minimal."

Analogous definitions of sufficiency and minimality apply with functions g for the inequalities problem, and with functions h for the equalities problem.

It is clear that whether or not a Jacobian condition is minimally sufficient for Lagrange regularity depends on the set U, the element $\bar{u}$, and the classes $\mathcal{G}$ and $\mathcal{H}$ of constraint functions under consideration, as well as on the class $\mathcal{F}$ of maximand functions. In what follows, the set U will always be an open subset of $\mathbb{R}^n$. For any $\bar{u} \in U$ we make these definitions:

$\mathcal{G}_D(\bar{u})$ is the set of functions $g : U \to \mathbb{R}^m$ that are differentiable at $\bar{u}$.

$\mathcal{G}_G(\bar{u})$ is the set of functions $g : U \to \mathbb{R}^m$ that have partial derivatives at $\bar{u}$ and are Gâteaux differentiable at $\bar{u}$ in a dense set of directions.

$\mathcal{G}_P(\bar{u})$ is the set of functions $g : U \to \mathbb{R}^m$ that have partial derivatives at $\bar{u}$.

$\mathcal{H}_D(\bar{u})$ is the set of functions $h : U \to \mathbb{R}^k$ that are differentiable at $\bar{u}$.

$\mathcal{H}_{DC}(\bar{u})$ is the set of functions $h : U \to \mathbb{R}^k$ that are differentiable at $\bar{u}$ and locally continuous at $\bar{u}$.

$\mathcal{H}_{C^1}(\bar{u})$ is the set of functions $h : U \to \mathbb{R}^k$ that are C^1 at $\bar{u}$.

$\mathcal{F}_D(\bar{u})$ is the set of functions $f : U \to \mathbb{R}^{m+k}$ that are differentiable at $\bar{u}$.

$\mathcal{F}_G(\bar{u})$ is the set of functions $f : U \to \mathbb{R}^{m+k}$ that have partial derivatives at $\bar{u}$ and are Gâteaux differentiable at $\bar{u}$ in a dense set of directions.

Our main objective in the rest of this section is to state a new Jacobian constraint qualification, the Jacobian Criterion. The purpose of Section V is to prove it is sufficient for Lagrange regularity with respect to some important classes $\mathcal{G}$, $\mathcal{H}$, and $\mathcal{F}$. The purpose of Section VI is to prove it is minimally sufficient with respect to those same classes.

We now state the main Jacobian condition of interest, in forms for the mixed, the inequalities, and the equalities problems.

The Jacobian Criterion. Suppose U is an open subset of $\mathbb{R}^n$ and $g : U \to \mathbb{R}^m$ and $h : U \to \mathbb{R}^k$ have partial derivatives at some $\bar{u} \in U$.[16] Let $a(1), \ldots, a(p)$ be the rows of the $p \times n$ Jacobian matrix $g_I'(\bar{u})$, and let $b(1), \ldots, b(k)$ be the rows of the $k \times n$ Jacobian matrix $h'(\bar{u})$.

A) (Mixed.) We say that (g, h) satisfies the *(Mixed-Problem) Jacobian Criterion at* $\bar{u}$ if one of the following two mutually exclusive conditions holds:[17]

a) $\operatorname{rank}(h'(\bar{u})) = k$ (i.e. $\operatorname{rank}(h'(\bar{u}))$ is maximal) (33*a*)

and there exists a $\xi \in \mathbb{R}^n$ such that:

$$g_I'(\bar{u})\xi > 0$$
$$h'(\bar{u})\xi = 0;$$

b) $\operatorname{wedge}(a(1), \ldots, a(p)) + \operatorname{span}(b(1), \ldots, b(k))) = \mathbb{R}^n$ (33*b*)

i.e.:

$$\{v \in \mathbb{R}^n : \exists t_{\,0 \leqq t \in \mathbb{R}^p}\ \exists z_{\,z \in \mathbb{R}^k}\ t_1 a(1) + \cdots + t_p a(p) + z_1 b(1) + \cdots + z_k b(k)\} = \mathbb{R}^n.$$

B) (Inequalities.) We say that g satisfies the *(Inequality-Problem) Jacobian Criterion at* $\bar{u}$ if one of the following two mutually exclusive conditions holds:[17]

a) I is nonempty and there exists a $\xi \in \mathbb{R}^n$ such that: (34*a*)

$$g_I'(\bar{u})\xi > 0;$$

b) $\operatorname{wedge}(a(1), \ldots, a(p)) = \mathbb{R}^n$, (34*b*)

i.e.:

$$\{v \in \mathbb{R}^n : \exists t_{\,0 \leqq t \in \mathbb{R}^p}\ t_1 a(1) + \cdots + t_p a(p)\} = \mathbb{R}^n.$$

C) (Equalities.) We say that h satisfies the *(Equality-Problem) Jacobian Criterion at* $\bar{u}$ if the classical *maximum rank condition*

$$\operatorname{rank}(h'(\bar{u})) = \min\{k, n\} \tag{35}$$

holds.

---

[16] For g we only need to assume existence of partial derivatives for components of g_I.

[17] I is defined in (16).

Remarks. (Mixed.) Alternative (33a) is Mangasarian's generalization[18] of the Arrow-Hurwicz-Uzawa Constraint Qualification.[19] The new alternative (33b) is natural in view of the fact that the basic Lagrange multiplier equation (18) says that $-f'(\bar{u})$ lies in the sum of the wedge and the span. We will see that the Criterion is sufficient for mixed Lagrange regularity. While it is not truly necessary for Lagrange mixed-regularity, we will show it is necessary in the sense of being a minimal *Jacobian* condition.

For the mixed case, the Jacobian Criterion requires that either the sum of the wedge and span is "small," with the wedge lying strictly on one side of some hyperplane and the span lying in it, or else it is "large," spanning the whole space. In fact, that observation proves that parts (a) and (b) are mutually exclusive.

Note that it is not sufficient to have two different vectors ξ satisfying respectively (33a) and (33b). See the Remark following Corollary 1a below.

(Inequalities.) When only inequalities are present, the Jacobian Criterion requires that either the wedge generated by the $g_I'(\bar{u})$ is "small," lying strictly on one side of some hyperplane, or else it is "large," spanning the whole space. Again, that observation proves that parts (a) and (b) are mutually exclusive. What the Criterion rules out are the "borderline" cases where neither is true — i.e., where whenever the spanned convex cone lies on one side of some hyperplane, it does not lie *strictly* on one side. Examples (14) above and (90) below are such examples, with $g^{1\prime}(0)$ and $g^{2\prime}(0)$ pointing in opposite directions. Indeed, the general exception to the Jacobian Criterion occurs when the $g^{i\prime}(\bar{u})$ all lie on one side of some hyperplane, but some $g^{j\prime}(\bar{u})$ points in the opposite direction from some convex combination of the other $g^{i\prime}(\bar{u})$.

(Equalities.) When only equality constraints are present, the Jacobian Criterion is the familiar hypothesis in the classical Lagrange Multiplier Theorem.

Removable Constraints. In many applications, Lagrange multipliers satisfying (13) are sought as a step in finding a constrained maximizing point $\bar{u}$. If the Jacobian Criterion fails, all is not lost. For it may be that some constraint can be removed, without changing the constraint set C near $\bar{u}$; and the Jacobian Criterion might be satisfied for the reduced set of constraints.

---

[18] The "modified Arrow-Hurwicz-Uzawa constraint qualification" in [35, pp. 172–173].

[19] Cf. the hypothesis of Theorem 3 in [4].

In such situations, theorems using the Jacobian rank conditions should be applied to the rank of a "maximally reduced" set of constraints.[20]

Example 1: $h^1 = h^2$;
Example 2: $g^1 \geqq 0 \Leftrightarrow g^2 \geqq 0$;
Example 3: $g^3 = h^3$.

In these examples, h^2, g^2, and g^3 can be eliminated. Note also that a non-binding constraint is always removable.

This dependence of Lagrange regularity on the constraint functions, rather than on the constraint set, is illustrated in Remark (iii), page 39, for path conditions rather than Jacobian conditions.

V The Jacobian Criterion Is Sufficient

We now prove that the Jacobian Criterion is sufficient for Lagrange regularity — i.e., sufficient to guarantee the existence of Lagrange multipliers.

Theorem 1. Let U be an open subset of $\mathbb{R}^n$, and let $\bar{u} \in U$.

A) (Equalities and inequalities.) The Jacobian Criterion (A)[21] is $(\bar{u}, \mathcal{G}_D(\bar{u}), \mathcal{H}_{DC}(\bar{u}), \mathcal{F}_D(\bar{u}))$-sufficient for Lagrange mixed-regularity.[22] In other words:

Suppose g is differentiable at $\bar{u}$, and h is differentiable at $\bar{u}$ and locally continuous at $\bar{u}$. If the Jacobian Criterion $(33a, b)$ holds for (g, h) at $\bar{u}$, then (g, h) is Lagrange mixed-regular for $(\bar{u}, \mathcal{F}_D(\bar{u}))$.

In particular, under the assumptions just made: if $\bar{u}$ maximizes $f : U \to \mathbb{R}$ subject to $g \geqq 0$ and $h = 0$, i.e., if:

$$g(\bar{u}) \geqq 0 \tag{36a}$$

$$h(\bar{u}) = 0 \tag{36b}$$

$$\forall u_{\,u \in U \;\&\; g(u) \geqq 0 \;\&\; h(u)=0}\; f(\bar{u}) \geqq f(u), \tag{36c}$$

---

[20] Cf. Condition R_1 in [2], p. 8, where $g^\dagger$ represents a reduced set of constraints. (Page 162 in [3].)

[21] Page 17(A).

[22] Sufficiency is defined in p. 15.

and if f is differentiable at $\bar{u}$, then there exists a $\lambda \in \mathbb{R}^m$ and a $\mu \in \mathbb{R}^k$ such that:

$$f'(\bar{u}) + \lambda^T g'(\bar{u}) + \mu^T h'(\bar{u}) = 0, \tag{37a}$$

i.e.,

$$\begin{aligned} \frac{\partial f}{\partial u_i}(\bar{u}) + \lambda_1 \frac{\partial g^1}{\partial u_i}(\bar{u}) + \ldots + \lambda_m \frac{\partial g^m}{\partial u_i}(\bar{u}) \\ + \mu_1 \frac{\partial h^1}{\partial u_i}(\bar{u}) + \ldots + \mu_k \frac{\partial h^k}{\partial u_i}(\bar{u}) = 0 \qquad (i = 1, \ldots, n), \end{aligned} \tag{37b}$$

and such that:

$$\lambda \geqq 0 \tag{38a}$$

$$\lambda_i = 0 \text{ if } i \in J. \tag{38b}$$

B) (Inequalities.)[23] The Jacobian Criterion (B)[24] is $(\bar{u}, \mathcal{G}_G(\bar{u}), \mathcal{F}_G(\bar{u}))$-sufficient for Lagrange inequality-regularity. In other words:

Suppose $g : U \longrightarrow \mathbb{R}^m$ is Gâteaux differentiable at $\bar{u}$ for a dense set of directions including the coordinate directions. If the Jacobian Criterion (34) holds for g at $\bar{u}$, then g is Lagrange inequality-regular for $(\bar{u}, \mathcal{F}_G(\bar{u}))$.[25]

In particular, under the assumptions just made: if

$$g(\bar{u}) \geqq 0 \tag{39a}$$

$$\forall u_{\,u \in U \,\&\, g(u) \geqq 0}\; f(\bar{u}) \geqq f(u), \tag{39b}$$

and if f is Gâteaux differentiable at $\bar{u}$, then there exists a $\lambda \in \mathbb{R}^m$ such that:

$$f'(\bar{u}) + \lambda^T g'(\bar{u}) = 0, \tag{40a}$$

i.e.:

$$\frac{\partial f}{\partial u_i}(\bar{u}) + \lambda_1 \frac{\partial g^1}{\partial u_i}(\bar{u}) + \ldots + \lambda_m \frac{\partial g^m}{\partial u_i}(\bar{u}) = 0 \qquad (i = 1, \ldots, n), \tag{40b}$$

and such that:[26]

$$\lambda \geqq 0 \tag{41a}$$

$$\lambda_j = 0 \text{ if } j \in J. \tag{41b}$$

---

[23] In this case g is present and h is absent. Equivalently, we can set $h = 0$ in part (A) delete refences to, and terms involving, μ or h.

[24] Page 17(B).

[25] A fortiori, then, g is Lagrange inequality-regular for Fréchet differentiable functions, i.e., for $(\bar{u}, \mathcal{F}_D(\bar{u}))$. Note that the assumption implies that g has partial derivatives at $\bar{u}$.

[26] J is defined in (16).

C) (Equalities.)[27] The Jacobian Criterion (C)[28] is $(\bar{u}, \mathcal{H}_{DC}(\bar{u}), \mathcal{F}_D(\bar{u}))$-sufficient for Lagrange mixed-regularity. In other words:

Suppose that h is differentiable at $\bar{u}$ and locally continuous at $\bar{u}$. If the Jacobian Criterion, the classical rank condition (35),

$$\operatorname{rank}(h'(\bar{u})) = \min\{k, n\} \tag{42}$$

holds for h at $\bar{u}$, then h is Lagrange regular for $(\bar{u}, \mathcal{F}_D(\bar{u}))$.

In particular, under the assumptions just made: if

$$h(\bar{u}) = 0 \tag{43a}$$

$$\forall u_{\ u \in U \ \& \ h(u)=0}\ f(\bar{u}) \geqq f(u), \tag{43b}$$

and if f is differentiable at $\bar{u}$, then there exists a $\mu \in I\!R^k$ such that:

$$f'(\bar{u}) + \mu^T h'(\bar{u}) = 0, \tag{44a}$$

i.e., there exist real numbers $\mu_1, \ldots, \mu_k$ such that:

$$\frac{\partial f}{\partial u_i}(\bar{u}) + \mu_1 \frac{\partial h^1}{\partial u_i}(\bar{u}) + \ldots + \mu_k \frac{\partial h^k}{\partial u_i}(\bar{u}) = 0 \qquad (i = 1, \ldots, n). \tag{44b}$$

Remarks. (Mixed Problem.) This theorem has weaker hypotheses than earlier results, in two respects. First, the Jacobian Criterion is weaker, and second, the differentiability conditions on the equality constraints h^i are weaker.[29] For details, see the Historical Comments and Comparisons.

The continuity hypothesis on h can be weakened significantly, as indicated in remark (iii) below on equalities; but it cannot be dispensed with altogether, as shown by the example below in remark (iv) on equalities.

(Inequalities Problem.) Part (B) solves a traditional problem, considered in Karush [**26**], in Kuhn and Tucker [**30**], and in Arrow, Hurwicz, and Uzawa [**4**]. The Jacobian constraint qualification we use, (34), is weaker than the Jacobian constraint qualifications in previous work because it provides a new alternative, (b). In addition, requiring only Gâteaux differentiability is weaker than the usual Fréchet differentiability assumption.[30]

---

[27] In this case h is present and g is absent. Equivalently, we can set $g = 0$ in part (A) and delete refences to, and terms involving, λ or g.

[28] Page 17(C).

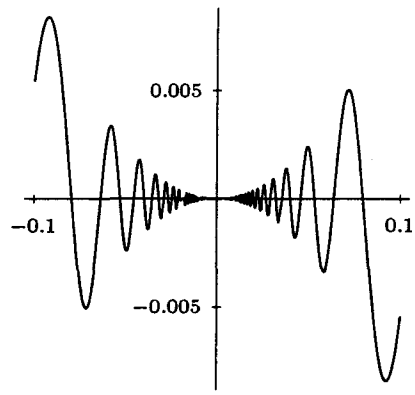
[29] The theorem holds for $\mathcal{H}_{DC}(\bar{u})$ instead of just for $\mathcal{H}_{C^1}(\bar{u})$).

[30] It holds for $(\mathcal{G}_G(\bar{u}), \mathcal{F}_G(\bar{u})$, instead of just for $(\mathcal{G}_D(\bar{u}), \mathcal{F}_D(\bar{u}))$.

(Equalities Problem.) This is the classical Lagrange Multiplier Theorem except we have weakened the usual C^1 hypothesis on h to differentiability at $\bar{u}$ and continuity in a neighborhood.[31] This allows, for example, $h = 0$ as an equality constraint, where:

$$h(u_1, u_2) = \begin{cases} u_2 - u_1^2 \sin(1/u_1), & \text{if } u_1 \neq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{45}$$

even though it is not C^1 at the origin. (See the figure.)



ii) When also $k \leqq n$, it is well known[32] that the Lagrange multiplier vector μ of (44) is unique. This can be seen by considering a linearly independent subset of k equations from the collection (44b), whose existence is guaranteed by the rank condition (35).

iii) The assumption that the constraint h is locally continuous is stronger than necessary.

First, if $k = n$, then the proof below for part (C) shows that no continuity is required. Second, if $k < n$, we can partition the vector of variables u_i into a k-vector y and an $(n-k)$-vector x, writing $u = (x, y)$, where $f_y(\bar{x}, \bar{u})$ is surjective. Then we can replace Theorem 1.C's hypothesis that h is locally continuous by the weaker assumption that, for every x near $\bar{x}$, the function $h(x, \cdot)$ is continuous;

---

(31) In particular, the theorem holds for constraints in $\mathcal{H}_{DC}(\bar{u})$ instead of just for those in $\mathcal{H}_{C^1}(\bar{u})$). Although many textbooks assume that the maximand f is also C^1, it is well known (e.g., [5]) that mere differentiability of f suffices.

(32) Cf. Bliss [7, p. 210] for $k < n$.

and the Non-C^1 Implicit Function Theorem can still be applied as in the proof of part (C) that we give below.

iv) As we have just seen, we can weaken the continuity hypothesis on h. But we cannot drop it completely, as the following "nonsubstitution" example shows.

Example. The function $h : \mathbb{R}^2 \to \mathbb{R}$:

$$h(x,y) = \begin{cases} x+y, & \text{if } x+y \neq 0 \\ x^2+y^2, & \text{if } x+y = 0 \end{cases} \tag{46}$$

is clearly differentiable at $(0,0)$, with $h'(0,0) = (1,1)$, even though it is not locally continuous — being discontinuous at all other points on the line where $x+y=0$. But h is not Lagrange regular. In particular, even though h is differentiable at the origin, its level set allows no "substitution," since the constraint set $C(h)$ contains only the origin $(0,0)$. So if we define the function $f : \mathbb{R} \to \mathbb{R}$ by $f(x,y) = y$, then f is maximized on $C(h)$ at $(0,0)$, but no real λ can satisfy:

$$\begin{aligned} \frac{\partial f}{\partial x}(0,0) &= \lambda \\ \frac{\partial f}{\partial y}(0,0) &= \lambda. \end{aligned} \tag{47}$$

Proof of Theorem 1. (A) (Mixed.) If condition (33b) holds, then it is immediate that λ and μ can be found satisfying (37) and (38). So we will assume that condition (33a) holds. Thus the index set[33] $I = \{1,\ldots,p\}$ of binding constraints is nonempty and $\text{rank}(h'(\bar{u})) = k$.

By the Reduction principle (23), it suffices to find $\lambda \in \mathbb{R}^p$ and $\mu \in \mathbb{R}^k$ satisfying (37) and (38a).

We complete the proof by contradiction. Suppose that, for some $f \in \mathcal{F}_D(\bar{u})$, there do not exist such $\lambda \in \mathbb{R}^p$ and $\mu \in \mathbb{R}^k$. Then by the Fundamental Lemma there exists a $\hat{z} \in \mathbb{R}^n$ such that:

$$g_I'(\bar{u})\hat{z} \geqq 0 \tag{48a}$$

$$h'(\bar{u})\hat{z} = 0 \tag{48b}$$

$$f'(\bar{u}) \cdot \hat{z} > 0. \tag{48c}$$

For all $s \in (0,1]$ define:

$$z(s) = \hat{z} + s\xi, \tag{49}$$

---

[33] I is defined in (16).

for any ξ satisfying (33a). With (48a) we then have:

$$g_I'(\bar{u})z(s) > 0, \tag{50}$$

and from (48c) we also have:

$$f'(\bar{u}) \cdot z(s) = f'(\bar{u}) \cdot \hat{z} + sf'(\bar{u}) \cdot \xi > 0 \qquad \text{for all small } s. \tag{51}$$

And from (33a) and (33b) we have:

$$h'(\bar{u})z(s) = 0 \qquad \text{for all } s \in (0,1]. \tag{52}$$

We define $\tilde{z} = z(s)$ for any such small s.

Since $k \geqq 1$, we can represent elements of $\mathbb{R}^n$ by $u = (x,y) \in \mathbb{R}^{n-k} \times \mathbb{R}^k = \mathbb{R}^n$. Then applying (2) to (50), (52), and (51):

$$g_u^I(\bar{u})\tilde{z} = g_x(\bar{x},\bar{y})\tilde{x} + g_y(\bar{x},\bar{y})\tilde{y} > 0 \tag{53a}$$

$$h_u(\bar{u})\tilde{z} = h_x(\bar{x},\bar{y})\tilde{x} + h_y(\bar{x},\bar{y})\tilde{y} = 0 \tag{53b}$$

$$f_u(\bar{u}) \cdot \tilde{z} = f_x(\bar{x},\bar{y})\tilde{x} + f_y(\bar{x},\bar{y})\tilde{y} > 0. \tag{53c}$$

Since $\text{rank}(h_u(\bar{u})) = k$, the Non-$C^1$ Implicit Function Theorem[34] implies there exists an open neighborhood $V \subseteqq \mathbb{R}^{n-k}$ of $\bar{x}$ and a function $y(\cdot) : V \to \mathbb{R}^k$ that is differentiable at $\bar{x}$, and such that:

$$y(\bar{x}) = \bar{y} \tag{54a}$$

$$h(x, y(x)) = 0 \qquad \text{for all } x \in V. \tag{54b}$$

$$y'(\bar{x}) = -(h_y(\bar{x},\bar{y}))^{-1} h_x(\bar{x},\bar{y}). \tag{54c}$$

By (53b) we also have:

$$\tilde{y} = -(h_y(\bar{x},\bar{y}))^{-1} h_x(\bar{x},\bar{y})\tilde{x}, \tag{55}$$

so:

$$\begin{aligned} \frac{d}{dt} y(\bar{x} + t\tilde{x})|_{t=0} &= y'(\bar{x})\tilde{x} \\ &= \tilde{y} \qquad \text{(by (54c) and (55))}. \end{aligned} \tag{56a}$$

Then the Chain Rule implies:

$$\begin{aligned} \frac{d}{dt} g_I(\bar{x} + t\tilde{x}, y(\bar{x} + t\tilde{x}))|_{t=0} &= g_{I_x}(\bar{x},\bar{y}) \cdot \tilde{x} + g_{I_y}(\bar{x},\bar{y}) \cdot \tilde{y} \\ &> 0 \qquad \text{(by (53a))} \end{aligned} \tag{56a}$$

$$\begin{aligned} \frac{d}{dt} f(\bar{x} + t\tilde{x}, y(\bar{x} + t\tilde{x}))|_{t=0} &= f_x(\bar{x},\bar{y}) \cdot \tilde{x} + f_y(\bar{x},\bar{y}) \cdot \tilde{y} \\ &> 0 \qquad \text{(by (53c))}, \end{aligned} \tag{56b}$$

---

[34] See the Appendix.

so for all small enough $t > 0$,

$$g(\bar{x} + t\tilde{x}, y(\bar{x} + t\tilde{x})) > 0 \tag{57a}$$

$$f(\bar{x} + t\tilde{x}, y(\bar{x} + t\tilde{x})) > f(\bar{x}, \bar{y}). \tag{57b}$$

By (54b) we also have $h(\bar{x}+t\tilde{x}, y(\bar{x}+t\tilde{x})) = 0$, so for small enough $t > 0$ there are points $(\bar{x}+t\tilde{x}, y(\bar{x}+t\tilde{x})) \in C(g,h)$ at which f attains a value higher than $f(\bar{x}, \bar{y})$. And that contradicts the maximality property (36c), completing the proof for the mixed case (A).

(B) (Inequalities.) This would be an immediate corollary of Theorem 1.A (with $h \equiv 0$), if we were willing to assume that g is Fréchet differentiable, rather than merely Gâteaux differentiable. We will give a proof under the weaker Gâteaux differentiability assumption by rewriting our proof of Theorem 1.A, avoiding all mention of h, and without applying the Non-C^1 Implicit Function Theorem.

If (34a) fails to hold, then the Jacobian Criterion implies that (34b) holds, and then we can take such a vector t as the desired λ. So we will assume that (34a) holds. Then the index set[35] $I = \{1, \ldots, p\}$ of binding constraints is nonempty.

By the Reduction principle, it suffices that for any $f \in \mathcal{F}_G(\bar{u})$ we can find a $\lambda \in \mathbb{R}^p$ and a $\mu \in \mathbb{R}^k$ satisfying $\lambda^T g_I'(\bar{u}) = -f'(\bar{u})$ and $\lambda \geqq 0$.[36] We complete the proof by contradiction. Suppose there does not exist such a λ. Then by the Fundamental Lemma there exists a $\tilde{z} \in \mathbb{R}^n$ such that:

$$g_I'(\bar{u})\hat{z} \geqq 0 \tag{58a}$$

$$f'(\bar{u}) \cdot \hat{z} > 0. \tag{58b}$$

For all $s \in (0, 1]$ define:

$$z(s) = \hat{z} + s\xi, \tag{59}$$

for any ξ satisfying (34a). By (58) and (34a) we have, for all small enough s:

$$\begin{aligned} &g_I'(\bar{u})z(s) > 0 \\ &f'(\bar{u}) \cdot z(s) > 0. \end{aligned} \tag{60}$$

Since these strict inequalities will not change if we slightly change the vector $z(s)$, and since g and f are Gâteaux differentiable in a dense set of directions,

---

[35] I is defined in (16).

[36] We are assured that the Jacobian $g_I'(\bar{u})$ is well defined since g is Gâteaux differentiable in directions that include the coordinate axes.

there exists a vector $\tilde{z}$ such that:

$$\begin{aligned} &g_I'(\bar{u})\tilde{z} > 0 \\ &f'(\bar{u}) \cdot \tilde{z} > 0 \end{aligned} \tag{61}$$

and such that g is Gâteaux differentiable in the direction $\tilde{z}$. We then have:

$$\begin{aligned} &\frac{d}{dt} g_I(\bar{u} + t\tilde{z})|_{t=0} = g_I'(\bar{u})\tilde{z} > 0 \\ &\frac{d}{dt} f(\bar{u} + t\tilde{z})|_{t=0} = f'(\bar{u}) \cdot \tilde{z} > 0. \end{aligned} \tag{62}$$

So for small movements from $\bar{u}$ in the direction of $\tilde{z}$, we remain in the constraint set $C(g)$ while increasing f. That contradiction of (39) completes the proof.

(C) (Equalities.) This result can be obtained as an immediate corollary of part (A) by setting $g \equiv 0$. However, because of the historical importance of the result, and the brevity allowed by our two mathematical pillars, we give a direct proof. If $\text{rank}(h'(\bar{u})) = n$, then by standard linear algebra the system

$$\mu_1 \frac{\partial h^1}{\partial u_i}(\bar{u}) + \cdots + \mu_k \frac{\partial h^k}{\partial u_i}(\bar{u}) = -\frac{\partial f}{\partial u_i}(\bar{u}) \qquad (i = 1, \ldots, n) \tag{63}$$

is solvable for μ, given any vector $f'(\bar{u})$. So, in view of (35), it remains to deal with the case that $\text{rank}(h'(\bar{u})) = k < n$.

The rest of our proof is again by contradiction. If, for some $f \in \mathcal{F}_D(\bar{u})$, there does not exist a $\mu \in I\!R^k$ satisfying (63) then by the Fundamental Lemma (using (2) to translate its result from matrix terms into derivatives) there exists a $\tilde{z} = (\tilde{x}, \tilde{y}) \in I\!R^n = I\!R^{n-k} \times I\!R^k$ such that:

$$h_u(\bar{u})\tilde{z} = h_x(\bar{x}, \bar{y})\tilde{x} + h_y(\bar{x}, \bar{y})\tilde{y} = 0 \tag{64a}$$

$$f_u(\bar{u}) \cdot \tilde{z} = f_x(\bar{x}, \bar{y}) \cdot \tilde{x} + f_y(\bar{x}, \bar{y})\tilde{y} > 0. \tag{64b}$$

Because $\text{rank}(h_u(\bar{u})) = k < n$, we can without loss of generality suppose that $\bar{u} = (\bar{x}, \bar{y}) \in I\!R^{n-k} \times I\!R^k = I\!R^n$ and

$$\text{rank } (h_y(\bar{x}, \bar{y})) = k. \tag{65}$$

Then by the Non-C^1 Implicit Function Theorem,[37] there exists an open neighborhood $V \subseteq I\!R^{n-k}$ of $\bar{x}$ and a function $y(\cdot) : V \to I\!R^k$ that is differentiable at $\bar{x}$, and such that:

$$y(\bar{x}) = \bar{y} \tag{66a}$$

$$h(x, y(x)) = 0 \qquad \text{for all } x \in V. \tag{66b}$$

$$y_x(\bar{x}) = -(h_y(\bar{x}, \bar{y}))^{-1} h_x(\bar{x}, \bar{y}). \tag{66c}$$

---

[37] See the Appendix.

By (64a) we also have:

$$\tilde{y} = -(h_y(\bar{x}, \bar{y}))^{-1} h_x(\bar{x}, \bar{y})\tilde{x}, \tag{67a}$$

hence:

$$\tilde{y} = y_x(\bar{x})\tilde{x}. \tag{67b}$$

Therefore:

$$\begin{aligned} \frac{d}{dt} y(\bar{x} + t\tilde{x})|_{t=0} &= y_x(\bar{x})\tilde{x} \\ &= \tilde{y}. \end{aligned} \tag{68}$$

Then the Chain Rule implies:

$$\begin{aligned} \frac{d}{dt} f(\bar{x} + t\tilde{x}, y(\bar{x} + t\tilde{x}))|_{t=0} &= f_x(\bar{x}, \bar{y}) \cdot \tilde{x} + f_y(\bar{x}, \bar{y}) \cdot \tilde{y} \\ &> 0 \qquad \text{(by (64b))}, \end{aligned} \tag{69}$$

so for all small enough $t > 0$:

$$f(\bar{x} + t\tilde{x}, y(\bar{x} + t\tilde{x})) > f(\bar{x}, \bar{y}). \tag{70}$$

Since $h(\bar{x} + t\tilde{x}, y(\bar{x} + t\tilde{x})) = 0$ by (66b), this contradicts the maximality property (43b). ∎

As a special case of Theorem 1.A, we have:

Corollary 1a. (Equalities and inequalities.) Let U be an open subset of $\mathbb{R}^n$, and let $\bar{u} \in U$. Suppose $g \in \mathcal{G}_D(\bar{u})$ and $h \in \mathcal{H}_{DC}(\bar{u})$, i.e. $g : U \to \mathbb{R}^m$ and $h : U \to \mathbb{R}^k$ are differentiable at $\bar{u}$, and h is locally continuous at $\bar{u}$. Let $g^1, \ldots, g^p$ be the binding inequality constraints at $\bar{u}$. If:

$$\operatorname{rank}\left(\begin{bmatrix} g_I'(\bar{u}) \\ h'(\bar{u}) \end{bmatrix}\right) = p + k, \tag{71}$$

then (g, h) is Lagrange mixed-regular for $(\bar{u}, \mathcal{F}_D(\bar{u}))$.

Remark (i). If $m = k \leqq n$ and $g^j(\bar{u}) = 0$ for $j = 1, \ldots, m$, then the Lagrange multiplier vector (λ, μ) whose existence is guaranteed by the corollary is unique. For in this case $\bar{u}$ a fortiori maximizes $f(u)$ subject to both $g(u) = 0$ and $h(u) = 0$, so remark (ii) on page 22 applies.[38]

---

[38] This also follows from [**37**, Theorem 3.1, p. 180], provided all constraints considered there vanish at $\bar{u}$.

Remark (ii). For (71) to hold it is not sufficient that $\text{rank}(g_I'(\bar{u})) = p$ and $\text{rank}(h'(\bar{u})) = k$. Although the (71) implies these two rank conditions, the following example shows that the converse is not true. Let:

$$\begin{aligned} h(x_1, x_2) &= -x_1 + x_2^2 \\ g(x_1, x_2) &= -x_1 + \frac{1}{2}x_2^2. \end{aligned} \tag{72}$$

Then

$$f(x_1, x_2) = x_1 + x_2 \tag{73}$$

is maximized at $(0,0)$ subject to the constraints $g \geqq 0$ and $h = 0$. And $\text{rank}(g'(\bar{u})) = 1$ and $\text{rank}h'(\bar{u})) = 1$, but the rank of the combined matrix is not equal to $1 + 1$.

Proof of Corollary 1a. Under the rank condition (71), it follows that $\text{rank}(h'(\bar{u})) = k$ and also that there exists a solution ξ of

$$\begin{bmatrix} g_I'(\bar{u})\xi \\ h'(\bar{u})\xi \end{bmatrix} = \begin{bmatrix} c \\ 0 \end{bmatrix}, \tag{74}$$

where c is any positive column vector in $\mathbb{R}^p$. Thus condition (33a) holds, so Theorem 1.A guarantees Lagrange regularity. ∎

As a special case of Theorem 1.B, we have:

Corollary 1b. (Inequalities.) Let U be an open subset of $\mathbb{R}^n$, and let $\bar{u} \in U$. Suppose $g \in \mathcal{G}_G(\bar{u})$, i.e., $g : U \to \mathbb{R}^m$ is Gâteaux differentiable at $\bar{u}$; and suppose $g^1, \ldots, g^p$ are the binding constraints at $\bar{u}$. If the *rank condition*

$$\text{rank}(g_I'(\bar{u})) = p \tag{75}$$

holds, then g is Lagrange inequality-regular for $(\bar{u}, \mathcal{F}_G(\bar{u}))$.

Remark (i). If $m \leqq n$ and $g^j(\bar{u}) = 0$ for $j = 1, \ldots, m$, then the Lagrange multiplier vector λ whose existence is guaranteed by Corollary 1b is unique. For then $\bar{u}$ a fortiori maximizes $f(u)$ subject to $g(u) = 0$, and hence remark (ii) on page 22 applies.[39]

---

[39] Again, this also follows from [37, Theorem 3.1, p. 180], provided all constraints considered there vanish at the maximizing point.

Remark (ii). The Jacobian $g'(\bar{u})$ may have full rank p even though the Jacobian $g'_I(\bar{u})$ of the binding constraints may have rank $< p$. So it is important to note that the rank condition (75) refers to the Jacobian of only the ***binding*** constraints Consider, for example:[40]

$$\begin{aligned} g^1(x_1, x_2) &= (1 - x_2)^3 - x_2 \\ g^2(x_1, x_2) &= x_1 \\ g^3(x_1, x_2) &= x_2. \end{aligned} \tag{76}$$

Then for $\bar{x} = (1, 0)$, the constraints g^1 and g^3 are binding, while g^2 is not. And:

$$g'(\bar{x}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{77}$$

which only has rank $2 < 3 = p$. So condition (75) fails, and Corollary 1b does not apply, even though $\text{rank}(g'(\bar{u})) = p$.

Remark (iii). Condition (75) is stronger than necessary for Lagrange regularity. Consider:

$$\begin{aligned} g^1(x_1, x_2) &= 6 - 3x_1 - 3x_2 \\ g^2(x_1, x_2) &= 6 - 4x_1 - 2x_2 \\ g^3(x_1, x_2) &= 6 - 2x_1 - 4x_2, \end{aligned} \tag{78}$$

for which all constraints are binding at the point $(2, 2)$, and:

$$g'(2, 2) = \begin{bmatrix} -3 & -3 \\ -4 & -2 \\ -2 & -4 \end{bmatrix}. \tag{79}$$

Clearly $\text{rank}(g'_I(2, 2)) = 2 < 3 = p$, so condition (75) does not hold. Yet $g'_I(2, 2)\xi > 0$ for $\xi = (1, 1)$, so g is Lagrange regular at $(2, 2)$ by Theorem 1.B.

Proof of Corollary 1b. Under the rank condition, there exists a solution ξ of $g'_I(\bar{u})\xi = e$, where $e = (1, \ldots, 1) > 0 \in \mathbb{R}^p$. Thus condition (34$a$) holds, so Theorem 1.B guarantees Lagrange regularity. ∎

---

[40] Cf. [**30**, p. 484].

(Equalities.) The rank condition for equalities parallel to those in (71) and (75) is the condition already stated in (35) of Theorem 1.C.

VI The Jacobian Criterion Is Minimal

In Section V we showed that the Jacobian Criterion is sufficient for Lagrange regularity. However, as shown by the Example below in the introduction to Section VII, p. 34, it is not necessary. Indeed, the Example shows that no Jacobian constraint qualification can be both necessary and sufficient for Lagrange regularity.

We will obtain a necessary and sufficient constraint qualification in Section VII (Theorems 3 and 4) below, but it will be a "path condition" rather than a Jacobian condition. Nevertheless, because the Jacobian conditions typically have computability properties that make them useful in practical applications, while path conditions are often more difficult to deal with, Jacobian conditions are of special interest. So it is important to show that the Jacobian Criterion of the previous sections is as weak as possible among Jacobian conditions.

We now show that, if one restricts oneself to Jacobian conditions that are sufficient for Lagrange regularity, then the Jacobian Criterion is minimal (over an appropriate class of constraint functions). Theorem 2.A, together with Theorem 1.A, will show that the mixed-problem Jacobian Criterion is a minimal Jacobian constraint qualification for the mixed problem over the class of differentiable functions g and h, where h is locally continuous.

And Theorem 2.B, together with Theorem 1.B, will show that the inequalities-problem Jacobian Criterion is a minimal Jacobian constraint qualification for the inequalities problem over the class of differentiable functions g.

Finally, Theorem 2.C, together with Theorem 1.C, will show that the equalities-problem Jacobian Criterion is a minimal Jacobian constraint qualification for the equalities problem over the class of differentiable functions h that are locally continuous.

Theorem 2.A. (Equalities and inequalities.) Let U be an open subset of $\mathbb{R}^n$ and let $\bar{u} \in U$. The Jacobian Criterion (A) is minimally $(\bar{u}, \mathcal{G}_D(\bar{u}), \mathcal{H}_{DC}(\bar{u}), \mathcal{F}_D(\bar{u}))$-sufficient for Lagrange mixed-regularity.[41]

---

[41] See pp. 15–16 for the definition of minimal sufficiency, and p. 10 for the definition of Lagrange mixed-regularity.

Remark. The theorem does not claim that the Mixed-Problem Jacobian Criterion is necessary for Lagrange mixed-regularity of a given *function pair* (g, h). It only claims the Criterion is minimal *in the class of Jacobian constraint qualifications*.[42] However, outside the class of Jacobian conditions, there are *path-type* conditions guaranteeing Lagrange regularity even in cases where the Jacobian Criterion is not satisfied. See, for example, the introduction to Section VII.

Proof of Theorem 2.A. By Theorem 1.A, the Jacobian Criterion (A)[43] is sufficient for Lagrange mixed-regularity. To see that the Jacobian Criterion (A) is minimal,[44] suppose there is some $g \in \mathcal{G}_D(\bar{u})$ and some $h \in \mathcal{H}_{DC}(\bar{u})$ that do not satisfy the Jacobian Criterion (A). We will show there is some $\hat{g} \in \mathcal{G}_D(\bar{u})$ and $\hat{h} \in \mathcal{H}_{DC}(\bar{u})$ with $\hat{g}'(\bar{u}) = g'(\bar{u})$ and $\hat{h}'(\bar{u}) = h'(\bar{u})$ for which $(\hat{g}, \hat{h})$ is not Lagrange regular for $(\bar{u}, \mathcal{F}_D(\bar{u}))$.

Without loss of generality we can assume that U is a neighborhood of the origin $0 \in \mathbb{R}^n$, and that $\bar{u} = 0$. Define the matrix $A = g'(0)$ and the matrix $B = h'(0)$. Since we are supposing that (g, h) violates the Jacobian Criterion (A), we know that (A, B) cannot satisfy either of these two mutually exclusive conditions:

$$\text{a) } \operatorname{rank}(B) = \min\{k, n\} \text{ and there exists a } \xi \in \mathbb{R}^n \text{ such that:} \qquad (80a)$$

$$A\xi > 0$$

$$B\xi = 0.$$

$$\text{b) } \operatorname{wedge}(\{a(1), \ldots, a(m)\}) + \operatorname{span}(\{b(1), \ldots, b(k)\}) = \mathbb{R}^n \qquad (80b)$$

i.e.:

$$\{t_1 a(1) + \cdots + t_m a(m) + z_1 b(1) + \cdots + z_k b(k) : \text{real } t_1, \ldots, t_p \geqq 0 \ \& \ z_1, \ldots, z_k \in \mathbb{R}\} = \mathbb{R}^n.$$

It will suffice to find $\hat{g}$ and $\hat{h}$ as above and such that $C(\hat{g}, \hat{h}) = \{0\}$. For then the failure of (b) implies that there is some $\gamma \in \mathbb{R}^n$ with:

$$\gamma \notin \operatorname{wedge}(\{a(1), \ldots, a(m)\}) + \operatorname{span}(\{b(1), \ldots, b(k)\}); \qquad (81)$$

In fact, as seen from the proof below, Theorem 1.A remains true even if the class of maximands is as narrow as the class of linear ones — i.e., if we replace $\mathcal{F}_D(\bar{u})$ by the class of linear functions.

(42) Cf. p. 15.

(43) Page 17.

(44) Page 15.

so if we define $f(x) = -\gamma \cdot x$ then f is maximized on $C(\hat{g}, \hat{h})$ at the origin 0, and $f'(0) = -\gamma$, which by (81) cannot satisfy the Lagrange regularity requirements (18) and (20).

We examine first the special case in which some row $a(i) = 0 \in \mathbb{R}^n$. For $x \in \mathbb{R}^n$, we define $\hat{g}^i(x) = -(x_1^2 + \cdots + x_n^2)$, so $a(i) = \hat{g}^{i\prime}(0)$, and $\hat{g}^i(x) \geqq 0$ if and only if $x = 0$; we define $\hat{g}^j = g^j$ for $j \neq i$. Thus $\hat{g} \in \mathcal{G}_D(\bar{u})$, $\hat{g}'(0) = A$, and $C(\hat{g}, h) = \{0\}$.

Analogously, if some row $b(j) = 0 \in \mathbb{R}^n$, then defining $\hat{h}^j(x_1, \ldots, x_n) = -(x_1^2 + \cdots + x_n^2)$ and $\hat{h}^i = h$ for $i \neq j$, a similar argument shows that $\hat{h} \in \mathcal{H}_{DC}(\bar{u})$, $\hat{h}'(0) = B$, and $C(g, \hat{h}) = \{0\}$.

It remains to consider the case that:

$$\begin{aligned} a(i) &\neq 0 \qquad \text{for all } i = 1, \ldots, m \\ b(j) &\neq 0 \qquad \text{for all } j = 1, \ldots, k. \end{aligned} \tag{82}$$

We know that (80a) fails. There are two ways it can fail: i) we might have $\text{rank}(B) < \min\{k, n\}$, or ii) there might exist no ξ with $A\xi > 0$ and $B\xi = 0$.

(i) First consider the case that $\text{rank}(B) < \min\{k, n\}$. Since $\text{rank}(B) \neq k$, then some row of B is a linear combination of the others. Without loss of generality, suppose that:

$$b(1) = z_2 b(2) + \cdots + z_k b(k) \tag{83}$$

for some real $z_2, \ldots, z_k$. Since $b(1) \neq 0$, we can also without loss of generality choose a basis so that

$$b(1) = (1, 0, \ldots, 0). \tag{84}$$

Now define:

$$\begin{aligned} \hat{g}^i(x_1, \ldots, x_n) &= a(i) \cdot x \qquad (i = 1, \ldots, m) \\ \hat{h}^1(x_1, \ldots, x_n) &= x_1 - (x_2^2 + \cdots + x_n^2) \\ \hat{h}^i(x_1, \ldots, x_n) &= b(i) \cdot x \qquad (i = 2, \ldots, k). \end{aligned} \tag{85}$$

Then $\hat{g}'(0) = A$ and $\hat{h}'(0) = B$, $\hat{g} \in \mathcal{G}_D(\bar{u})$ and $\hat{h} \in \mathcal{H}_{DC}(\bar{u})$; and if $\hat{g}(x) \geqq 0$ & $\hat{h}(x) = 0$ it follows from (83), (84), and (85) that $x_1 = \cdots = x_n = 0$, so again the constraint set contains just the origin: $C(\hat{g}, \hat{h}) = \{0\}$.

(ii) The other way that (80a) could fail is through absence of a $\xi \in \mathbb{R}^n$ with $A\xi > 0$ and $B\xi = 0$. In that case the Theorem of the Alternative,[45] implies

[45] See the Appendix, applying the special case of part (II) when C and γ are absent.

existence of a $t \in \mathbb{R}^m$ and $v \in \mathbb{R}^k$ such that:

$$\begin{aligned} &t \geq 0 \\ &t^T A + v^T B = 0 \in \mathbb{R}^n. \end{aligned} \tag{86a}$$

Without loss of generality we can rewrite this as:

$$a(1) = -(w_2 a(2) + \cdots + w_m a(m) + z_1 b(1) + \cdots + z_k b(k)), \tag{86b}$$

for some real z_i and some real $w_j \geqq 0$. Because of (82) we can, also without loss of generality, choose a basis so that:

$$a(1) = (1, 0, \ldots, 0) \in \mathbb{R}^n. \tag{87}$$

Now define:

$$\begin{aligned} &\hat{g}^1(x_1, \ldots, x_n) = x_1 - (x_2^2 + \cdots + x_n^2) \\ &\hat{g}^i(x_1, \ldots, x_n) = a(i) \cdot x \qquad (i = 2, \ldots, m) \\ &\hat{h}^i(x_1, \ldots, x_n) = b(i) \cdot x \qquad (i = 1, \ldots, k). \end{aligned} \tag{88}$$

Then $\hat{g}'(0) = A$ and $\hat{h}'(0) = B$, $\hat{g} \in \mathcal{G}_D(\bar{u})$ and $\hat{h} \in \mathcal{H}_{DC}(\bar{u})$, and if $\hat{g}(x) \geqq 0$ & $\hat{h}(x) = 0$ it follows from (86), (87), and (88) that $x_1 = \cdots = x_n = 0$; so again the constraint set contains just the origin: $C(\hat{g}, \hat{h}) = \{0\}$. ∎

Theorem 2.B. (Inequalities.) Let U be an open subset of $\mathbb{R}^n$ and let $\bar{u} \in U$. The Jacobian Criterion (B) is minimally $(\bar{u}, \mathcal{G}_G(\bar{u}), \mathcal{F}_G(\bar{u}))$-sufficient for Lagrange inequality-regularity.[46]

Proof. This follows from obvious modifications of our proof for Theorem 2.A, or as a corollary of that theorem if we set $h = 0$. ∎

Finally, the Jacobian Criterion (C), the classical rank condition (35), is a minimally sufficient Jacobian constraint qualification for the classical Lagrange equality-constrained problem:

---

[46] See pp. 15–16 for the definition of minimal sufficiency, and p. 10 for the definition of Lagrange inequality-regularity.

In fact, as seen from the proof below, Theorem 2.B remains true even if the class of maximands is as narrow as the class of linear (hence infinitely differentiable) ones — i.e., if we replace $\mathcal{F}_G(\bar{u})$ by the class of linear functions.

Theorem 2.C. (Equalities.) Let U be an open subset of $\mathbb{R}^n$ and let $\bar{u} \in U$. The Jacobian Criterion (C),

$$\operatorname{rank}(h'(\bar{u})) = \min\{k, n\}, \tag{89}$$

is minimally $(\bar{u}, \mathcal{H}_D(\bar{u}), \mathcal{F}_D(\bar{u}))$-sufficient for Lagrange equality-regularity.[47]

Proof. This follows from obvious modifications of our proof for Theorem 2.A, or as a corollary of that theorem if we set $g = 0$. ∎

VII The Tangency-Path Criterion

Why consider more than the Jacobian? Because the Lagrange multiplier property (18) is a local property, one might expect that Jacobian properties would suffice to characterize conditions under which Lagrange regularity holds for (g, h). However, they cannot determine Lagrange regularity in all cases, as we see from the following example.

Example. Consider the inequalities $g \geqq 0$ where:

$$\begin{aligned} g^1(u_1, u_2) &= \begin{cases} u_2 - u_1^2 \sin(1/u_1), & \text{if } u_1 \neq 0 \\ 0 & \text{otherwise} \end{cases} \\ g^2(u_1, u_2) &= -u_2, \end{aligned} \tag{90}$$

or the equalities $h = 0$ where:

$$\begin{aligned} h^1(u_1, u_2) &= \begin{cases} u_2 - u_1^2 \sin(1/u_1), & \text{if } u_1 \neq 0 \\ 0 & \text{otherwise} \end{cases} \\ h^2(u_1, u_2) &= -u_2. \end{aligned} \tag{91}$$

It will be evident from Theorem 3 (or Theorem 5) that, for any function f that is maximized at $\bar{u} = (0, 0)$ subject to the constraints (90) or (91) there do exist Lagrange multipliers satisfying (18) and (20). By contrast, the function g of (14)

---

[47] See pp. 15–16 for the definition of minimal sufficiency, and p. 11 for the definition of Lagrange equality-regularity.

Again, as seen from the proof below, Theorem 2.C remains true even if the class of maximands is as narrow as the class of linear (hence infinitely differentiable) ones — i.e., if we replace $\mathcal{F}_D(\bar{u})$ by the class of linear functions.

is not Lagrange inequality-regular, even though it has the same Jacobian $g'(0,0)$ as (90) above; and the function h of (15) is not Lagrange-equality regular, even though it has the same Jacobian $h'(0,0)$ as (91).[48]

So Jacobian conditions alone cannot characterize Lagrange regularity of (g,h) — none can be both necessary and sufficient. Even though the Jacobian Criterion was minimal among Jacobian constraint qualifications, a finer tool than Jacobian conditions is needed for a complete characterization. So we turn to path conditions.

The tangent cone. To describe derivatives of paths, we use the tangent cone. For any subset S of $U \subseteq \mathbb{R}^n$, we say that a vector $v \in \mathbb{R}^n$ is *tangent to S at $\bar{u}$* if:

either:

a) there exists a sequence of points $u^i \in S$ such that: (92a)

$$\text{i) } u^i \neq \bar{u} \qquad \text{for all } i = 1,2,3,\ldots$$

$$\text{ii) } \|u^i - \bar{u}\| \xrightarrow[i\to\infty]{} 0$$

$$\text{iii) } \frac{u^i - \bar{u}}{\|u^i - \bar{u}\|} \xrightarrow[i\to\infty]{} \frac{v}{\|v\|};$$

or:

b) $\bar{u}$ is an isolated point of S and $v = 0$. (92b)

We define:

$$T_{\bar{u}}S = \{rv \in U : v \text{ is tangent to } S \text{ at } \bar{u} \;\&\; r \text{ is a nonnegative real number}\}. \tag{93}$$

As this is clearly a cone, $T_{\bar{u}}S$ is called the *tangent cone of S at $\bar{u}$*.

A path interpretation of the tangent cone. For an alternative view of the tangent cone concept, we now give an equivalent formulation in terms of derivatives of certain paths. This provides useful intuition. Also, when we later introduce a new constraint qualification based on the tangent cone, the path interpretation will help us link the new qualification to the traditional constraint qualifications of Karush [26], Kuhn and Tucker [30], and Arrow, Hurwicz and Uzawa [4], which are based on derivatives of paths.

---

[48] Cf. the discussion of (14) and (15), p. 9.

We characterize the tangent cone as a set of derivatives of paths as follows. First we say that a function $\phi : [0,1] \to I\!R^n$ is a *path at $\bar{u}$ attending S* if $\phi(0) = \bar{u}$ and $\phi(t)$ lies in S for arbitrarily small positive t:

$$\forall \varepsilon_{\ \varepsilon>0}\ \exists t_{\ 0<t<\varepsilon}\ \phi(t) \in S. \tag{94}$$

Then we can view the $T_{\bar{u}}S$ as the derivatives of paths at $\bar{u}$ attending S:

Proposition on Paths and Tangent Cones. a) If a path $\phi : [0,1] \to I\!R^n$ at $\bar{u}$ attends S and is differentiable at 0, then $\phi'(0) \in T_{\bar{u}}S$.

b) Conversely, if $v \in T_{\bar{u}}S$, then there is a (not necessarily continuous) path $\phi : [0,1] \to I\!R^n$ at $\bar{u}$ attending S with $v = \phi'(0)$.

Proof. (a) We assume the conditions of (a). If ϕ is constant in a neighborhood of 0, then $\phi(t) = \bar{u}$ for all t, so $\phi'(0) = 0$, which belongs to $T_{\bar{u}}S$. If ϕ is not constant near 0, then there exist $t \to 0$ with $\phi(t) \neq \phi(0)$, and then we have a sequence of $t > 0$ with $t \searrow 0$ and $\phi(t) \in S$ with:

$$\begin{aligned} \frac{\phi(t) - \phi(0)}{\|\phi(t) - \phi(0)\|} &= \frac{\phi'(0)t + o(t)}{\|\phi'(0)t + o(t)\|} \\ &= \frac{\phi'(0) + \dfrac{o(t)}{t}}{\|\phi'(0) + \dfrac{o(t)}{t}\|} \xrightarrow[t \searrow 0]{} \frac{\phi'(0)}{\|\phi'(0)\|} \in T_{\bar{u}}S, \end{aligned} \tag{95}$$

so $\phi'(0) \in T_{\bar{u}}S$.

(b) Suppose $v \in T_{\bar{u}}S$. If $v = 0$, then $\phi \equiv 0$ lies in S and satisfies $v = \phi'(0)$. If $v \neq 0$,then there exist $u^i \in S$ satisfying (92). Then on this path ϕ:

$$\phi(t) = \begin{cases} \bar{u} + t\dfrac{u^i - \bar{u}}{\|u^i - \bar{u}\|}, & \text{if } \|u^{i+1} - \bar{u}\| < t \leqq \|u^i - \bar{u}\| \\ \bar{u}, & \text{if } t = 0 \end{cases} \tag{96}$$

we have $\phi(t^i) = u^i \in S$ for $t_i = \|u^i - \bar{u}\|$ so ϕ attends S. And $\phi'(0) = v$. ∎

We cannot strengthen the statement in part (b) of the Proposition that ϕ attends S to the requirement that the values $\phi(t)$ all lie in S. Simple examples rule that out.

Also, we cannot guarantee continuity of ϕ at 0 in part (b) of the Proposition. This is apparent from the example for (91) and the figure. There the tangent cone to the constraint set at $\bar{u} = (0,0)$ is just the horizontal axis. If we start away from the origin, then the only way to approach the origin from within the constraint set is to hop discontinuously from intersection to intersection, converging to the origin, on the horizontal axis, with a horizontal tangent vector.[49]

Conversion of equalities to inequalities. Our mixed maximization problem, as stated at the beginning of Section III, is concerned with maximizing a function f subject to both inequality constraints $g^i \geqq 0$ and equality constraints $h^i = 0$. It is sometimes convenient to convert each equality $h^i = 0$ into two inequalities $\tilde{g}^{i_1} \geqq 0$ and $\tilde{g}^{i_2} \geqq 0$, where

$$\begin{aligned} \tilde{g}^{i_1} &= h^i \\ \tilde{g}^{i_2} &= -h^i. \end{aligned} \tag{97}$$

We write the new inequalities $\tilde{g} \geqq 0$ together with the original inequalities $g \geqq 0$ by $\tilde{g}(h)$, which we call the *associated inequalities* for (g,h).[50] Clearly the original constrained maximization problem is now equivalent to maximizing a function f on U subject to $\tilde{g}(h) \geqq 0$.

We can now state a path criterion for Lagrange regularity in terms of the tangent cone concept.

Tangency-Path Criterion. To characterize Lagrange regularity, we state the following condition in terms of tangent cones; a path interpretation will be given later.

Suppose U is an open subset of $\mathbb{R}^n$, and let $(g,h) : U \to \mathbb{R}^m \times \mathbb{R}^k$. We say that (g,h) satisfies the *Tangency-Path Criterion at* $\bar{u}$ if:

---

[49] Therefore only requiring paths to have tangents at the origin allows more generality than in Kuhn and Tucker, where the definitions assumed differentiable paths. This answers affirmatively the question raised by Arrow and Hurwicz [4], who noted that their differentiability hypothesis was weaker than that in Kuhn and Tucker [30], and who wondered if this in fact provided extra generality.

[50] In particular, when h is absent, then $\tilde{g}(h)$ is understood to be g, and when g is absent, then $\tilde{g}(h)$ is understood to contain the inequalities specified in (97).

i) all the $\tilde{g}(h)^i$ have partial derivatives at $\bar{u}$,[51]
ii) for any $v \in I\!R^n$,[52]

$$\tilde{g}(h)_I{}'(\bar{u})v \geqq 0 \quad \Rightarrow \quad v \in \mathrm{cl}(\mathrm{ch}(T_{\bar{u}}C(\tilde{g}(h)))). \tag{98a}$$

We can also write (ii) as:

$$L(\tilde{g}(h)) \subseteqq V(\tilde{g}(h)), \tag{98b}$$

where we define define:

$$L(\tilde{g}(h)) = \{v \in I\!R^n : \tilde{g}(h)_I{}'(\bar{u})v \geqq 0\}, \tag{99}$$

and:

$$V(\tilde{g}(h)) = \mathrm{cl}(\mathrm{ch}(T_{\bar{u}}C(\tilde{g}(h)))). \tag{100}$$

A path interpretation of the Criterion. In view of the Proposition on Paths and Tangent Cones, property (ii) in (98a) can also be stated in terms of paths. Paraphrased in terms of binding constraints g^i, (98a.ii) says that if a vector v has a nonnegative inner product $g^{i\prime}(\bar{u}) \cdot v$ with each $g^{i\prime}(\bar{u})$, then v is a limit of positive linear combinations of tangents of paths attending[53] the constraint set. In particular, if the g^i were differentiable[54] then any direction v in which all the g^i had positive derivatives would be, if not a direction (i.e., tangent) of a "feasible" path, then at least a positive linear combination of directions of attending paths.

Remark (i). One might consider the simpler condition obtained from replacing $\mathrm{cl}(\mathrm{ch}(T_{\bar{u}}C(\tilde{g}(h))))$ by its subset $\mathrm{ch}(T_{\bar{u}}C(\tilde{g}(h)))$:[55]

$$L(\tilde{g}(h)) \subseteqq \mathrm{ch}(T_{\bar{u}}C(\tilde{g}(h))). \tag{101}$$

However, that would impose a stronger requirement on constraints than does (98b). Consider, for example, the inequality constraints $g^i : I\!R^3 \to I\!R$ defined by:

$$\begin{aligned} g^1(x,y,z) &= z^2 - y^2 - (x - |z|)^2 \\ g^2(x,y,z) &= x. \end{aligned} \tag{102}$$

---

[51] I.e., all g^i and all h^j have partial derivatives at $\bar{u}$.

[52] I is defined in (16).

[53] See the definition page 36.

[54] And not merely possessing partial derivatives.

[55] Cf. [15].

Both g^i are binding at the origin $(0,0,0)$, and have partial derivatives (in fact they are differentiable) at the origin:

$$g'(0,0,0) = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}. \tag{103}$$

The constraint set $C(g)$ is a closed cone,[56] so it is easy to see that $T_{(0,0,0)}C(g) = C(g)$. But $\mathrm{ch}(C(g)) \subsetneqq \mathrm{cl}(\mathrm{ch}(C(g)))$, since the former set is the open half-space defined by $x > 0$, together with the line where $x = y = 0$, while the latter is the closed half-space defined by $x \geqq 0$.[57] Since (103) implies that $L(g)$ is the latter closed half-space, the Tangency-Path condition is satisfied under our definition (98*b*), but not if we define it using the stronger requirement (101) — i.e., using $\mathrm{ch}(T_{(0,0,0)}C(g))$ rather than $\mathrm{cl}(\mathrm{ch}(T_{(0,0,0)}C(g)))$.[58]

Remark (ii). If we replace g^1 by

$$\hat{g}^1(x,y,z) = \begin{cases} g^1(x,y,z)/\sqrt{(x^2+y^2+z^2)}, & \text{if } (x,y,z) \neq (0,0,0) \\ 0, & \text{otherwise,} \end{cases} \tag{104}$$

then $(\hat{g}^1, g^2)$ has the same partial derivatives at the origin $(0,0,0)$ as in (103), but it is no longer Fréchet differentiable at the origin, even though it defines the same constraint set as (g^1, g^2). This illustrates the fact that Theorem 3 guarantees Lagrange regularity even when full differentiability does not apply. The example still exhibits a discrepancy between $\mathrm{cl}(\mathrm{ch}(T_0C(g))$ and $\mathrm{ch}(T_0C(g))$.

Remark (iii). Without g^2, the system (102) would not be Lagrange regular, even though the constraint set would be the same. This emphasizes that Lagrange regularity is a property of constraint functions, not just of constraint sets.[59]

---

[56] Cf. [**16**, p. 9].

[57] See [**16**], *loc. cit.*

[58] See also (163), p. 59.

[59] Cf. the comments on Removable Constraints above, p. 18, for Jacobian conditions. Also see [**2**, footnote 2, p. 2], where again the form of constraints influences whether the rank constraint qualification holds or not — indeed, whether the inequality constraint is Lagrange regular or not.

We will show that the Tangency-Path Criterion is necessary and sufficient for Lagrange regularity. By the conversion rule (15), it clearly suffices to show this for inequality constraints $g^i \geqq 0$. By the definition of the Tangency-Path Criterion, the functions g^i in the following theorem have partial derivatives; but it is not assumed that g is Fréchet, or even Gâteaux differentiable.[60]

Theorem 3. (Inequalities.) Let U be an open subset of $\mathbb{R}^n$, and let $\bar{u} \in U$. Suppose $g : U \to \mathbb{R}^m$. If the Tangency-Path Criterion[61] holds at $\bar{u}$, then g is Lagrange inequality-regular for $(\bar{u}, \mathcal{F}_D(\bar{u}))$. In other words, if $\bar{u}$ maximizes $f : U \to \mathbb{R}$ subject to $g \geqq 0$, i.e., if:

$$g(\bar{u}) \geqq 0 \tag{105a}$$

$$\forall u_{\,u \in U \;\&\; g(u) \geqq 0}\; f(\bar{u}) \geqq f(u), \tag{105b}$$

and if f is differentiable at $\bar{u}$, then there exists a $\lambda \in \mathbb{R}^m$ such that:

$$f'(\bar{u}) + \lambda^T g'(\bar{u}) = 0, \tag{106a}$$

i.e.:

$$\frac{\partial f}{\partial u_i}(\bar{u}) + \lambda_1 \frac{\partial g^1}{\partial u_i}(\bar{u}) + \ldots + \lambda_m \frac{\partial g^m}{\partial u_i}(\bar{u}) = 0 \qquad (i = 1, \ldots, n), \tag{106b}$$

and such that:

$$\lambda \geqq 0 \tag{107a}$$

$$\lambda_j = 0 \;\text{ if } j \in J. \tag{107b}$$

Proof. For each j with $j \in J$, define:

$$\lambda_j = 0. \tag{108}$$

Then (107b) is satisfied. If the index set[62] I of binding constraints is empty, we are done, since then $\bar{u}$ is in the interior of the constraint set, hence the usual

---

[60] Gould and Tolle [18, p. 167], presupposing differentiability and continuity conditions that are not made in Theorems 3, 4, 5, or 6 below, state a criterion that is the dual of a condition stronger than the Tangency-Path Criterion. (See Section X below.)

Our Theorem 3 is a strengthing of Arrow, Hurwicz, and Uzawa's Theorem 1 [4], as well as to the "if" part of Gould and Tolle's theorem. Our Theorem 4 is a strengthening of Theorem 2 in Arrow, Hurwicz, and Uzawa [4]; it plays the same role, under weaker hypotheses, as the "only if" part of Gould and Tolle's theorem.

[61] Page 37.

[62] I is defined in (16).

calculus arguments imply

$$\frac{\partial f}{\partial u_j}(\bar{u}) = 0 \qquad (j = 1, \ldots, n), \tag{109}$$

which together with (108) satisfies (106). So we assume that $I = \{1, \ldots, p\}$ is nonempty.

By the Reduction principle,[63] it suffices to find a $\lambda \in I\!R^p$ satisfying $\lambda^T g_I'(\bar{u}) = -f'(\bar{u})$ and $\lambda \geqq 0$. Arguing by contradiction, suppose there does not exist such a λ. Then the Fundamental Lemma implies there exists a $\tilde{z} \in I\!R^n$ such that:

$$g_I'(\bar{u})\tilde{z} \geqq 0 \tag{110a}$$

$$f'(\bar{u}) \cdot \tilde{z} > 0. \tag{110b}$$

By the Tangency-Path Criterion,[64] property (110a) implies the existence of a sequence of $w^i \in \mathrm{ch}(T_{\bar{u}}C(g))$ with:

$$w^i \xrightarrow[i \to \infty]{} \tilde{z}. \tag{111}$$

For all large enough i, therefore, (110b) implies:

$$f'(\bar{u}) \cdot w^i > 0. \tag{112}$$

By (7a), for each i there exist nonnegative $t_{i,1}, \ldots, t_{i,q}$ summing to 1, and there exist of $w^{i,1}, \ldots, w^{i,q} \in T_{\bar{u}}C(g)$ with:[65]

$$w^i = t_{i,1}w^{i,1} + \cdots + t_{i,q}w^{i,q}. \tag{113}$$

It follows from (112) and (113) that:

$$f'(\bar{u}) \cdot w^{i,j} > 0 \tag{114}$$

for some large i and some $j = 1, \ldots, n$. Since $w^{i,j} \in T_{\bar{u}}C(g)$, and since $w^{i,j} \neq 0$ by (114), it follows from the definition (93) that there exists a sequence of points $u^k \in C(g)$ such that:

$$\begin{aligned} &\text{i) } u^k \neq \bar{u} \qquad \text{for all } k = 1, 2, 3, \ldots \\ &\text{ii) } \|u^k - \bar{u}\| \xrightarrow[k \to \infty]{} 0 \\ &\text{iii) } \frac{u^k - \bar{u}}{\|u^k - \bar{u}\|} \xrightarrow[k \to \infty]{} \frac{w^{i,j}}{\|w^{i,j}\|}. \end{aligned} \tag{115}$$

---

(63) Page 11.

(64) Page 37.

(65) Actually, Carathéodory's Theorem ensures that we can choose $q = n + 1$ (cf. [**39**], p. 155 (Theorem 17.1)). In fact, since $T_{\bar{u}}C(g)$ is a cone, we can choose $q = n$ ([**39**], p. 156, (Theorem 17.1.2)).

Since f is differentiable,

$$f(u^k) = f(\bar{u}) + f'(\bar{u}) \cdot (u^k - \bar{u}) + o(||u^k - \bar{u}||) \tag{116}$$

for all k, so:

$$\frac{f(u^k) - f(\bar{u})}{||u^k - \bar{u}||} = f'(\bar{u}) \cdot \frac{(u^k - \bar{u})}{||u^k - \bar{u}||} + \frac{o(||u^k - \bar{u})}{||u^k - \bar{u}||}. \tag{117}$$

Now by (115(iii)) the first term on the right hand side converges to $f(\bar{u}) \cdot w^{i,j}$, which is positive by (114), and the second term converges to 0. So for all large k we have $u^k \in C(g)$ and (applying (2) to $f'(\bar{u}) \cdot w^{i,j} > 0$) we have $f(u^k) > f(\bar{u})$, which is a contradiction of the constrained maximization hypothesis (105b). ∎

Theorem 3 shows that the Tangency-Path Criterion[66] is strong enough for Lagrange regularity. Now we show it is not too strong.

Theorem 4. (Inequalities.) Let U be an open subset of $\mathbb{R}^n$, and let $\bar{u} \in U$. Let $g : U \to \mathbb{R}^m$. If g is Lagrange inequality-regular for $(\bar{u}, \mathcal{F}_D(\bar{u}))$,[67] then g satisfies the Tangency-Path Criterion.[68]

Remark. The necessity property in Theorem 4 is analogous to the minimality properties of the Jacobian Criterion (Theorems 2.A, 2.B, 2.C). However, the Path Criterion is strictly weaker than the Jacobian Criterion, as shown by the Example, page 34. In fact, in view of Theorem 4, no further weaking is possible for $g \in \mathcal{G}_P(\bar{u})$ with constraint qualifications expressible in terms of initial derivatives of paths attending the constraint set $C(g)$, and for maximand functions in $\mathcal{F}_D(\bar{u})$.

---

[66] See definition page 37.

[67] See definition page 10.

[68] See definition page 37.

Proof. Suppose that g is Lagrange inequality-regular for $(\bar{u}, \mathcal{F}_D(\bar{u}))$,[69] so the g^i have partial derivatives at $\bar{u}$. To verify the Tangency-Path Criterion (98), we must show that: for any $z \in \mathbb{R}^n$,[70]

$$g_I'(\bar{u})z \geqq 0 \quad \Rightarrow \quad z \in V(g). \tag{118}$$

Without loss of generality, we will suppose that $\bar{u}$ is the origin.

a) If there are no binding constraints g^i, then $0 \in B_\varepsilon(0) \subseteqq C(g)$ for some $\varepsilon > 0$, so $V(g) = \mathbb{R}^n$, and we are done. So we will suppose that there are binding constraints, i.e.,[71] the set $I = \{1, \ldots, p\}$ is nonempty.

b) For a proof by contradiction, suppose the Tangency-Path Criterion (118) is not true, so there exists a $z \in L(g) \setminus V(g)$. Now $V(g) = \mathrm{cl}(\mathrm{ch}(T_{\bar{u}}C(g))))$ is a closed convex cone, not equal to $\mathbb{R}^n$ since it does not contain z; so by standard separating hyperplane theorems[72] there exists a $q \in \mathbb{R}^n$ such that:

$$q \neq 0 \tag{119a}$$

$$q \cdot u < 0 \qquad \text{for all } u \in V(g) \text{ with } -u \notin V(g) \tag{119b}$$

$$q \cdot u = 0 \qquad \text{for all } u \in V(g) \text{ with } -u \in V(g) \tag{119c}$$

$$q \cdot z > 0 \tag{119d}$$

$$q \cdot q = 1. \tag{119e}$$

We define the hyperplane H_q by:

$$H_q = \{u \in \mathbb{R}^n : q \cdot u = 0\}, \tag{120}$$

so q is orthogonal to H_q. And we define the subspace orthogonal to H_q:

$$Q = \{tq : t \in \mathbb{R}\}, \tag{121}$$

so $\mathbb{R}^n$ is the direct sum:

$$\mathbb{R}^n = H_q \oplus Q. \tag{122}$$

c) In what follows we shall define a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ that is maximized on $C(g)$ at $\bar{u} = 0$, and for which $f'(0) = q$. Then by Lagrange

---

(69) See definition page 10.

(70) I is defined in (16).

(71) See the definition of binding constraint, page 9, and the notation established in (16) and (17).

(72) Cf. [27], p. 315, Theorem 2.7.

inequality-regularity there must exist $\lambda_i \geqq 0$ such that $f'(0) = -\sum_{i=1}^{p} \lambda_i g^{i\prime}(0)$, so:

$$q \cdot z = f'(0) \cdot z \tag{123a}$$

$$= -\sum_{i=1}^{p} \lambda_i g^{i\prime}(0) \cdot z \tag{123b}$$

$$\leqq 0 \qquad \text{(by (99), since } z \in L(g)), \tag{123c}$$

which contradicts (119d).

d) Finding a function f with the properties mentioned in (c) is equivalent to finding a function $f : \mathbb{R}^n \to \mathbb{R}$ that is differentiable at 0, has $f_u(0) = q$, and is maximized on $C(g)$ at 0. That is because the matrix corresponding to the linear transformation $f_u(0)$ is represented in the standard basis of $\mathbb{R}^n$ by the Jacobian matrix of f.[73]

e) Our intuition in defining f is simple. If all of $C(q)$ lies "below" the hyperplane H_q, then $f(x + tq) = t$ clearly satisfies the requirements of part (c) above. Now properties (119b,c) almost imply that $V(g)$, hence its subset $C(g)$, lies below the hyperplane H_q. However they do allow $C(g)$ itself to rise "above" H_q — though only "gradually," much as the function $y = x^2$ rises above the x-axis. So we take as the graph of our function f (through the function $\tilde{\gamma}$ below) the "upper boundary" of $C(g)$, and show it has (like the function x^2) a zero derivative.

f) Next, some definitions:

$$\begin{aligned} B^k &= \{u \in U : \|u\| \leqq \frac{1}{2^k}\} \\ C^k &= C(g) \cap B^k \\ H_q^k &= \{u \in H_q : \|u\| \leqq \frac{1}{2^k}\}. \end{aligned} \tag{124}$$

g) To define the function f we first define, for each $k \in \mathbb{N}$, a function $\gamma^k : H_q \to \mathbb{R} \cup \{-\infty\}$; then we change γ^k into $\tilde{\gamma}^k$ so that its values belong to the closure of the ball B^k; finally, we define f in terms of $\tilde{\gamma}^k$.

---

[73] Cf. [40, Theorem 9.17, page 215].

We define for each $k \in I\!N$, the function $\gamma^k : H_q \to I\!R \cup \{-\infty\}$ by:[74]

$$\gamma^k(x) = \begin{cases} \sup\{t : x + tq \in C^k\}, & \text{if } \exists y[x + y \in C^k] \\ -\infty, & \text{otherwise.} \end{cases} \tag{125}$$

Because $C^k \subseteqq B^k$, the values of γ^k are either finite or $-\infty$. We note that:

$$\gamma^k(x) \geqq t \qquad \text{for all } x + tq \in C^k. \tag{126}$$

To bound the function below, we define:

$$\tilde{\gamma}^k(x) = \max\{0, \gamma^k(x)\} \qquad \text{for all } x \in H_q, \tag{127}$$

so:

$$\tilde{\gamma}^k \geqq \gamma^k \qquad \text{and} \qquad \tilde{\gamma}^k \geqq 0. \tag{128}$$

h) Next we show that for all large enough $k \in I\!N$:

$$\tilde{\gamma}^k(0) = 0. \tag{129}$$

First we note that for large enough $k \in I\!N$:

$$\neg\exists t_{\,0<t\in I\!R}\; tq \in \mathrm{cl}(C^k). \tag{130}$$

Otherwise there would exist a sequence of points $x^k + t^k q) \in C^k$ with $t^k > 0$ and:

$$\frac{||x^k||}{||t^k q||} \leqq \frac{1}{2^k} \xrightarrow[k\to\infty]{} 0 \tag{131}$$

and $x^k + t^k q \xrightarrow[k\to\infty]{} 0$ (since $x^k + t^k q \in C^k$). Then by definition (92a) and the compactness of the unit ball, there is a subsequence (we use the same index k for convenience) on which:

$$\frac{x^k + t^k q}{||x^k + t^k q||} \xrightarrow[k\to\infty]{} \text{some } \bar{w} \in T_0 C(g). \tag{132}$$

Since $\bar{\bar{w}} \in T_0 C(g) \subseteqq V(g)$, properties (119b,c) imply:

$$q \cdot \bar{w} \leqq 0. \tag{133}$$

But using (131) we see that, in the maximum norm we have, for large enough k:

$$\frac{x^k + t^k q}{||x^k + t^k q||} = \frac{x^k + t^k q}{||t^k q||} \xrightarrow[k\to\infty]{} \frac{q}{||q||}, \tag{134}$$

---

[74] We use the supremum rather than the maximum because the sets C^k are not necessarily closed. That is because the constraint functions g^i are only assumed to have partial derivatives, and only at the origin, so they are not necessarily continuous in any neighborhood of 0.

so $\bar{w} = q/\|q\|$. Then (133) implies $q \cdot q \leqq 0$, so $q = 0$, contradicting (119a) and completing the proof of (130). So for all large enough $k \in I\!N$ we have:

$$tq \in \mathrm{cl}(C^k) \;\Rightarrow\; t \leqq 0, \tag{135}$$

and therefore (129) holds.

i) Now we show that, for all sufficiently large $k \in I\!N$, the function $\tilde{\gamma}^k$ is differentiable at $0 \in H_q$, and:

$$\tilde{\gamma}^k \text{ is differentiable at } 0 \in H_q \qquad \text{and} \qquad d\tilde{\gamma}^k(0) = 0. \tag{136}$$

Let $x^i \in H_q$ with $0 \neq \|x^i\| \xrightarrow[i\to\infty]{} 0$. We must show that:

$$\frac{|\tilde{\gamma}^k(x^i) - \tilde{\gamma}^k(0)|}{\|x^i - 0\|} \xrightarrow[i\to\infty]{} 0 \in I\!R, \tag{137}$$

which by (128) and (129) means:

$$\frac{\tilde{\gamma}^k(x^i)}{\|x^i\|} \xrightarrow[i\to\infty]{} 0. \tag{138}$$

It suffices to consider infinite subsequences of all the i's for which $\tilde{\gamma}^k(x^i) > 0$, hence $\tilde{\gamma}^k(x^i) = \gamma^k(x^i)$. And it suffices to show that any such "positive" subsequence itself has a subsequence that satisfies (138). Now on any subsequence of a positive subsequence (we'll use "j" to remind us) we have:

$$0 < \tilde{\gamma}^k(x^j) = \gamma^k(x^j) \xrightarrow[j\to\infty]{} 0, \tag{139}$$

since otherwise this subsequence itself would have a subsequence converging to some $t > 0$; but then (by the definition (125)) there would exist $x^j + t^j q \in C^k$ converging to $tq \in \mathrm{cl}(C^k)$, contradicting (130).

Now (139) implies:

$$0 \neq \|x^j + \gamma^k(x^j)q\| \xrightarrow[j\to\infty]{} 0, \tag{140}$$

so by compactness of the unit ball, there is a subsequence of the j's on which we have convergence:

$$\frac{x^j + \gamma^k(x^j)q}{\|x^j + \gamma^k(x^j)q\|} \xrightarrow[j\to\infty]{} \text{some } \bar{\bar{w}}. \tag{141}$$

We will show that (138) holds on any such subsequence (141). First we note that:

$$\bar{\bar{w}} \in T_0 C(g). \tag{142}$$

In view of (140) and (141), that would be immediate from the definition (92) if $x^j + \gamma^k(x^j)q \in C^k$; but the definition (125) only requires that $\gamma^k(x^j)$ be the sup

of t for which $x^j + tq \in C^k$.[75] Nevertheless, it follows from (125) and (141) that there are t^j with $0 \neq x^j + t^j q \in C^k$, $||x^j + t^j q|| \to 0$, and:

$$\frac{x^j + t^j q}{||x^j + t^j q||} \xrightarrow[j\to\infty]{} \bar{\bar{w}}, \tag{143}$$

so (142) holds.

From (142) and (119b,c) we see that:

$$0 \geqq q \cdot \bar{\bar{w}}. \tag{144}$$

Then (141) and (144) imply:

$$\begin{aligned} 0 \geqq q \cdot \bar{\bar{w}} \xleftarrow[j\to\infty]{} & \; q \cdot \frac{x^j + \gamma^k(x^j)q}{||x^j + \gamma^k(x^j)q||} \\ &= \frac{\gamma^k(x^j)}{||x^j + \gamma^k(x^j)q||} q \cdot q \qquad (\text{since } x^j \in H_q) \\ &= \frac{\gamma^k(x^j)}{||x^j|| + ||\gamma^k(x^j)q||} \qquad (\text{using (119e) and the sum norm}) \\ &\geqq 0 \qquad (\text{since } \gamma^k(x^j) > 0), \end{aligned} \tag{145}$$

which implies (138).

j) Now we define $f : \mathbb{R}^n \to \mathbb{R}$ by:[76]

$$\begin{aligned} f(u) &= t - \tilde{\gamma}^K(x) \\ &\qquad \text{for all } u \in \mathbb{R}^n \text{ with } u = x + tq \;\;\&\;\; x \in H_q \;\;\&\;\; t \in \mathbb{R}, \end{aligned} \tag{147}$$

where K is any k so large that (129) and (136) hold, as in parts (h) and (i). Clearly $0 \in \mathbb{R}^n$ maximizes f on C^K, since $f(0) = 0$ by (147) and (129), and if $x + tq \in C^K \subseteqq C(g)$ then $f(x + tq) \leqq 0$ by (147), (126), and (128).

By our argument in parts (c) and (d) above, it only remains to show that f is differentiable at the origin $0 \in \mathbb{R}^n$, with $f_u(0) = q$. And for that it suffices to show:

$$\begin{aligned} f(x + tq) - f(0) &= f_u(0) \cdot (x + tq) + o(||x + tq||) \\ &= q \cdot (x + tq) + o(||x + tq||). \end{aligned} \tag{148}$$

---

[75] See footnote 74, page 45.

[76] If we wanted f to have a unique maximum at the origin, we could instead define it by:

$$f(u) = t - \tilde{\gamma}^K(x) - ||x||^2, \tag{146}$$

where $|| \cdot ||$ is the Euclidean norm.

Since $f(0) = 0$ and $q \cdot x = 0$ for $x \in H_q$, this becomes:

$$f(x + tq) = t + o(||x + tq||), \tag{149}$$

i.e.,

$$\frac{\gamma^K(x)}{||x + tq||} \xrightarrow[||x+tq||\to 0]{} 0. \tag{150}$$

Using the sum norm,[77] this becomes:

$$\frac{\gamma^K(x)}{||x|| + ||tq||} \xrightarrow[||x+tq||\to 0]{} 0, \tag{151}$$

which follows immediately from (136). ∎

While one could combine Theorems 3 and 4 into an "if and only if" theorem, we have separated them to make the proofs more readable. Because the remaining results are corollaries of these theorems, we state them in the symmetric fashion.

Theorem 5. (Equalities.) Let U be an open subset of $\mathbb{R}^n$, and let $\bar{u} \in U$. Suppose $h : U \to \mathbb{R}^k$. Then h is Lagrange equality-regular for $(\bar{u}, \mathcal{F}_D(\bar{u}))$[78] if and only if $\tilde{g}(h)$ satisfies the Tangency-Path Criterion at $\bar{u}$. In other words, if $\bar{u}$ maximizes $f : U \to \mathbb{R}$ subject to $h = 0$:

$$h(\bar{u}) = 0 \tag{152a}$$

$$\forall u_{\ u \in U \ \&\ h(u)=0}\ f(\bar{u}) \geqq f(u), \tag{152b}$$

and if f is differentiable at $\bar{u}$, then there exists a $\mu \in \mathbb{R}^k$ such that:

$$f'(\bar{u}) + \mu^T h'(\bar{u}) = 0, \tag{153a}$$

i.e., there exist real numbers $\mu_1, \ldots, \mu_k$ such that:

$$\frac{\partial f}{\partial u_i}(\bar{u}) + \mu_1 \frac{\partial h^1}{\partial u_i}(\bar{u}) + \ldots + \mu_k \frac{\partial h^k}{\partial u_i}(\bar{u}) = 0 \qquad (i = 1, \ldots, n) \tag{153b}$$

if and only if $\tilde{g}(h)$ satisfies the Tangency-Path Criterion[79] at $\bar{u}$.

---

[77] See (4c), page 6.

[78] See definition page 11.

[79] Page 37.

Proof. This follows immediately from Theorems 3 and 4, by converting the equalities to inequalities, as in (97). ∎

And finally we have the most general path theorem:

Theorem 6. (Equalities and inequalities.) Let U be an open subset of $\mathbb{R}^n$, and let $\bar{u} \in U$. Suppose $g : U \to \mathbb{R}^m$ and $h : U \to \mathbb{R}^k$. Then (g, h) is Lagrange regular for $(\bar{u}, \mathcal{F}_D(\bar{u}))$, i.e., for functions f differentiable at $\bar{u}$, if and only if $\tilde{g}(h)$ satisfies the Tangency-Path Criterion[80] at $\bar{u}$. In other words, if $\bar{u}$ maximizes $f : U \to \mathbb{R}$ subject to $h = 0$:

$$h(\bar{u}) = 0 \tag{154a}$$

$$\forall u_{\,u \in U \;\&\; h(u)=0}\; f(\bar{u}) \geqq f(u), \tag{154b}$$

and if f is differentiable at $\bar{u}$, then there exists a $\lambda \in \mathbb{R}^m$ and a $\mu \in \mathbb{R}^k$ such that:

$$f'(\bar{u}) + \lambda^T g'(\bar{u}) + \mu^T h'(\bar{u}) = 0, \tag{155a}$$

i.e.,

$$\begin{aligned} \frac{\partial f}{\partial u_i}(\bar{u}) + \lambda_1 \frac{\partial g^1}{\partial u_i}(\bar{u}) + \ldots + \lambda_m \frac{\partial g^m}{\partial u_i}(\bar{u}) \\ + \mu_1 \frac{\partial h^1}{\partial u_i}(\bar{u}) + \ldots + \mu_k \frac{\partial h^k}{\partial u_i}(\bar{u}) = 0 \qquad (i = 1, \ldots, n), \end{aligned} \tag{155b}$$

and such that:

$$\lambda \geqq 0 \tag{156a}$$

$$\lambda_i = 0 \;\text{ if } i \in J, \tag{156b}$$

if and only if $\tilde{g}(h)$ satisfies the Tangency-Path Criterion[81] at $\bar{u}$.

Proof. This again follows from Theorems 3 and 4, by converting equalities to inequalities, as in (97). ∎

---

[80] Page 37.

[81] Page 37.

VIII Comparison of Jacobian and Tangency-Path Conditions

The Tangency-Path Criterion has the advantage of characterizing Lagrange regularity (Theorems 3, 4, 5, and 6). But it is an existential assertion — a statement that, for every vector z of a certain kind, there exists a path whose initial tangent is z, and which attends[82] the constraint set. And that may be hard to verify in particular instances.

On the other hand, the Jacobian Criteria have the advantage of being computable, in a sense we will describe below. But they do not completely characterize Lagrange regularity. (See the remarks concerning (90) and (91), page 34).) Being minimal conditions within the class of Jacobian conditions (as in Theorems 2.A, 2.B, and 2.C) is not the same as being necessary as well as sufficient for Lagrange regularity of constraint functions.

As to computability, let us explain what we mean when we say that the Jacobian criteria are computable. The idea is simple: the criteria can be applied using only simple rules and elementary arithmetic operations (addition, multiplication, and division). More specifically, beginning with a particular Jacobian matrix — $h'(\bar{u})$, for example — we know rules, or algorithms, from algebra that enable us to calculate the rank of the matrix. If the elements of the matrix were integers, then we could formalize the notion of algorithm and talk of a recursive function (or a Turing machine, or some other standard equivalent computability concept) yielding the result of the calculation. If the elements were rational numbers, we could represent them by pairs of integers, and again talk of computability in terms of recursive functions.

If, on the other hand, some of the matrix elements are irrational numbers, then we are in a context more general than basic recursion theory. While it is possible to formalize notions of computability in such a context,[83] we will be content with a few intuitive observations.

The first problem we face in explaining any notion of computability for real numbers is to determine how the numbers are presented. One approach would consider them to be presented simply as primitive entities *sui generis*. If we take the latter approach, and if we take as primitive notions the usual algebraic operations and relations: addition, multiplication, division, equality, and greater than, then we can "compute" the rank of a matrix, and determine if the rank has a given value. In any particular instance, then we can determine, in a "computable" manner, whether or not our Jacobian criterion (35) is true.

---

(82) Page 36.

(83) Cf. [8] and [38], for example.

Similar remarks apply to the computability of (34a), for example. We know that the existence of a solution ξ to such a system of inequalities can be determined in an algorithmic fashion by "elimination of quantifiers." That is a consequence of Tarski's celebrated theorem on the decidability of real closed fields [**45**]. In fact, because we are concerned with a system of *linear* inequalities, there are very simple algorithms (Fourier elimination) for eliminating the quantifiers.[84]

Similar remarks also apply to the computability of (34b). If we write $g'_I(\bar{u})$ as a $p \times n$ matrix A, then that wedge property is clearly equivalent to the statement that, for each vector $\tilde{e}_i$ $(i = 1, \ldots, 2n)$ that is either a unit coordinate vector or the negative of one, there exists a $t \in \mathbb{R}^p$ such that

$$\begin{aligned} &A^T t = \tilde{e}_i \\ &t \geqq 0. \end{aligned} \tag{157}$$

Again, simple Fourier elimination provides an algorithm for determining whether or not (157) is solvable for t.

The same kind of computability arguments we have applied to (35) and (34) can be applied to show that our other Jacobian criterion (33) is a computable condition. By contrast, it does not seem likely that such simple algorithms exist to determine whether the Tangency-Path Criterion holds. Of course, to prove that would require a formal definition of computability, taking us beyond the scope of the present paper.

---

[84] Cf. [**29**], [**43**].

IX Appendix

Our understanding of constraint qualifications, and our proofs of Lagrange regularity rest on one algebraic theorem and one analysis theorem. The algebraic theorem is a theorem of the alternative, or a transposition theorem.

Theorem of the Alternative. Let A, B, and C be matrices whose components are from $\mathbb{R}$. Suppose that

A has a rows and n columns, and $\alpha \in \mathbb{R}^a$,
B has b rows and n columns, and $\beta \in \mathbb{R}^b$,
C has c rows and n columns, and $\gamma \in \mathbb{R}^c$,

where $0 < a \in \mathbb{N}$, $0 < b \in \mathbb{N}$, $0 < c \in \mathbb{N}$, and $0 < n \in \mathbb{N}$.

Then:

I) Exactly one of (1) or (2) is true:

1) There exists $r = (r_1, ..., r_n) \in \mathbb{R}^n$ solving:

$$Ar > \alpha$$
$$Br \geqq \beta$$
$$Cr = \gamma.$$

2) There exists $u \in \mathbb{R}^a$, $v \in \mathbb{R}^b$, and $z \in \mathbb{R}^c$, such that both (a) and (b) hold:

a) $u^T A + v^T B + z^T C = 0$

b) either (i) or (ii) hold:

i) $u \geq 0 \;\&\; v \geqq 0 \;\&\; u^T\alpha + v^T\beta + z^T\gamma \geqq 0$

or

ii) $u = 0 \;\&\; v \geqq 0 \;\&\; v^T\beta + z^T\gamma > 0.$

II) When some, but not all, of (A, α) or (B, β) or (C, γ) are not present, then the same alternatives (1) and (2) in part (I) above hold with these modifications:

In (I,1):

remove the first row if A and α are not present;
remove the second row if B and β are not present;
remove the first row if C and γ are not present.

In (2,a):

set $A = 0$ and $u = 0$ if A and α are not present;
set $B = 0$ and $v = 0$ if B and β are not present;
set $C = 0$ and $z = 0$ if C and γ are not present.

In (2,b):

only case (ii) occurs if A and α are not present;
set $\beta = 0$ and $v = 0$ if B and β are not present;
set $\gamma = 0$ and $z = 0$ if C and γ are not present.

The theorem can be proved from the Transposition Theorem of Motzkin [**36**], [**35**, pp. 28(2)–29] which is a homogeneous version (equivalent to setting $\alpha = 0$, $\beta = 0$, and $\gamma = 0$). A proof by Fourier elimination, along the lines of [**43**, pp. 1–20] or [**29**] yields a proof of the remark above concerning ordered fields and subfields.

The other pillar of our approach is an implicit function theorem with weaker than standard differentiability hypotheses. This is a special case of Theorem 1 of [**24**].

A Non-C^1 Implicit Function Theorem. Let $X \times Y$ be an open subset of $\mathbb{R}^n \times \mathbb{R}^k$ and $(\bar{x}, \bar{y}) \in X \times Y$. Suppose $\psi : X \times Y \to \mathbb{R}^k$ is differentiable at $(\bar{x}, \bar{y})$, and suppose that

$$\psi(\bar{x}, \bar{y}) = 0; \tag{158a}$$

$$\psi(x, \cdot\,) \text{ is continuous on } Y, \text{ for all } x \in X; \tag{158b}$$

$$\psi_y(\bar{x}, \bar{y}) \text{ is surjective; i.e.,} \tag{158c}$$

$$\det\left(\begin{bmatrix} \dfrac{\partial \psi^1(\bar{x}, \bar{y})}{\partial y_1} & \cdots & \dfrac{\partial \psi^1(\bar{x}, \bar{y})}{\partial y_k} \\ \vdots & & \vdots \\ \dfrac{\partial \psi^k(\bar{x}, \bar{y})}{\partial y_1} & \cdots & \dfrac{\partial \psi^k(\bar{z})}{\partial y_k} \end{bmatrix}\right) \neq 0.$$

Then:

a) There exists an open neighborhood $X_0 \times Y_0 \subseteqq X \times Y$ of $(\bar{x}, \bar{y})$ and a function $\phi : X_0 \to Y_0$ such that:

$$\psi(x, \phi(x)) = 0 \qquad \text{for all } x \in X_0 \tag{159a}$$

$$\phi(\bar{x}) = \bar{y}. \tag{159b}$$

b) Every function $\phi : X_0 \to Y_0$ satisfying (159) is differentiable at $\bar{x}$, with:

$$\phi'(\bar{x}) = (\psi_y(\bar{x}, \bar{y}))^{-1} \psi_x(\bar{x}, \bar{y}). \tag{160}$$

X Historical Comments and Comparisons

The beginnings. Lagrange illustrated for one equality constraint, and suggested for multiple constraints a "general principle":[85]

> When a function of several variables is to have a maximum or a minimum, and when there are one or more equations among these variables, it will suffice to add to the proposed function the functions which must vanish, each multiplied by an undetermined quantity, and then to seek the maximum or minimum as if the variables were independent; the equations that one will find, combined with the given equations, will serve to determine all the unknowns.

Lagrange refers to maximizing or minimizing what we now call the Lagrangean function, but his analytics deal only with first order conditions. Modern theory embodies the maximization (or minimization) view in duality saddle point theorems, when certain convexity assumptions hold. In particular, there is now a bifurcation in the theory of constrained maximization. On the one hand, assuming these convexity properties, theorems (such as Kuhn and Tucker's Theorem 3 [**30**]) assert that constrained maxima correspond to saddle points of the Lagrangean function. These theorems do not generally require differentiability hypotheses (cf. Uzawa's [**46**]). On the other hand, a second class of results obtains the existence of Lagrange Multipliers through assuming differentiability properties but avoiding convexity assumptions. This second class of results provides an important working tool for applied mathematics, but is of particular interest to economists, since by avoiding convexity hypotheses it permits the analysis of such phenomena as increasing returns to scale.

Euler [**14**] has been credited ([**9**], [**11**], [**17**], [**28**]) with originating a principle ("Euler's rule"), precursor of the Lagrange approach, for extremizing functions subject to constraints. In the context of the isoperimetric problem of the calculus of variations, Euler's rule states that minimizing $\int_{t_0}^{t_1}(F+\lambda G)dt$ for some multiplier λ yields the same first order conditions as minimizing $\int_{t_0}^{t_1} F dt$ subject to $\int_{t_0}^{t_1} G dt =$ constant. But according to Carathéodory [**11**, p. 177], Lagrange was the first to recognize the fundamental significance of the parameters that came to be known as the Lagrange multipliers.

---

[85] Our translation of the end of Section 58, in Chapter XI of the Second Part of [**32**]. An earlier exposition of the idea is contained in [**19**], QUATRIEME SECTION, paragraphs 1-8, pages 44–49.

Modern treatments of the Lagrange Multiplier Theorem for equalities are found in Bolza [**9**],[86] Carathéodory [**11**], Bliss [**6**], [**7**], as well as many recent textbooks. An earlier proof, applicable to real analytic functions, is found in Weierstrass [**49**].

Classical, Karush, and Kuhn-Tucker approaches. In order to establish existence of Lagrange multipliers, modern treatments of equality-constrained optimization, such as those mentioned above, assume the constraint functions are C^1 and have a Jacobian of maximum rank. The C^1 hypothesis has been needed to apply the classical implicit function theorem, which uses a C^1 hypothesis.

Hancock [**20**, p. 150] suggested handling constrained maximization problems with inequality constraints ("limitations") by converting them into equality problems. A constraint $g(x) \geqq 0$ would be replaced by the equality $g(x) + z^2 = 0$.

The same conversion device was used by Valentine [**47**] in the treatment of a calculus of variations minimization problem with inequality side conditions. In Corollary 3:4, p. 9 (of the dissertation) and p. 415 (of the reprinting), he shows that the Lagrange coefficients for the inequality constraints are nonpositive (corresponding to nonnegative constraints in maximization problems).

Karush [**26**] also used the squared slack variable device to prove one of his results (cf. his Theorem 3:1, pp. 11–13). His remarks noted that, with a "normality" condition and C^2 constraints, the Lagrange coefficients would be nonpositive in his minimization problem.[87]

Karush's work went unnoticed for many years, although it contains most of the basic concepts and many of the results of later work. His Theorem 3:2, for example, uses what he calls "Condition Q" and today is called the "Kuhn-Tucker Constraint Qualification," to obtain the existence of Lagrange multipliers of appropriate sign. Condition Q also marks a transition from Jacobian rank conditions to path conditions. And his Theorem 3:3 shows that a certain positivity Jacobian condition, is sufficient for his constraint qualification to hold; this condition is essentially the same as our (34*a*) and as the inequality in Arrow-Hurwicz-Uzawa's Theorem 3. And he notes its computability by observing that a standard algebraic theorem "... provides a useful method for determining in a finite number of; steps whether or not such an admissible vector ... does exist" [**26**, p. 11]. Although Karush assumes that his inequality constraint functions are

---

(86) Bolza's formulation makes clear that C^1 is a sufficient smoothness condition for Lagrange regularity. Earlier contributions appear to postulate real-analyticity.

(87) Neither Valentine nor Karush seem to refer to Hancock, and Karush does not refer to Valentine.

C^1, it is clear from his proof that he only uses differentiability at the maximizing point.

Fritz John [**25**] formulated first order necessary conditions for a minimum of a C^1 function f of finitely many variables subject to a (finite or infinite) set of C^1 inequality constraints $g_\alpha \geqq 0$ $(\alpha \in A)$. He showed that there exists a vector $(\lambda_0, \lambda_{\alpha_1}, \ldots, \lambda_{\alpha_m}) \neq 0$ with $\lambda_0 \geqq 0$ and $\lambda_{\alpha_1}, \ldots, \lambda_{\alpha_m} > 0$, such that, at a minimizing point $\bar{x}$:

$$\frac{\partial L}{\partial x_i}(\bar{x}) = 0 \qquad (i = 1, \ldots, n), \tag{161}$$

where L is defined by:[88]

$$L(x) = \lambda_0 f - (\lambda_{\alpha_1} g^{\alpha_1} + \cdots + \lambda_{\alpha_r} g^{\alpha_r}). \tag{162}$$

Since he did not postulate any constraint qualifications, it may happen that $\lambda_0 \neq 0$. (By contrast, Karush and Kuhn and Tucker, assuming constraint qualifications, are able to obtain $\lambda_0 = 1$.)[89]

As far as first order necessary consequences of constrained maximization is concerned, Kuhn and Tucker's paper [**30**] independently retraced part of the path traveled earlier by Karush. Their Theorem 1 is similar to Karush's Theorem 3:2, although Kuhn and Tucker are interested in *nonnegative* solutions of the inequality constraints, and they provide more information about the Lagrange multipliers (analogous to our condition (20b)); they also reduced the C^1 hypothesis on constraints to differentiability. Kuhn later became aware of Karush's work and earlier references to it ([**13**], [**37**], [**44**], et al.), and related research by others. See [**31**] for his very informative historical account, and some general comments on the several quite divergent interests that converged in similar theorems. Because they use a constraint qualification, they are able to obtain a result for a Lagrangean of the form $f(x) + \lambda_1 g^1(x) + \cdots + \lambda_m g^m(x)$, as distinct from $\lambda_0 f(x) + \lambda_1 g^1(x) + \cdots + \lambda_m g^m(x)$, which would be the Fritz John analogue.

Kuhn and Tucker only assumed differentiability of the constraint functions, in contrast to the C^1 assumptions of Karush and Fritz John. Their differentiability assumption remains stronger, however, than required by our Tangency-Path Criterion.

---

[88] For the case of equality constraints, Bolza [9, p. 547] attributes to Hilbert a similar Lagrangean expression, with λ_0 multiplying the maximand and equal to 1 or 0 according to the rank of the constraint's Jacobian.

[89] In his Theorem 3:1, Karush also had a similar result, in the absence of any constraint qualification.

Arrow, Hurwicz, and Uzawa's Constraint Qualification W [**4**] weakened the Kuhn-tucker constraint qualification. Their Theorem 1 proved that W is sufficient for Lagrange regularity, and their Theorem 2 proved that W is necessary for Lagrange regularity when the constraint set is convex.

Pennisi [**37**] considered maximization constrained by both inequalities and equalities. He obtained what we call Lagrange regularity, implicitly assuming all inequality constraints to be effective (satisfied with equality at the maximizing point), and assuming "normality" in the sense of Bliss.

Mangasarian and Fromovitz [**34**] also addressed maximization problems with both equality and inequality constraints. Without a constraint qualification, they obtained(90) an analogue of the Fritz John result, in which the coefficient of the maximand is allowed to be zero. Then, they introduced a new Jacobian-type constraint qualification, combining a condition of Arrow, Hurwicz, and Uzawa for inequality problems with the classical rank condition for equality problems; their new condition guaranteed that the coefficient of the maximand could be taken equal to one. In both results, the constraints were assumed to be continuously differentiable.

In Mangasarian [**35**, p. 173, Theorem 6, part (iv)] the second result of [**34**] (under the "modified Arrow-Hurwicz-Uzawa constraint qualification") is obtained with the hypothesis on the inequality constraints reduced from continuous differentiability to differentiability. In our Theorem 1.A above it is reduced still further to differentiability and local continuity at the constrained maximizing point.

We have mentioned how our results compare with a few of those above. For more detailed comparisons, we will distinguish two categories: (a) Jacobian conditions, and (b) path conditions.

(a) Jacobian criteria. Here the main innovations are three-fold. First, we reduce the smoothness requirements on equality constraints from C^1 in a neighborhood to differentiability at the maximizing point and continuity in a neighborhood (cf. Theorem 1(A,B), and Corollary 1A). In this way the hypotheses for equality-constrained problems become more like the hypotheses for inequality-constrained problems mentioned below.(91) The weaker smoothness condition is made possible by the Non-C^1 Implicit Function Theorem.(92)

(90) See [**34**, pp. 41–42]

(91) This is in the spirit of the differentiability hypotheses of Theorem 3 and Corollary 6 of [**4**].

(92) Which, in fact, was developed with these applications in mind [**24**].

Second, we introduce the notion of minimal Jacobian conditions. And third, we provide new Jacobian criteria and prove they are minimal.

Theorem 1.A uses a Jacobian Criterion weaker than earlier Jacobian constraint qualifications, and it also relaxes the requirement that h be C^1 to mere differentiability at the maximum point and continuity locally, as indicated above. It is a generalization of Mangasarian's Theorem 6(iv), page 173 of [**35**] in several respects. First, our Mixed-Problem Jacobian Criterion (33a) is weaker: part (a) of Theorem 1.A corresponds to Mangasarian's "modified Arrow-Hurwicz-Uzawa constraint qualification," but if (a) is not satisfied, then our part (b) provides another alternative (which, according to Theorem 2.A, makes the combination the weakest possible Jacobian condition).

Kuhn and Tucker are explicit about the need for a constraint qualification, giving an example ([**30**, pp. 483–484]) in which Lagrange regularity fails in its absence. Our examples (14) and (15), for which Lagrange multipliers may not exist, are similar in spirit. Slater also had an example illustrating the need for a constraint qualification; although it was proposed in the context of concave programming and the saddle Point Equivalence Theorem, rather than Lagrange regularity, it is applicable in the present setting as well.

Theorem 1.B uses a Jacobian Criterion weaker than earlier Jacobian constraint qualifications for inequality-constrained problems, and it only requires Gâteaux differentiability at $\bar{u}$. If we consider Mangasarian's Theorem 6(iv), [**35**, p 173], when it is restricted to the case of inequality constraints, then our Theorem 1.B has a weaker constraint qualification (Mangasarian's "modified Arrow-Hurwicz-Uzawa constraint qualification" again amounts to part (a) of our Criterion).

Theorem 1.C generalizes the classical Lagrange Multiplier Theorem by reducing the traditional C^1 hypothesis on the constraint functions to mere differentiability at the constrained maximizer.

The paper introduces the concept of Lagrange regularity for matrices, and uses it in Theorems 2.A, 2.B, and 2.C, to establish that the Jacobian Criteria are "minimal" Jacobian constraint qualifications. We are not aware of earlier results along these lines.

(b) Path criteria. We developed our Tangency-Path Criterion[(93)] based in part on Hestenes' use of the tangent cone ([**21**, pp. 25 ff.], [**22**, pp. 203 ff.]) and in part by analogy with Constraint Qualification W of Arrow, Hurwicz, and Uzawa [**4**]. Subsequently we discovered that the paper by Gould and Tolle [**18**]

(93) Page 37.

contains a closely related, but stronger constraint qualification. Although their condition is stated in terms of dual (polar) cones, and although it is stated for both equalities and inequalities, if we state it in our terminology and express the equalities through pairs of inequalities, it amounts to $L(g) = V(g)$, in contrast to our weaker condition $L(g) \subseteqq V(g)$. Furthermore, they assume that the constraint functions are continuous in a neighborhood of the constrained maximum and are differentiable at the constrained maximum, while we only require existence of partial derivatives at the constrained maximum and impose no continuity requirements beyond that.

By itself, the inclusion condition (98b) is weaker than Gould and Tolle's equality condition. This can be seen from the following example:[94]

$$\begin{aligned} g^1(x_1, x_2, y) &= \begin{cases} y, & \text{for } x_1 = 0 \\ -1, & \text{otherwise} \end{cases} \\ g^2(x_1, x_2, y) &= \begin{cases} -y, & \text{for } x_2 = 0 \\ -1, & \text{otherwise.} \end{cases} \end{aligned} \tag{163}$$

This pair (g^1, g^2) fails to satisfy the Fréchet differentiability and continuity conditions that Gould and Tolle impose, as well as their $L(g) = V(g)$ constraint qualification. Nevertheless, it is partially differentiable at the origin $(0, 0, 0)$, and satisfies our criterion since $L(g) \subsetneqq V(g)$ and so by our Theorem 3 it is Lagrange inequality-regular (as a direct proof also shows). However, if one imposed Gould and Tolle's stronger continuity and differentiability assumptions[95] on constraints, then our subset condition would imply their equality condition (by [**1**, Lemma 4]).

Despite the stronger hypotheses of Gould and Tolle, and the more demanding constraint qualification, the proofs we had developed for our Theorems 3 and 4 turned out to be similar to corresponding parts of Gould and Tolle's proof. Indeed, it appears that parts of their proofs could be used to yield alternative proofs of Theorems 3 and 4.

Reducing smoothness requirements and weakening the constraint qualifications are really two sides of the same coin. Under the weak partial derivative hypothesis, the weak constraint qualification (the Tangency-Path Criterion) is necessary and sufficient for Lagrange regularity (Theorems 3 and 4 above); under the stronger differentiability hypothesis, the stronger constraint qualification is necessary and sufficient for Lagrange regularity (Gould and Tolle's Theorem).

---

[94] See also (102), p. 39, and (104), p. 39.

[95] Stronger than the existence of partial derivatives postulated in the Tangency-Path Criterion.

The tangent cone was applied to optimization problems by Hestenes [**21**], [**22**], Abadie [**1**], and Varaiya [**48**]. It was defined by Bouligand [**10**, paragraph 68, pp.65–66] (as the *contingent* set).

To the best of our knowledge, the Tangency-Path Criterion, in its present form, is new. As a sufficient condition (Theorem 3) for Lagrange regularity for inequalities (hence, by conversion, also equalities), it is weaker than any of the previously proposed constraint qualifications. Furthermore, the Criterion applies to a wider class of constraint functions, because it requires only the existence of partial derivatives, and only at the maximizing point.

Generalized derivative notions. Finally we mention another direction in which one can extend the notion of Lagrange regularity. Clarke uses the concept of generalized gradient to state a "Lagrange Multiplier Rule," [**12**, Theorem 6.1.1, page 228]. He does not assume the existence of partial derivatives, but he does assume that the functions f, g_i, h_j are "Lipschitz near any given point." This hypothesis is neither stronger nor weaker than ours, and the conclusion is weaker. Instead of the Lagrange Multiplier equality, he obtains the weaker condition "$0 \in \partial_x L(x, \lambda, r, s, k)$," where ∂_x denotes the subdifferential at x.

References

1. J. Abadie, *On the Kuhn-Tucker Theorem*, North-Holland, New York, NY, 1967, Nonlinear Programming.

2. K. J. Arrow and L. Hurwicz, *Reduction of Constrained Maxima to Saddle-Point Problems*, in Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, vol. V, University of California Press, Berkeley and Los Angeles, 1956, 1–20.

3. ———, eds., *Studies in Resource Allocation Processes*, Cambridge University Press, Cambridge, 1978.

4. K. J. Arrow, L. Hurwicz and H. Uzawa, *Constraint Qualifications in Maximization Problems*, Naval Research Logistics Quarterly **8** (1961), 175–191.

5. A. Avez, *Differential Calculus*, John Wiley & Sons, New York, NY, 1986.

6. G. A. Bliss, *Normality and Abnormality in the Calculus of Variations*, Transactions of the American Mathematical Society **43** (1938), 365–376.

7. ———, *Lectures on the Calculus of Variations*, University of Chicago Press, Chicago, 1946.

8. L. Blum, M. Shub and S. Smale, *On a Theory of Computation and Complexity over the Real Numbers: NP-Completeness, Recursive Functions and Universal Machines*, Bulletin of the American Mathematical Society **21** (1989), 1–46.

9. O. Bolza, *Vorlesungen über Variationsrechnung*, Chelsea Publishing Company, New York, NY, no date, author's preface dated 1909, second edition.

10. G. Bouligand, *Introduction à la Géométrie Infinitésimale Directe*, Librairie Vuibert, Paris, 1932.

11. C. Carathéodory, *Calculus of Variations and Partial Differential Equations of the First Order*, Chelsea Publishing Company, New York, NY, 1982, originally published as *Variationsrechnung und Partielle Differentialgleichungen erster Ordnung*, B. G. Teubner, Berlin, 1935.

12. F. H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley & Sons, New York, NY, 1983.

13. M. A. El-Hodiri, *The Karush Characterization of Constrained Extrema of Functions of a Finite number of Variables*, UAR Ministry of Treasury Research Memoranda, Series A, No. 3, July, 1967.

14. L. Euler, *Methodus Inveniendi Lineas Curvas, in Leonhardi Euleri Opera Omnia, series prima, XXIV, vol. 34*, Swiss Society of Natural Sciences, Bern, 1952, reprinted from *Methodus Inveniendi Lineas Curvas*, Lausanne & Geneva, 1744.

15. J. P. Evans, *On Constraint Qualifications in Nonlinear Programming*, Center for Mathematical Studies in Business and Economics, Graduate School of Business, University of Chicago, Report 6917, Chicago, May, 1969.

16. W. Fenchel, *Convex Cones, Sets and Functions*, Princeton University. Department of Mathematics, Princeton, 1953, from notes by D. W. Blackett of lectures at Princeton University, Spring term, 1951.

17. H. H. Goldstine, *A History of the Calculus of Variations*, Springer-Verlag, New York, NY, 1980.

18. F. J. Gould and J. W. Tolle, *A Necessary and Sufficient Qualification for Constrained Optimization*, SIAM Journal of Applied Mathematics **20** (1971), 164–172.

19. J. L. de la Grange, *Méchanique Analitique*, Chez la Veuve Desaint, Libraire, Paris, 1788. See also [**33**].

20. H. Hancock, *Theory of Maxima and Minima*, Ginn and Company, Boston, 1917.

21. M. R. Hestenes, *Calculus of Variations and Optimal Control Theory*, John Wiley & Sons, New York, NY, 1966.

22. ———, *Optimization Theory*, John Wiley & Sons, New York, NY, 1975.

23. J. R. Hicks, *Value and Capital*, Oxford University Press, Oxford, 1939.

24. L. Hurwicz and M. K. Richter, *Implicit Functions and Diffeomorphisms without C^1*, Discussion Paper No. 279, Department of Economics, University of Minnesota, 1994.

25. F. John, *Extremum Problems with Inequalities as Subsidiary Conditions*, in Studies and Essays: Courant Anniversary Volume, K. O. Friedrichs, O. E. Neugebauer and J. J. Stoker, eds., Interscience Publishers, New York, 1948, 187–204.

26. W. Karush, *Minima of Functions of Several Variables with Inequalities as Side Conditions*, 1939, Master of Science Dissertation, Department of Mathematics, University of Chicago.

27. V. L. Klee, Jr., *Separation Properties of Convex Cones*, Proceedings of the American Mathematical Sociaty **6** (1955), 313–318.

28. A. Kneser, *Variationsrechnung*, Teubner, Leipzig, 1899-1916, originally published as *Heft 5*, 1904.

29. H. W. Kuhn, *Solvability and Consistency for Linear Equations and Inequalities*, American Mathematical Monthly **63** (1956), 217–232.

30. H. W. Kuhn and A. W. Tucker, *Nonlinear Programming*, in Second Berkeley Symposium on Mathematical Statistics and Probability, J. Neyman, ed., University of California Press, 1951, 481–492.

31. H. W. Kuhn, *Nonlinear Programming: A Historical View*, Nonlinear Programming, American Mathematical Society, Providence, RI, 1976, in *Nonlinear Programming*, ed. by Richard W. Cottle and C. E. Lemke.

32. J. L. Lagrange, *Théorie des Fonctions Analytiques*, Courcier, Paris, 1813, Nouvelle édition.

33. ———, *Mécanique Analytique*, Gauthier-Villars et Fils, Bureau des Longitudes, de l'École Polytechnique, Paris, 1888, in volumes XI and XII of *Ouevres*. See also [**19**].

34. O. L. Mangasarian and S. Fromovitz, *The Fritz John Necessary Optimality Conditions in the Presence of Equality and Inequality Constraints*, Journal of Mathematical Analysis and Applications **17** (1967), 37–47.

35. O. L. Mangasarian, *Nonlinear Programming*, McGraw-Hill, New York, NY, 1969.

36. T. S. Motzkin, *Beiträge zur Theorie der linearen Ungleichungen*, Basle, 1934, (Dissertation).

37. L. L. Pennisi, *An Indirect Sufficiency Proof for the Problem of Lagrange with Differential Inequalities as Added Side Conditions*, Transactions of the American Mathematical Society **74** (1953), 177–198.

38. M. B. Pour-El and J. I. Richards, *Computability in Analysis and Physics*, Springer-Verlag, New York, NY, 1989.

39. R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.

40. W. Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, New York, NY, 1976, third edition.

41. P. A. Samuelson, *Foundations of Economic Analysis*, Harvard University Press, Cambridge, MA, 1947.

42. M. Slater, *Lagrange Multipliers Revisited: A Contribution to Nonlinear Programming*, Cowles Commission Discussion Paper, Math. 403, November, 1950.

43. J. Stoer and C. Witzgall, *Convexity and Optimization in Finite Dimensions I*, Springer-Verlag, New York, NY, 1970.

44. A. Takayama, *Mathematical Economics*, The Dryden Press, Hinsdale, IL, 1974.

45. A. Tarski, *A Decision Method for Elementary Algebra and Geometry*, University of California Press, Berkeley, 1951, Second Edition, Revised.

46. H. Uzawa, *The Kuhn-Tucker Theorem in Concave Programming*, in Studies in Linear and Non-Linear Programming, K. J. Arrow, L. Hurwicz and H. Uzawa, eds., Stanford University Press, Stanford, 1958, 32–37.

47. F. A. Valentine, *The Problem of Lagrange with Differential Inequalities as Added Side Conditions*, in Contributions to the Calculus of Variations 1933–1937: Theses Submitted to the Department of Mathematics of the University of Chicago, University of Chicago Press, Chicago, 1937, 403–447.

48. P. P. Varaiya, *Nonlinear Programming in Banach Space*, SIAM J. Appl. Math. **15** (1967), 284–293.

49. K. Weierstrass, *Mathematische Werke*, Akademische Verlagsgesellschaft, Leipzig, 1927, edited from notes of Weierstrass' 1875-188 lectures.