Ec 121a
Theory of Value

KC Border
Fall 2020

Lecture 9: Quasiconcavity and Maximization

9.1 Maximization and Inequality Constraints

When I was a student, the following theorem was known as the Kuhn–Tucker Theorem. It is now known as the Karush–Kuhn–Tucker Theorem. See the references at the end and Kuhn [5, 6] for a discussion of the history.

9.1.1 Theorem (Karush–Kuhn–Tucker) *Let $f, g_1, \ldots, g_m \colon \boldsymbol{R}^{\mathrm{n}}_{+} \to \boldsymbol{R}$ be differentiable at x^*, and let x^* be a constrained local maximizer of f subject to $g(x) \geqq 0$ and $x \geqq 0$.*

Let $B = \{i : g_i(x^*) = 0\}$ denote the set of binding functional constraints, and let $Z = \{j : x_j = 0\}$ denote the set of binding nonnegativity constraints on the variables. Assume that x^* satisfies the Kuhn–Tucker Constraint Qualification (see Section 9.3 below). Then there exists $\lambda^* \in \boldsymbol{R}^{\mathrm{m}}$ such that

$$f'(x^*) + \sum_{i=1}^{m} \lambda_i^* {g_i}'(x^*) \leqq 0,$$

$$x^* \cdot \left(f'(x^*) + \sum_{i=1}^{m} \lambda_i^* {g_i}'(x^*) \right) = 0,$$

$$\lambda^* \geqq 0,$$

$$\lambda^* \cdot g(x^*) = 0.$$

9.2 Karush–Kuhn–Tucker Theorem for Minimization

9.2.1 Theorem (Karush–Kuhn–Tucker) *Let $f, g_1, \ldots, g_m \colon \boldsymbol{R}^{\mathrm{n}}_{+} \to \boldsymbol{R}$ be differentiable at x^*, and let x^* be a constrained local minimizer of f subject to $g(x) \geqq 0$ and $x \geqq 0$.*

Let $B = \{i : g_i(x^*) = 0\}$, and let $Z = \{j : x_j^* = 0\}$,. Assume that x^* satisfies the Kuhn–Tucker Constraint Qualification. Then there exists $\lambda^* \in \boldsymbol{R}^{\mathrm{m}}$ such that

$$f'(x^*) - \sum_{i=1}^{m} \lambda_i^* {g_i}'(x^*) \geqq 0,$$

$$x^* \cdot \left(f'(x^*) - \sum_{i=1}^{m} \lambda_i^* {g_i}'(x^*) \right) = 0,$$

$$\lambda^* \geqq 0,$$

$$\lambda^* \cdot g(x^*) = 0.$$

Proof: Minimizing f is the same as maximizing $-f$. The Kuhn–Tucker conditions for this imply that there exists $\lambda^* \in \boldsymbol{R}^{\mathrm{m}}_+$ such that

$$-f'(x^*) + \sum_{i=1}^{m} \lambda_i^* g_i{}'(x^*) \leqq 0,$$

and the conclusion follows by multiplying this by -1. ∎

9.3 Constraint Qualifications

9.3.1 Definition *Let $f, g_1, \ldots, g_m \colon \boldsymbol{R}^{\mathrm{n}}_+ \to \boldsymbol{R}$. Let*

$$C = \{x \in \boldsymbol{R}^{\mathrm{n}} : x \geqq 0,\ g_i(x) \geqslant 0,\ i = 1, \ldots, m\}.$$

denote the ***constraint set****. Consider a point $x^* \in C$ and let*

$$B = \{i : g_i(x^*) = 0\} \text{ and } Z = \{j : x_j^* = 0\},$$

index the set of binding functional constraints and the set of binding nonnegativity constraints at x^*. The point x^* satisfies the ***Kuhn–Tucker Constraint Qualification*** *if $f, g_1, \ldots, g_m$ are differentiable at x^*, and for every $v \in \boldsymbol{R}^{\mathrm{n}}$ satisfying*

$$\begin{aligned} v_j = v \cdot e^j &\geqslant 0 \quad j \in Z, \\ v \cdot g_i{}'(x^*) &\geqslant 0 \quad i \in B, \end{aligned}$$

there is a continuous curve $\xi \colon [0, \varepsilon) \to \boldsymbol{R}^{\mathrm{n}}$ satisfying

$$\begin{aligned} \xi(0) &= x^*, \\ \xi(t) &\in C \qquad \text{for all } t \in [0, \varepsilon), \\ D\xi(0) &= v, \end{aligned}$$

Consistent notation?

where $D\xi(0)$ is the one-sided directional derivative at 0.

This condition is actually a little weaker than Kuhn and Tucker's condition. They assumed that the functions $f, g_1, \ldots, g_m$ were differentiable everywhere and required ξ to be differentiable everywhere. You can see that it may be difficult to verify it in practice.

To better understand the hypotheses of the theorem, let's look at a classic example of its failure (cf. Kuhn and Tucker [7]).

9.3.2 Example (Failure of the Kuhn–Tucker Constraint Qualification) Let $f \colon \boldsymbol{R}^2 \to \boldsymbol{R}$ via $f(x, y) = x$ and $g \colon \boldsymbol{R}^2 \to \boldsymbol{R}$ via $g(x, y) = (1 - x)^3 - y$. The curve $g = 0$ is shown in Figure 9.1, and the constraint set in Figure 9.2.

Clearly $(x^*, y^*) = (1, 0)$ maximizes f subject to $(x, y) \geqq 0$ and $g \geqslant 0$. At this point we have $g'(1, 0) = (0, -1)$ and $f' = (1, 0)$ everywhere. Note that no λ (nonnegative or not) satisfies

$$(1, 0) + \lambda(0, -1) \leqq (0, 0).$$

Fortunately for the theorem, the Constraint Qualification fails at $(1, 0)$. To see this, note that the constraint $g \geqslant 0$ binds, that is $g(1, 0) = 0$ and the second coordinate of (x^*, y^*) is zero. Suppose $v = (v_x, v_y)$ satisfies

$$v \cdot g'(1, 0) = v \cdot (0, -1) = -v_y \leqslant 0 \qquad \text{and} \qquad v \cdot e^2 = v_y \geqslant 0,$$

that is, $v_y = 0$. For instance, take $v = (1, 0)$. The constraint qualification requires that there is a path starting at $(1, 0)$ in the direction $(1, 0)$ that stays in the constraint set. Clearly no such path exists, so the constraint qualification fails.

□

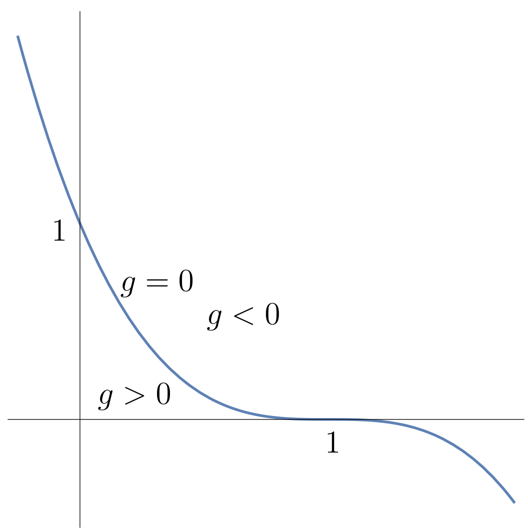

Figure 9.1. The function $g(x, y) = (1 - x)^3 - y$.

The next result, which may be found in Arrow, Hurwicz, and Uzawa [2, Corollaries 1, 4, 6, pp. 183–184], provides a tractable sufficient condition for the KTCQ.

9.3.3 Theorem (Constraint Qualifications) *In Theorem 9.1.1, the KTCQ may be replaced by any of the conditions below.*

1. *Each g_i is convex. (This includes the case where each is linear.)*
2. *Each g_i is concave and there exists some $\hat{x} \gg 0$ for which each $g_i(\hat{x}) > 0$.*
3. *The set $\{e^j : j \in Z\} \cup \{{g_i}'(x^*) : i \in B\}$ is linearly independent.*

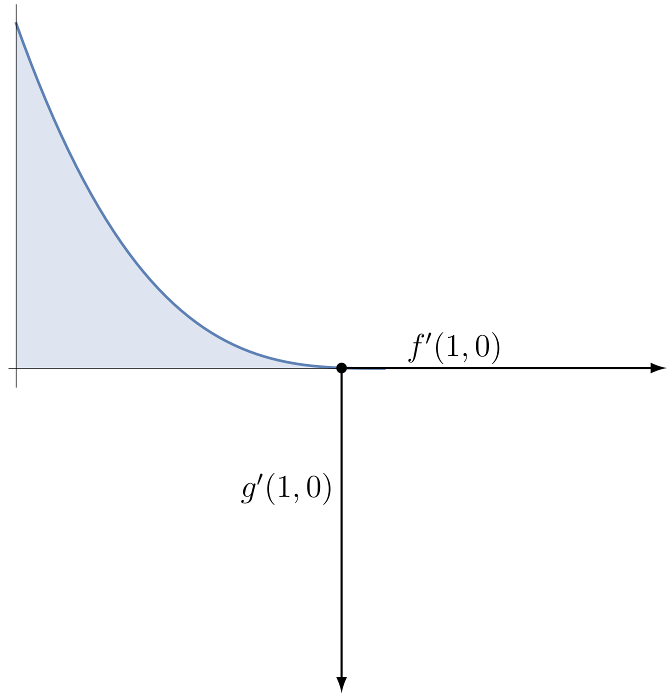

Figure 9.2. This constraint set violates the Constraint Qualification.

9.4 Quasisaddlepoint Theorem

The following theorem and its proof may be found in Arrow and Enthoven [1].

9.4.1 Theorem (Arrow–Enthoven) *Let $f, g_1, \ldots, g_m \colon \boldsymbol{R}^{\mathrm{n}}_{+} \to \boldsymbol{R}$ be differentiable and quasiconcave. Suppose $x^* \in \boldsymbol{R}^{\mathrm{n}}_{+}$ and $\lambda^* \in \boldsymbol{R}^{\mathrm{m}}$ satisfy the constraints $g(x^*) \geqq 0$ and $x^* \geqq 0$ and the Kuhn–Tucker–Lagrange first order conditions:*

$$f'(x^*) + \textstyle\sum_{j=1}^{m} \lambda_i^* {g_j}'(x^*) \leqq 0$$

$$x^* \cdot \Big(f'(x^*) + \textstyle\sum_{j=1}^{m} \lambda_i^* {g_j}'(x^*)\Big) = 0$$

$$\lambda^* \geqq 0$$

$$\lambda^* \cdot g(x^*) = 0.$$

Say that a variable x_j is ***relevant*** *if it may take on a strictly positive value in the constraint set. That is, if there exists some $\hat{x} \geqq 0$ (which may depnd on j) satisfying $\hat{x}_j > 0$ and $g(\hat{x}) \geqq 0$.*

Suppose at least one of the following conditions is satisfied:

1. $\dfrac{\partial f(x^*)}{\partial x_j} \neq 0$ *for some relevant variable x_j.*

2. *$f'(x^*) \neq 0$ and f is twice differentiable in a neighborhood of x^*.*

3. *f is concave.*

Then x^* maximizes $f(x)$ subject to the constraints $g(x) \geqq 0$ and $x \geqq 0$.

9.5 Cost Function for Linear Production Function

With this constant returns to scale production function, all inputs are perfect substitutes for each other (provided units are chosen properly).

$$y = \alpha_1 x_1 + \cdots + \alpha_n x_n,$$

where each $\alpha_i > 0$, $i = 1, \ldots, n$.

The Lagrangean for the cost minimization problem is

$$\sum_{i=1}^{n} w_i x_i - \lambda \left(\sum_{i=1}^{n} \alpha_i x_i - y \right)$$

and the naïve first order conditions are

$$\frac{\partial L}{\partial x_i} = w_i - \lambda \alpha_i = 0 \qquad i = 1, \ldots, n,$$

which taken at face value imply $\dfrac{w_1}{\alpha_1} = \cdots = \dfrac{w_n}{\alpha_n}$, which is unlikely since these are all exogenous. This is a red flag that signals that the nonnegativity constraints are binding and that you need to examine the Kuhn–Tucker first order conditions. They are

$$w_i - \lambda \alpha_i \geqslant 0 \qquad i = 1, \ldots, n,$$

and

$$x_i > 0 \implies w_i - \lambda \alpha_i = 0 \qquad \text{and} \qquad w_i - \lambda \alpha_i > 0 \implies x_i = 0.$$

In addition, $\lambda \geqslant 0$ and $\lambda \left(\sum_{i=1}^{n} \alpha_i x_i - y \right) = 0$.

Thus

$$\frac{w_i}{\alpha_i} \geqslant \lambda \qquad i = 1, \ldots, n.$$

The question is, can we have strict inequality for each i? The answer is no, as that would imply $x_i = 0$ for each i and the output would be zero, not $y > 0$. So the solution must satisfy

$$\hat{\lambda} = \min_i \frac{w_i}{\alpha_i}.$$

Let i^* satisfy $\hat{\lambda} = \frac{w_{i^*}}{\alpha_{i^*}}$. That is, i^* is a factor that maximizes "bang per buck." Then the conditional factor demand given by:

$$\hat{x}_i = \begin{cases} \dfrac{y}{\alpha_i}, & i = i^* \\ 0, & i \neq i^* \end{cases}$$

minimizes cost, and the cost function is

$$c(y,w) = y \cdot \min\left\{\frac{w_1}{\alpha_1},\cdots,\frac{w_n}{\alpha_n}\right\}.$$

This is the cost function even if i^* is not unique, but when there is more than one such i^*, the conditional factor demand is no longer a unique input vector, but rather a set of cost minimizing input vectors. In fact, the set of cost minimizing input vectors is the convex set:

$$\operatorname{co}\left\{\frac{y}{\alpha_i}e^i : \frac{w_i}{\alpha_i} = \hat\lambda = \min_j \frac{w_j}{\alpha_j}\right\}.$$

Note that even though the production function is very smooth, the cost function fails to be differentiable (for $n \geqslant 2$). This is to be expected since the bordered Hessian of the production function is given by

$$\begin{bmatrix} f_{11} & \dots & f_{1n} & f_1 \\ \vdots & & \vdots & \vdots \\ \vdots & & \vdots & \vdots \\ f_{n1} & \dots & f_{nn} & f_n \\ f_1 & \dots & f_n & 0 \end{bmatrix} = \begin{bmatrix} 0 & \dots & 0 & \alpha_1 \\ \vdots & & \vdots & \vdots \\ \vdots & & \vdots & \vdots \\ 0 & \dots & 0 & \alpha_n \\ \alpha_1 & \dots & \alpha_n & 0 \end{bmatrix},$$

which is singular for $n \geqslant 2$. (It has rank 2.)

9.6 Saddlepoint Theorem

9.6.1 Definition *Let $\varphi\colon X \times Y \to \boldsymbol{R}$. A point (x^*, y^*) in $X \times Y$ is a* ***saddlepoint of φ (over $X \times Y$)*** *if it satisfies*

$$\varphi(x,y^*) \leqslant \varphi(x^*,y^*) \leqslant \varphi(x^*,y) \quad \text{for all} \quad x \in X,\ y \in Y.$$

9.6.2 Definition *Given $f, g_1, \dots, g_m\colon C \to \boldsymbol{R}$, the associated* ***Lagrangean*** *$L\colon C \times \Lambda \to \boldsymbol{R}$ is defined by*

$$L(x,\lambda) = f(x) + \sum_{j=1}^{m} \lambda_j g_j(x) = f(x) + \lambda \cdot g(x),$$

where Λ is an appropriate subset of $\boldsymbol{R}^{\mathrm{m}}$. (Usually $\Lambda = \boldsymbol{R}^{\mathrm{m}}$ or $\boldsymbol{R}^{\mathrm{m}}_+$.) The components of λ are called ***Lagrange multipliers****.*

9.6.3 Theorem *Let $C \subset \boldsymbol{R}^{\mathrm{n}}$ be convex, and let $f, g_1, \dots, g_m\colon C \to \boldsymbol{R}$ be concave. Assume in addition that* ***Slater's Condition****,*

$$\exists \bar{x} \in C \quad g(\bar{x}) \gg 0, \tag{$\boldsymbol{S}$}$$

is satisfied. Then

x^* maximizes f subject to the constraints $g_j(x) \geqslant 0$, $j = 1, \dots, m$

if and only if

there exist real numbers $\lambda_1^*, \dots, \lambda_m^* \geqslant 0$ such that $x^*, \lambda_1^*, \dots, \lambda_m^*$ is a saddlepoint of the Lagrangean for $x \in C$, $\lambda \geqq 0$. That is,

$$L(x, \lambda^*) \leqslant L(x^*, \lambda^*) \leqslant L(x^*, \lambda) \quad x \in C, \quad \lambda \geqq 0, \tag{1}$$

where $L(x, \lambda) = f(x) + \lambda \cdot g(x)$.

Furthermore, in this case

$$\sum_{j=1}^{m} \lambda_j^* g_j(x^*) = 0. \tag{2}$$

In other words, for a concave programming problem, the optimal x^* maximizes the Lagrangean $L(\cdot, \lambda^*)$. The rôle of the Lagrange multipliers is to provide conversion factors or prices to convert a constrained maximization problem to an unconstrained maximization problem.

The next example shows what can go wrong when Slater's Condition fails.

9.6.4 Example In this example, due to Slater [8], $C = \boldsymbol{R}$, $f(x) = x$, and $g(x) = -(1-x)^2$. Note that Slater's Condition fails because $g \leqslant 0$. The constraint set $[g \geqslant 0]$ contains only 1. Therefore f attains a constrained maximum at 1. There is however no saddlepoint at all of the Lagrangean

$$L(x, \lambda) = x - \lambda(1-x)^2 = -\lambda + (1 + 2\lambda)x - \lambda x^2.$$

To see this, observe the first order condition for a maximum in x is $\frac{\partial L}{\partial x} = 0$, or $1 + 2\lambda - 2\lambda x = 0$, which implies $x > 1$ since $\lambda \geqslant 0$. But for $x > 1$, $\frac{\partial L}{\partial \lambda} = -(1-x)^2 < 0$, so no minimum with respect to λ exists. □

9.6.1 The rôle of Slater's Condition

In this section we present a geometric argument that illuminates the rôle of Slater's Condition in the saddlepoint theorem. Let us consider the argument underlying its proof. In the framework of Theorem 9.6.3, define the function $h\colon C \to \boldsymbol{R}^{m+1}$ by

$$h(x) = \Big(g_1(x), \dots, g_m(x), f(x) - f(x^*)\Big)$$

and set

$$H = \{h(x) : x \in C\} \quad \text{and} \quad \hat{H} = \Big\{y \in \boldsymbol{R}^{m+1} : \exists x \in C \quad y \leqq h(x)\Big\}.$$

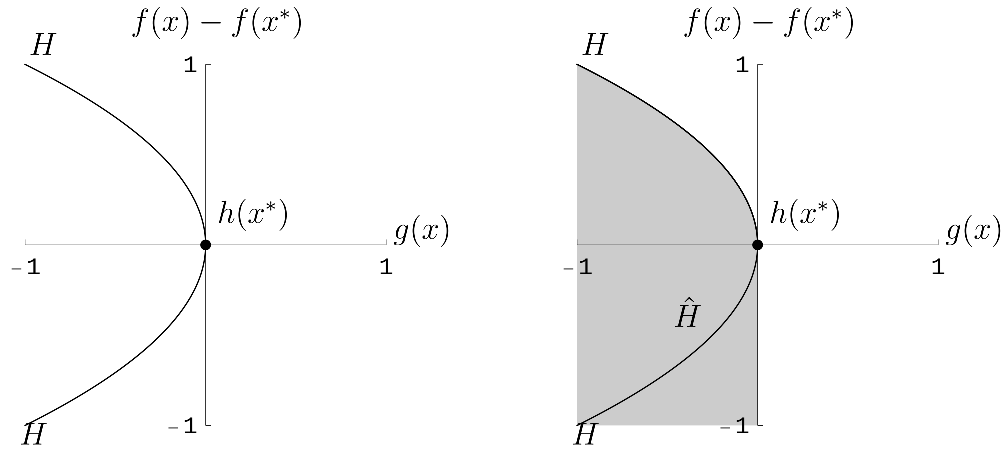

Figure 9.3. The sets H and $\hat{H}$ for Slater's example.

Then $\hat{H}$ is a convex set bounded in part by H. Figure 9.3 depicts the sets H and $\hat{H}$ for Slater's example 9.6.4, where $f(x)-f(x^*)$ is plotted on the vertical axis and $g(x)$ is plotted on the horizontal axis. Now if x^* maximizes f over the convex set C subject to the constraints $g_j(x) \geqslant 0$, $j = 1,\ldots,m$, then $h(x^*)$ has the largest vertical coordinate among all the points in H whose horizontal coordinates are nonnegative.

The semipositive $m+1$-vector $\hat{\lambda}^* = (\lambda_1^*,\ldots,\lambda_m^*,\mu^*)$ from Theorem 9.6.3 is obtained by separating the convex set $\hat{H}$ and $\boldsymbol{R}^{m+1}_{++}$. It has the property that

$$\hat{\lambda}^* \cdot h(x) \leqslant \hat{\lambda}^* h(x^*)$$

for all $x \in C$. That is, the vector $\hat{\lambda}^*$ defines a hyperplane through $h(x^*)$ such that the entire set $\hat{H}$ lies in one half-space. It is clear in the case of Slater's example that the hyperplane is a vertical line, since it must be tangent to H at $h(x^*) = (0,0)$. The fact that the hyperplane is vertical means that μ^* (the multiplier on f) must be zero.

If there is a non-vertical hyperplane through $h(x^*)$, then μ^* is nonzero, so we can divide by it and obtain a full saddlepoint of the true Lagrangean. This is where Slater's condition comes in.

In the one dimensional, one constraint case, Slater's Condition reduces to the existence of $\bar{x}$ satisfying $g(\bar{x}) > 0$. This rules out having a vertical supporting line through x^*. To see this, note that the vertical component of $h(x^*)$ is $f(x^*) - f(x^*) = 0$. If $g(x^*) = 0$, then the vertical line through $h(x^*)$ is simply the vertical axis, which cannot be, since $h(\bar{x})$ lies to the right of the axis. If $g(x^*) > 0$, then $\hat{H}$ includes every point below $h(x^*)$, so the only line separating $\hat{H}$ and $\boldsymbol{R}^2_{++}$ is horizontal, not vertical. See Figure 9.4.

In Figure 9.4, the shaded area is included in $\hat{H}$. For instance, let $C = (-\infty, 0]$, $f(x) = x$, and $g(x) = x + 1$. Then the set $\hat{H}$ is just $\{y \in \boldsymbol{R}^2 : y \leqq (0,1)\}$.

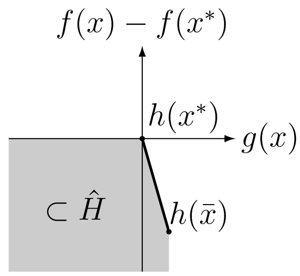

Figure 9.4. Slater's condition guarantees a non-vertical supporting line.

If f and the g_js are linear, then Slater's Condition is not needed to guarantee a non-vertical supporting line. Intuitively, the reason for this is that for the linear case, the set $\hat{H}$ is polyhedral, so even if $g(x^*) = 0$, there is still a non-vertical line separating $\hat{H}$ and $\boldsymbol{R}^{\mathrm{m}}_{++}$. The proof of this fact relies on results about linear inequalities. It is subtle because Slater's condition rules out a vertical supporting line. In the linear case, there may be a vertical supporting line, but if there is, there is also a non-vertical supporting line that yields a Lagrangean saddlepoint. As a case in point, consider $C = (-\infty, 0]$, $f(x) = x$, and $g(x) = x$. Then the set $\hat{H}$ is just $\{y \in \boldsymbol{R}^2 : y \leqq 0\}$, which is separated from $\boldsymbol{R}^2_{++}$ by every semipositive vector.

References

[1] K. J. Arrow and A. C. Enthoven. 1961. Quasi-concave programming. *Econometrica* 29(4):779–800. http://www.jstor.org/stable/1911819

[2] K. J. Arrow, L. Hurwicz, and H. Uzawa. 1961. Constraint qualifications in maximization problems. *Naval Research Logistics Quarterly* 8(2):175–191. DOI: 10.1002/nav.3800080206

[3] F. John. 1948. Extremum problems with inequalities as subsidiary conditions. In K. O. Friedrichs, O. E. Neugebauer, and J. J. Stoker, eds., *Studies and Essays: Courant Anniversary Volume*, pages 187–204. New York: Interscience. The volume is also subtitled: "Presented to R. Courant on his 60th Birthday, January 8, 1948.".

[4] W. Karush. 1939. Minima of functions of several variables with inequalities as side conditions. Master's thesis, Department of Mathematics, University of Chicago.

[5] H. W. Kuhn. 1976. Nonlinear programming: A historical view. *SIAM–AMS Proceedings* IX:1–26. Reprinted as [6].

[6] ——— . 1982. Nonlinear programming: A historical view. *ACM SIGMAP Bulletin* 31:6–18. Reprinted from SIAM–AMS Proceedings, volume IX, pp. 1–26.

[7] H. W. Kuhn and A. W. Tucker. 1951. Nonlinear programming. In J. Neyman, ed., *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability II, Part I*, pages 481–492. Berkeley: University of California Press. Reprinted in [**?**, Chapter 1, pp. 3–14].
http://projecteuclid.org/euclid.bsmsp/1200500249

[8] M. L. Slater. 1950. Lagrange multipliers revisited: A contribution to non-linear programming. Discussion Paper Math. 403, Cowles Commission. Reissued as Cowles Foundation Discussion Paper #80 in 1959.
http://cowles.econ.yale.edu/P/cd/d00b/d0080.pdf