OPTIMIZATION THEORY IN A NUTSHELL

Daniel McFadden, © 1990, 2003

UNCONSTRAINED OPTIMIZATION

1. Consider the problem of maximizing a function $f:\mathbb{R}^n \to \mathbb{R}$ within a set $\mathbf{A} \subseteq \mathbb{R}^n$. Typically, **A** might be all of $\mathbb{R}^n$, or the non-negative orthant, or a half-space. (More complicated sets of feasible points will be treated explicitly as constraints on the maximization.) In general, there is no guarantee that the function f will achieve a maximum on **A**, but this is guaranteed if **A** is compact (i.e., closed and bounded) and f is continuous. Suppose that f does have a maximum on **A**, and denote it x^o. There is no guarantee that this maximum is unique; however, when the set **A** is convex and the function f is strictly concave on **A** (i.e., $x,y \in \mathbf{A}$, $x \neq y$, $0 < \theta < 1$ implies $f(\theta x+(1-\theta)y) > \theta f(x) + (1-\theta)f(y)$), a maximand x^o is necessarily unique.

A vector $y \in \mathbb{R}^n$ points into **A** (from a vector $x^0 \in \mathbf{A}$) if for all sufficiently small positive scalars θ, $x^0 + \theta y \in \mathbf{A}$.

2. Assume that f is twice continuously differentiable, and that x^o achieves the maximum of f on **A**. Then, a Taylor's expansion gives

(1) $$f(x^o) \geq f(x^o+\theta y) = f(x^o) + \theta f_x(x^o)\cdot y + (\theta^2/2)y' f_{xx}(x^o)y + R(\theta^2)$$

for all y that point into **A** and sufficiently small scalars $\theta > 0$, where $R(\varepsilon)$ is a remainder satisfying $\lim_{\varepsilon\to 0} R(\varepsilon)/\varepsilon = 0$. Taking θ sufficiently small so that squared and remainder terms are negligible, this inequality implies

(2) $$0 \geq f_x(x^o)\cdot y \quad \text{for all directions y that point into } \mathbf{A}.$$

If x^o is in the interior of **A**, then (1) holds for all y and -y, implying the first-order-condition (FOC)

(3) $$0 = f_x(x^o).$$

If **A** is the non-negative orthant, then a component of $f_x(x^o)$ is zero if the corresponding component of x^o is positive, and is non-positive if the corresponding component of x^o is zero: $f_x(x^o) \leq 0$ and $x^o\cdot f_x(x^o) = 0$.

For all $y \neq 0$ that point into **A** and have $f_x(x^o) \cdot y = 0$, (1) implies (taking θ small) that

(4) $\quad 0 \geq y' f_{xx}(x^o) y.$

This is then a necessary second-order condition (SOC) for a maximum; it says that f_{xx} is negative semi-definite subject to constraint. If the inequality in (4) is strict, then this second-order condition is sufficient for x^o to be a unique maximum of f within **A** and some open ball containing x^o. If x^o is interior to **A**, then (4) must hold for all y (i.e., f_{xx} is negative semidefinite).

3. In the case that f is to be minimized over **A**, the inequality in (1) is reversed for a minimand x^o implying the FOC

(5) $\quad 0 \leq f_x(x^o) \cdot y \quad$ for all y pointing into **A**,

or $f_x(x^o) = 0$ in the case of x^o interior to **A**. The SOC becomes

(6) $\quad 0 \leq y' f_{xx}(x^o) y$ for all y pointing into **A** such that $f_x(x^o) \cdot y = 0$.

Again, the SOC for an interior minimum is that $y' f_{xx}(x^o) y \leq 0$ for all y.

CONSTRAINED OPTIMIZATION

4. Consider the problem of maximizing $f: \mathbb{R}^n \rightarrow \mathbb{R}$ subject to constraints

$$\begin{aligned} & g^i(x) - b_i \leq 0 \quad \text{for } i = 1,\ldots, k \\ (7) \quad & h^j(x) - c_j = 0 \quad \text{for } j = 1,\ldots, m \end{aligned}$$

where $k+m < n$. One may have either $k = 0$ yielding a classical equality constrained optimization problem, $m = 0$ yielding a mathematical programming problem, or both positive, so there is a mix of equality and inequality constraints. If the domain of x is restricted to a set **A**, as in paragraph 1, this is incorporated into the constraints rather than being handled directly. The system of constraints (7) will often be written in vector notation as

(8) $\quad g(x) - b \leq 0 \quad$ and $\quad h(x) - c = 0$

k×l k×l k×l m×l m×l m×l

The set of x′s satisfying (8) is termed the feasible set. The functions f, g, and h are all assumed to be twice continuously differentiable.

The Lagrangian method is useful for solving this constrained optimization problem. Define a non-negative k-vector p and an unsigned m-vector q of undetermined Lagrange multipliers. These are also called shadow prices as the construction below will justify. Define the Lagrangian

(9) $L(x,p,q) = f(x) - p\cdot[g(x) - b] - q\cdot[h(x) - c];$

note that the constraints enter the expression so that when (8) holds, the terms in brackets are non-positive. A point (x^o,p^o,q^o) is said to be a (global) Lagrangian Critical Point (LCP) if for all $x \in \mathbb{R}^n$, $q \in \mathbb{R}^m$ and non negative $p \in \mathbb{R}^k$,

(10) $L(x,p^o,q^o) \le L(x^o,p^o,q^o) \le L(x^o,p,q).$

The vector (x^o,p^o,q^o) is said to be a local LCP if

(11) $L_x(x^o,p^o,q^o) = 0$, $L_q(x^o,p^o,q^o) = 0$,

$L_p(x^o,p^o,q^o) \le 0$ and $p\cdot L_p(x^o,p^o,q^o) = 0$, and

$z'\, L_{xx}(x^o p^o,q^o)z \le 0$ if z satisfies $\mathbf{g}_x^*(x^o)z = 0$ and $h_x(x^o)z = 0$,

where $g^*(x) - b_* = 0$ denotes the set of inequality constraints that are binding (i.e., hold with equality) at x^o, and $g^\#(x) - b_\# < 0$ denotes the remaining non-binding constraints at x^o. The LCP conditions are also called the Kuhn-Tucker conditions.

7. The relation of the Lagrangian to the original constrained optimization problem is this: If (x^o,p^o,q^o) is a global LCP, then x^o is a global maximum of f(x) subject to the constraints (8); if it is a local LCP, then x^o is a local maximum. In the other direction, if x^o is a local maximum of f(x) subject to (8), and a technical condition called a constraint qualification holds, then there exist (p^o,q^o) such that (x^o,p^o,q^o) define a local LCP. Globally, if the functions g(x) are convex, the functions h(x) are linear, the function f(x) is concave, x^o solves the constrained maximization problem, and the constraint qualification is satisfied, then there exist (p^o,q^o) such that (x^o,p^o,q^o) defines a global LCP. These last conditions are sufficient for a global LCP, but they are not necessary: there are problems where the concavity or convexity assumptions do not hold, but it is still possible to find Lagrangian multipliers to define a LCP.

The constraint qualification at x^o is that the matrix whose rows are the partial derivatives of the binding constraints be of full rank; i.e., if there are k_1 binding constraints $g^*(x^o) - b_* = 0$, plus the equality constraints $h(x^o) - c = 0$, and partial derivatives are denoted $g_i^* = \partial g^*/\partial x_i$ and $h_i = \partial h/\partial x_i$, then the $(k_1 + m) \times n$ matrix

(12) $$B = \begin{bmatrix} g_1^* \; g_2^* \cdots g_n^* \\ h_1 \; h_2 \cdots h_n \end{bmatrix}$$

has rank $k_1 + m$.

8. First, the result that a global LCP solves the constrained maximization problem will be explained. Suppose (x^o,p^o,q^o) is a global LCP. Writing out the definition of the Lagrangian in (10), one obtains from the second inequality that

(13) $$(p - p^o)\cdot[g(x^o) - b] + (q - q^o)\cdot[h(x^o) - c] \le 0.$$

But this can hold for all q only if $h(x^o) - c = 0$, and can hold for all non-negative p only if $g(x^o) - b \le 0$. Then, x^o satisfies the constraints (8). Taking $p = 0$ in (13), and using the inequalities just established, one has

(14) $$p^o\cdot[g(x^o) - b] = 0;$$

these are called the <u>complementary slackness</u> conditions, and state that if a constraint is not binding, then its Lagrangian multiplier is zero. Putting these results together, the Lagrangian at the LCP satisfies

(15) $$L(x^o, p^o, q^o) = f(x^o) - p^o\cdot[g(x^o) - b] - q^o\cdot[h(x^o) - c] = f(x^o).$$

Now write out the first inequality in (10), using (15), to get

(16) $$f(x) - p^o\cdot[g(x)- b] - q^o\cdot[h(x) - c] \le f(x^o).$$

Then, any x in the feasible set has $g(x) - b \le 0$ and $h(x) - c = 0$, implying that $f(x) \le f(x^o)$. This establishes that x^o is a solution to the optimization problem.

To show that a local LCP (x^o, p^o, q^o) yields a local maximum to the constrained optimization problem, first note that $0 = L_q(x^o, p^o, q^o) = h(x^o) - c$, $0 \ge L_p(x^o, p^o, q^o) = g(x^o) - b$, and $0 = p^o\cdot L_p(x^o,p^o,q^o) = p^o\,[g(x^o) - b]$, imply that x^o is feasible, and that $L(x^o,p^o,q^o) = f(x^o)$. A

Taylor's expansion of $L(x^o+\theta z,p^o,q^o)$ yields

(17) $L(x^o+\theta z,p^o,q^o) = L(x^o,p^o,q^o) + \theta L_x(x^o,p^o,q^o)\cdot z + (\theta^2/2)z'L_{xx}(x^o,p^o,q^o)z + R(\theta^3)$;

the $R(\theta^3)$ term is a residual. Similar expansions could be made separately for the objective function f and the binding constraints g^* and h. Note that the non-binding constraints stay non-binding for θ small. Writing (17) out, using $L(x^o,p^o,q^o) = f(x^o)$ and $L_x(x^o,p^o,q^o) = 0$, yields

(18) $f(x^o+\theta z) - p^o[g(x^o+\theta z) - b] - q^o[h(x^o+\theta z) - c] = f(x^o) + (\theta^2/2)z'L_{xx}(x^o, p^o, q^o)z + R(\theta^3)$

A point $x^o + \theta z$ satisfying the constraints, with θ small, must (by a Taylor's expansion of these constraints) satisfy $\mathbf{g}_x^*(x^o)z = 0$ and $h_x(x^o)z = 0$. Then, the negative semidefiniteness of L_{xx} subject to these constraints implies in (18) that

(19) $f(x^o+\theta z) \le f(x^o) + (\theta^2/2)z'L_{xx}(x^o,p^o,q^o)z + R(\theta^3) \le f(x^o) + R(\theta^3)$.

If L_{xx} is negative definite subject to these constraints then this SOC is sufficient for x^o to be a local maximum.

9. The result that a local solution to the constrained optimization problem corresponds to a local LCP, provided a constraint qualification holds, will now be explained. Suppose x^o is a local solution to the constrained optimization problem, and suppose that at x^o the binding constraints are denoted $g^*(x^o) - b_* = 0$ and $h(x^o) - c = 0$. The fact that x^o is a local solution to the optimization problem implies that $0 \ge f_x(x^o)\cdot z$ for all z satisfying $\mathbf{g}_x^*(x^o)\cdot z = 0$ and $h_x(x^o)z = 0$, or from (12), $\mathbf{B}z = 0$. Assume the constraint qualification, so that $\mathbf{B}$ has rank $k_1 + m$. Then, all the columns of the matrix $\mathbf{I} - \mathbf{B}'(\mathbf{BB}')^{-1}\mathbf{B}$, and their negatives, are potential z vectors satisfying $\mathbf{B}z = 0$, and hence must satisfy $0 = [\mathbf{I} - \mathbf{B}'(\mathbf{BB}')^{-1}\mathbf{B}]f_x(x^o)$. Define the Lagrange multipliers

(20) $\begin{bmatrix}\mathbf{p}^{*o}\\ \mathbf{q}^o\end{bmatrix} = (\mathbf{BB}')^{-1}\mathbf{B}f_x(x^o)$,

and $p^{\#o} = 0$ for the non-binding constraints $g^{\#}(x^o) - b_{\#} < 0$. Define the Lagrangian as in equation (9). Then,

(21) $\quad L_x(x^o,p^o,q^o) = f_x(x^o) - \mathbf{g}_x^*(x^o)'p^{*o} - h_x(x^o)'q^o$

$$= [\mathbf{I} - \mathbf{B}'(\mathbf{B}\mathbf{B}')^{-1}\mathbf{B}]f_x(x^o) = 0.$$

The construction guarantees that $L_q(x^o,p^o,q^o) = 0$, $L_p(x^o,p^o,q^o) \le 0$, and $p^o \cdot L_p(x^o,p^o,q^o) = 0$. Finally, the Taylor's expansion (17) of the Lagrangian, written out in the form (18), establishes that $z'L_{xx}(x^o,p^o,q^o)z \le 0$ for all z satisfying $\mathbf{g}_x^*(x^o) \cdot z = 0$ and $h_x(x^o) \cdot z = 0$. Therefore, the constrained maximum corresponds to a local LCP satisfying (11).

10. Sufficient conditions for a global solution x^o to the constrained optimization problem to correspond to a global LCP are that the objective function f is concave, the inequality constraint functions g(x) are convex, the equality constraint functions h(x) are linear, and a local constraint qualification, that the array of derivatives of the binding constraints be of maximum rank, holds. Since x^o is also a local solution to the optimization problem, it corresponds to a local LCP (x^o,p^o,q^o), with the Lagrangian multipliers given by (20). The curvature assumptions on f, g, and h imply

(22) $\quad f(x^o + z) \le f(x^o) + f_x(x^o)\ z$

$$g(x^o + z) \ge g(x^o) + g_x(x^o)\ z$$

$$h(x^o + z) = h(x^o) + h_x(x^o)\ z.$$

Substitute these into the definition of the Lagrangian to obtain

(23) $\quad L(x^o + z,p^o,q^o) \le f(x^o) + [f_x(x^o) - p^o \cdot g_x(x^o) - q^o \cdot h_x(x^o)]z.$

From (20), the term in brackets is zero. Then $L(x,p^o,q^o) \le L(x^o,p^o,q^o)$. The properties of p^o and q^o from the argument for a local LCP imply the inequality $L(x^o,p^o,q^o) \le L(x^o,p,q)$. Hence, (x^o,p^o,q^o) is a global LCP.

11. Lagrangian multipliers have the following interpretation: $\mathbf{p_i^o}$ equals the rate of increase in the optimized objective function for each unit increase in the constant b_i in the constraint $g^i(x^o) - b_i \le 0$, and $\mathbf{q_i^o}$ equals the rate of change in the objective function for each unit increase of c_i in the equality constraint $h^i(x) - c_i = 0$. In other words, a Lagrangian multiplier measures the value of relaxing a corresponding constraint by one unit. For this reason, Lagrangian multipliers are sometimes called shadow prices.

To explain this result, recall that the value of the Lagrangian at a LCP equals the value of the objective function. These values will in general depend on the levels of the constraints; i.e., b and c. Make this dependence explicit by writing the Lagrangian as L(x,p,q,b,c), a solution to the constrained optimization problem as $x^o = X(b,c)$, and the associated Lagrangian multipliers as $p^o = P(b,c)$ and $q^o = Q(b,c)$. Consider a change $\theta \triangle b$ in b, where θ is a small scalar. Suppose there is a global LCP when the constraint constant is b, and another when the constraint constant is b + $\theta \triangle b$ Then, one has the inequality

(24) $$f(X(b,c)) \equiv L(X(b,c),P(b,c),Q(b,c)) \geq L(X(b+\theta \triangle b,c),P(b,c),Q(b,c))$$
$$\equiv f(X(b+\theta \triangle b)) - P(b,c)\,[g(X(b+\theta \triangle b)) - b] \geq f(X(b+\theta \triangle b)) - P(b,c) \cdot [b + \theta \triangle b - b];$$

the first inequality comes from the LCP property, the identity comes from writing out the definition of L, and the last inequality comes from the constraint condition $g(X(b+\theta \triangle b)) \leq b + \theta \triangle b$. Similarly,

(25) $$f(X(b+\theta \triangle b,c)) \equiv L(X(b+\theta \triangle b,c), P(b+\theta \triangle b\ c), Q(b + \theta \triangle b,c)$$
$$\geq L(X(b,c),P(b+\theta \triangle b,c),Q(b+\theta \triangle b,c)$$
$$\equiv f(X(b)) - P(b+\theta \triangle b, c) \cdot [g(X(b) - b - \theta \triangle b] \geq f(X(b)) - P(b+\theta \triangle b, c)\,[b - b - \theta \triangle b]$$

The inequalities (24) and (25) imply for $\theta > 0$ that

(26) $$P(b+\theta \triangle b, c) \triangle b \leq \frac{f(X(b + \Delta b, c)) - f(X(b,c))}{\theta} \leq P(b, c) \triangle b.$$

If the Lagrangian multiplier is continuous in b, then taking $\theta \to 0$ in this inequality implies that $\partial f(X(b,c))/\partial b$ exists, and that

(27) $$\partial f(X(b,c))/\partial b = P(b,c).$$

A similar argument establishes that $\partial f(X(b,c))/\partial c = Q(b,c)$ whenever the multiplier is continuous in c.

Another way to establish the shadow price result is to assume differentiability. Start from the identity $L(X(b,c),P(b,c),Q(b,c),b,c) \equiv f(X(b,c))$. Differentiate this with respect to b, using the composite function rule, to obtain

(28) $$\partial f(X(b,c))/\partial b \equiv L_X(X(b,c),P(b,c),Q(b,c),b,c) \cdot X_b(b,c)$$
$$+ L_p(X(b,c),P(b,c),Q(b,c),b,c)\ P_b(b,c)$$
$$+ L_q(X(b,c),P(b,c),Q(b,c),b,c)\ Q_b(b,c)$$

$$+ L_b(X(b,c),P(b,c),Q(b,c),b,c)$$

The first three terms are zero, from the first-order conditions (11) for a local LCP, plus the fact that if $\partial L/\partial p_i < 0$, then $p_i = P^i (b+\theta\Delta b, c) = 0$ for small θ, so that $\partial P^i/\partial b$ 0. The last term in (26) equals P(b,c), again establishing (27). This argument is sometimes called the <u>envelope theorem</u>: if a function L(x,b) is maximized in x for each b at x = X(b), then the derivative of the maximized function L(X(b),b) with respect to b is the same as the partial derivative of L(x,b) with respect to b, with x fixed and then evaluated at X(b).

12. Constrained minimization rather than constrained maximization can be handled simply by maximizing the negative of the objective function. Then, all the previous results also apply to this case. The usual setup for the minimization problem is

(29) $$\min_x f(x) \qquad \text{subject to } g(x) - b \geq 0 \text{ and } h(x) - c = 0.$$

The associated Lagrangian is

(30) $$L(x,p,q,b,c) = f(x) - p\cdot[g(x) - b] - q\cdot[h(x) - c].$$

A global LCP (x^o,p^o,q^o) has $p^o \geq 0$ and satisfies

(31) $$L(x^o,p,q,b,c) \leq L(x^o,p^o,q^o,b,c) \leq L(x, p^o,q^o,b,c)$$

for all x, q, and $p \geq 0$. A local LCP satisfies

(32) $$L_x(x^o,p^o,q^o,b,c) = 0,\ L_q(x^o,p^o,q^o,b,c) = 0,$$
$$L_p(x^o,p^o,q^o,b,c) \leq 0 \text{ and } p\cdot L_p(x^o,p^o,q^o,b,c) = 0, \text{ and}$$
$$z'L_{xx}(x^o,p^o,q^o,b,c)\, z \geq 0 \text{ if z satisfies } g^*_x(x^o)z = 0 \text{ and } h_x(x^o)z = 0,$$

where g^* denotes the inequality constraints that are binding.