WENDNER: INTERMEDIATE MICROECONOMICS

Important Principles of Constrained Optimization (Classical Programming & some Hints for Nonlinear Programming)

1. Ingredients of the Classical Programming Problem

The problem of classical programming is that of choosing N choice variables so as to maximize the objective function $f(x)$ subject to $M < N$ *equality* constraints.

$$\max_x f(x) \quad \text{subject to} \quad g(x) - b = 0 \,, \tag{1}$$

where

$$g(x) = \begin{pmatrix} g_1(x_1,...,x_N) \\ g_2(x_1,...,x_N) \\ \vdots \\ g_M(x_1,...,x_N) \end{pmatrix} , \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{pmatrix} , \quad 0 = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} .$$

So, $g(x) - b$ is a vector valued function from $\mathbb{R}^N \to \mathbb{R}^M$. In contrast to the unconstrained optimization problem, here, the opportunity space is constrained to $X = \{x \in \mathbb{R}^N \,|\, g(x) = b\}$, where b is called the *constraint vector*, and b_i, $i = 1, ..., M$ are called constraint constants.

2. Theorem on Lagrange Multipliers

Problem (1) can be solved for by means of the Lagrange function:

$$\mathcal{L} = f(x) + \lambda_1(b_1 - g_1(x)) + ... + \lambda_M(b_M - g_M(x)) \,, \tag{2}$$

where $\lambda = (\lambda_1, ..., \lambda_N)$ are called the *Lagrange multipliers*. Every Lagrange multiplier λ_i measures the change in $f(x)$ – at the optimum – upon a marginal change in the $i-$th constraint constant. Let (x^*, λ^*) be the solution of problem (1). Then, $\lambda_i^* = \partial f(x^*)/\partial b_i$.

Query. Set up a simple utility maximization problem. What exactly is the associated Lagrange multiplier telling us?

Constraint qualification (for equality constraints). Notice that we generally want to have (i) $N > M$, (ii) constraints are linearly independent of each other (at an optimum)[1], and (iii) $f(x)$, $g(x)$ are twice continuously differentiable functions. In this case, we are having necessary and sufficient

[1] In this case, the rank of the Jacobian of $g(x^*)$ equals M.

conditions that parallel those of the unconstrained case (see my notes on unconstrained optimization).

Considering the Lagrange multipliers, the optimization problem involves N choice variables and M Lagrange multiplier variables.

Theorem 1 (Lagrange multipliers) *The following first order conditions are necessary for an optimum.*

$$\frac{\partial \mathcal{L}}{\partial x_i} = \frac{\partial f}{\partial x_i} - \sum_{m=1}^{M} \lambda_m \frac{\partial g_m}{\partial x_i} = 0 \quad (N \text{ conditions}) \tag{3}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = b - g(x) = 0 \quad (M \text{ conditions}). \tag{4}$$

So, in an optimum, there exists a vector of M Lagrange multipliers λ such that the gradient of $f(x)$ is a linear combination of the gradients of $g(x)$: $\partial f(x)/\partial x_i = \sum_{m=1}^{M} \lambda_m (\partial g_m/\partial x_i)$, $i = 1, ...N$.

A few remarks are in order. First, the first order conditions, given by Theorem 1, constitute necessary conditions. I.e., suppose (x^*, λ^*) is a solution to the optimization problem. Then, (x^*, λ^*) imply the first order conditions (3) and (4) (but not vice versa). Second, the value of the Lagrange multipliers is not constrained in any way (in contrast to the case with inequality constraints).[2] However, a constraint is not binding, at the optimum, if its value equals zero. Third, if $N = 2$, $M = 1$, a direct implication of Theorem 1 is that — at an optimum — the gradients of both the objective function and the constraints are pointing to the same direction and are proportional to each other (with λ being the proportionality factor). Fourth, as the constraints are *equality* constraints, the opportunity set is of dimension $N - M$![3] Fifth, $f(x^*) = \mathcal{L}(x^*, \lambda^*)$. Notice, if $M = 0$, Theorem 1 reduces to the theorem on first order conditions for the case without constraints.

3. Second Order Conditions

The necessary and sufficient second order conditions coincide with those of the unconstrained case — with $f(x)$ being replaced $\mathcal{L}(x, \lambda)$. The Hessian[4] of $\mathcal{L}(x, \lambda)$, Hb is called a *bordered* Hessian, as the constraints "border" the

[2]Remember, for the case of inequality constraints, the Kuhn-Tucker Conditions require $\lambda^* \geq 0$!

[3]Suppose, $N = 2, M = 1$. Then, the unconstrained domain would be $\mathbb{R}^2$. However, the constraint reduces the opportunity set to a curve in $\mathbb{R}^2$.

[4]Consider $y \in \mathbb{R}^K$. The gradient of a function $L(y)$ is the column vector of the K first partial derivatives of $L(.)$ with respect to y – so, the kth row of the gradient at y is $\partial L(y)/\partial y_k$, $k = 1, ..., K$. The Hessian of a function $L(y)$ is the Jacobian of

Hessian of the objective function. The Lagrange function is given by:
$\mathcal{L} = f(x_1,...,x_N)+\lambda_1\,g_1(x_1,...,x_N)+\lambda_2\,g_2(x_1,...,x_N)+\cdots+\lambda_M\,g_M(x_1,...,x_N)$:

$$Hb = \left(\begin{array}{ccc|ccc} \mathcal{L}_{\lambda_1\lambda_1} & \cdots & \mathcal{L}_{\lambda_1\lambda_m} & \mathcal{L}_{\lambda_1 x_1} & \cdots & \mathcal{L}_{\lambda_1 x_n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathcal{L}_{\lambda_m\lambda_1} & \cdots & \mathcal{L}_{\lambda_m\lambda_m} & \mathcal{L}_{\lambda_m x_1} & \cdots & \mathcal{L}_{\lambda_m x_n} \\ \hline \mathcal{L}_{x_1\lambda_1} & \cdots & \mathcal{L}_{x_1\lambda_m} & \mathcal{L}_{x_1 x_1} & \cdots & \mathcal{L}_{x_1 x_n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathcal{L}_{x_n\lambda_1} & \cdots & \mathcal{L}_{x_n\lambda_m} & \mathcal{L}_{x_n x_1} & \cdots & \mathcal{L}_{x_n x_n} \end{array}\right)$$

The four quadrants of this matrix are counted in the following way:

$$\left(\begin{array}{c|c} II & I \\ \hline III & IV \end{array}\right)$$

Certainly, quadrant IV represents the $N \times N$ (unbordered) Hessian matrix of the objective function, H, which we already encountered in course of unconstraint optimization. Next, quadrant II is an $M \times M$ nullmatrix. Quadrant I is the Jacobian of the system of constraint functions $g(x)$, which, by the constraint qualification is of (row) rank M. Denote this Jacobian by G. Then, quadrant III is given by the transpose of G, G^T:

$$Hb = \left(\begin{array}{c|c} 0 & G \\ \hline G^T & H \end{array}\right) .$$

Query. What is the dimension of the matrix Hb? Why is Hb a symmetric matrix.

Query. What exactly is a Jacobian matrix? What exactly is the Jacobian of $g(x)$?

Theorem 2 (Second order conditions) *For the equality-constrained maximization problem, a necessary condition for (x^*, λ^*) to represent a maximum is the negative semidefiniteness of Hb. A sufficient condition for (x^*, λ^*) to represent a maximum is the negative definiteness of Hb.*

In order to determine definiteness of Hb, we will see later, that we need to consider the leading principal minors of Hb. For the time being, we might gain some insight by considering the following case.

its gradient. I.e., the Hessian is a quadratic and, due to Young's Theorem, symmetric matrix of the second order partial derivatives. The kth row of the Hessian is then: $\partial^2 L(y)/(\partial y_k \partial y_1)$, $\partial^2 L(y)/(\partial y_k \partial y_2)$, ..., $\partial^2 L(y)/(\partial y_k \partial y_K)$.

Suppose, $g(x) - b = 0$ is a linear equality constraint. Then, the second order sufficient condition reduces to the requirement of $f(x)$ being a strictly quasiconcave function. Moreover, the latter requirement is implied by $f(x)$ being a strictly concave function. So, if $g(x) - b = 0$ is a linear system of equations, and $f(x)$ is strictly concave, then we know that the second order sufficient conditions are satisfied. Hence, the first order conditions (3), (4) are both necessary and sufficient for (x^*, λ^*) to represent a maximum.

4. Inequality Constraints: Nonlinear Programming (Sketch)

With inequality constraints, the world changes quite a bit. One reason is that the opportunity set's dimension is not reduced to $N - M$. In the nonlinear programming problem we are having $M + N$ inequality constraints:

$$g(x) - b \leq 0\,, \quad x \geq 0\,. \tag{5}$$

When a weak inequality constraint becomes binding as an equality, we face boundary solutions, for which the usual first order conditions do not need to hold. Moreover, $\mathcal{L}(x, \lambda) = f(x) + \lambda(b - g(x))$, i.e., in contrast to the classical programming problem, $\mathcal{L}(x, \lambda) \geq f(x)$!

Necessary first order conditions for an inequality-constrained optimization problem — in an optimum — are the following:

$$\frac{\partial \mathcal{L}(x^*, \lambda^*)}{\partial x} \leq 0\,, \quad \frac{\partial \mathcal{L}(x^*, \lambda^*)}{\partial \lambda} \geq 0 \tag{6}$$

$$\frac{\partial \mathcal{L}(x^*, \lambda^*)}{\partial x}\, x^* = 0\,, \quad \lambda^* \, \frac{\partial \mathcal{L}(x^*, \lambda^*)}{\partial \lambda} = 0 \tag{7}$$

$$x^* \geq 0\,, \quad \lambda^* \geq 0\,. \tag{8}$$

Now consider boundary conditions. These first order conditions imply the *complementary slackness conditions*:

$$\frac{\partial f(x^*)}{\partial x_n} - \lambda^* \, \frac{\partial g(x^*)}{\partial x_n} < 0 \text{ implies } x_n^* = 0,\ n = 1, ..., N \tag{9}$$

$$g_m(x^*) < b_m \text{ implies } \lambda_m^* = 0,\ m = 1, ..., M\,. \tag{10}$$

Please, take a day or two to think about these complementary slackness conditions!

You should be aware of the complementary slackness conditions, however, you will not need to deal with them a lot during the bakk program. If you proceed to the masters program, however, you will desperately need to inquire a bit more into the "Kuhn-Tucker Conditions".