Lecture #7

Taylor's Theorem

Theorem (Taylor's Theorem) 7.1: Let $f: U \to R$, where U is an open subset of R^N. Suppose that U contains the straight line segment from a to b, where $a \neq b$, and suppose that f is r times differentiable, where $r \geq 1$. Then there exists a point q on the line segment from a to b such that $q \neq a$, $q \neq b$, and

$$f(b) = f(a) + Df(a)(b-a) + \frac{1}{2}D^2f(a)(b-a, b-a) + \frac{1}{3!}D^3f(a)(b-a, b-a, b-a)$$
$$+ \dots + \frac{1}{(r-1)!}D^{r-1}f(a)(b-a, \dots, b-a) + \frac{1}{r!}D^rf(q)(b-a, \dots, b-a). \tag{7.1}$$

Proof: Let $F: [0, 1] \to R$ be defined by $F(t) = f(a + t(b-a))$. Then

$$\frac{dF(t)}{dt} = Df(a + t(b-a))(b-a)$$

$$\frac{d^2F(t)}{dt^2} = D^2f(a + t(b-a))(b-a, b-a)$$

$$\vdots$$

$$\frac{d^rF(t)}{dt^r} = D^rf(a + t(b-a))(b-a, \dots, b-a).$$

By the univariate Taylor's theorem 5.9 applied to F, there is an α such that $0 < \alpha < 1$ and

$$F(1) = F(0) + \frac{dF(0)}{dt} + \frac{1}{2}\frac{d^2F(0)}{dt^2} + \frac{1}{3!}\frac{d^3F(0)}{dt^3}$$
$$+ \dots + \frac{1}{(r-1)!}\frac{d^{r-1}F(0)}{dt^{r-1}} + \frac{1}{r!}\frac{d^rF(\alpha)}{dt^r}. \tag{7.2}$$

Let $q = a + \alpha(b-a)$ and substitute in equation 7.2 $f(a)$ for $F(0)$, $Df(a)(b-a)$ for $\frac{dF(0)}{dt}$, ..., $D^{r-1}f(a)(b-a, \dots, b-a)$ for $\frac{d^{r-1}F(0)}{dt^{r-1}}$, and $Df^r(q)(b-a, \dots, b-a)$ for $\frac{d^rF(\alpha)}{dt^r}$. The result is equation 7.1. ■

In order to understand the notation in Taylor's theorem, notice that

$$Df(a)(b-a) = \sum_{n=1}^{N} \frac{\partial f(a)}{\partial x_n}(b_n - a_n),$$

$$D^2 f(a)(b-a, b-a) = \sum_{m=1}^{N}\sum_{n=1}^{N} \frac{\partial^2 f(a)}{\partial x_m \partial x_n}(b_m - a_m)(b_n - a_n),$$

and

$$D^r(f(a))(b-a, \ldots b-a) = \sum_{n_1=1}^{N}\sum_{n_2=1}^{N} \ldots \sum_{n_r=1}^{N} \frac{\partial^r f(a)}{\partial x_{n_1} \partial x_{n_2} \ldots \partial x_{n_r}}(b_{n_1} - a_{n_1}) \ldots (b_{n_r} - a_{n_r}).$$

Taylor's theorem can strengthened if we assume that rth derivative of f is continuous.

Definition: If $f: U \to R^K$, where U is an open subset of R^N and r is a positive integer, then f is r times continuously differentiable if it is r times differentiable and all the partial derivatives of order r of all K component functions, f_k, of f are continuous.

Theorem 7.2: Let $f: U \to R$, where U is an open subset of R^N. Suppose that U contains the straight line segment from a to b, where $a \neq b$, and suppose that f is r times continuously differentiable, where $r \geq 1$. If

$$R(a, b) = \frac{1}{r!} D^r f(q)(b-a, \ldots, b-a)$$

is the remainder term in Taylor's formula 7.1, then

$$\lim_{b \to a} \frac{R(a, b)}{||b-a||^{r-1}} = 0.$$

Proof: For the moment, let $||x||_M$ be the maximum norm of any vector. That is,

$$||x||_M = \max\{|x_n| : x_n \text{ is a component of } x\}.$$

Recall the $D^r f(q)$ is an r-linear form and so can be represented by a vector with N^r components. Since $D^r f(q)$ is continuous with respect to q, there exists a $\delta > 0$, such that $||D^r f(q) - D^r f(a)||_M < 1$, if $||q-a||_M < \delta$. Then if $||q-a||_M < \delta$

$$|D^r f(q)(b-a, \ldots, b-a)|$$

$$\leq |D^r f(q)(b-a, \ldots, b-a) - D^r f(a)(b-a, \ldots, b-a)| + |D^r f(a)(b-a, \ldots, b-a)|$$

$$\leq (||b-a||_M N)^r + ||D^r f(a)||_M (||b-a||_M N)^r$$

Therefore

$$\frac{|D^r f(q)(b-a, \ldots, b-a)|}{||b-a||_M^{r-1}} \leq \frac{(||b-a||_M N)^r + ||D^r f(a)||_M (||b-a||_M N)^r}{||b-a||_M^{r-1}}$$

$$\leq (1 + ||D^r f(a)||_M) N^r ||b-a||_M,$$

and so

$$\lim_{q \to a} \frac{|D^r f(q)(b-a, \ldots, b-a)|}{||b-a||_M^{r-1}} = 0.$$

Since $||b-a|| \geq ||b-a||_M$, this equation proves the theorem. ■

Remark: If $r = 1$, the above theorem has the same conclusion as lemma 6.3.

The next theorem is the multivariate analogue of theorem 5.10.

Theorem 7.3: Let $f: U \to R$, where U is an open subset of R^N and suppose that f is twice continuously differentiable. If $c \in U$ is such that $Df(c) = 0$ and the symmetric bilinear form $D^2 f(c)$ is negative definite, then f has a local maximum at c. If $D^2 f(c)$ is positive definite, f has a local minimum at c.

Proof: The matrix representation of $D^2 f(c)$ is negative definite if and only if its leading principal minors have determinants that alternate in sign with the first being negative. Since these determinants are continuous functions of x, if $D^2 f(c)$ is negative definite, $D^2 f(x)$ is negative definite for x sufficiently close to c. That is, there is a positive number ε such that $D^2 f(x)$ is negative definite if $||x - c|| < \varepsilon$. Suppose that $x \in U$ is such that $0 < ||x - c|| < \varepsilon$. By Taylor's theorem 7.1, there is a vector q on the line segment from c to x such that

$$f(x) = f(c) + Df(c)(x-c) + \frac{1}{2} D^2 f(q)(x-c, x-c).$$

Since $||q - c|| \leq ||x - c|| < \varepsilon$, $D^2 f(q)$ is negative definite and so $D^2 f(q)(x-c, x-c) < 0$. By assumption, $Df(c) = 0$, so that the above equation implies that $f(x) < f(c)$.

A similar argument shows that if $Df(c) = 0$ and $D^2 f(c)$ is positive definite, then f has a

The Inverse and Implicit Function Theorems

The inverse and implicit function theorems elaborate the idea that continuously differentiable functions behave locally like their derivatives. The proofs of these theorems are long and difficult and will not be provided. The focus will be on what the theorems mean.

The inverse function theorem says that if Df(c) is invertible, then $f: U \to R^M$, where U is an open subset of R^N, has an inverse defined on an open set containing f(c). Recall that Df(c) is a linear transformation from R^N to R^M and it is invertible if and only if there is a linear transformation $(Df(c))^{-1}$ from R^M to R^N such that $(Df(c))^{-1}Df(c) = I = Df(c)(Df(c))^{-1}$, where I is the identity function from R^N to R^N. Of course, M = N, if Df(c) is invertible.

Inverse Function Theorem 7.4: Let $f: U \to R^N$, where U is an open subset of R^N, and suppose that f is continuously differentiable on U. Suppose that Df(c) is invertible, where $c \in U$. Then there exists an open subset V of U that contains c and is such that W = f(V) is an open subset of R^N containing f(c) and there exists a function $g: W \to V$ that is continuously differentiable on W and is such that $f \circ g(w) = w$ and $g \circ f(v) = v$, for all $w \in W$ and $v \in V$. If $v \in V$, then $Dg(f(v)) = (Df(v))^{-1}$.

I introduce the implicit function theorem by describing the corresponding result for linear functions. Let $T: R^{N+K} \to R^K$ be a linear function with Kx(N+K) matrix representation C with respect to the standard bases of R^{N+K} on R^K. Suppose that C has rank K. (Since the maximum rank of C is K, C may be said to have full rank.) Hence C has K independent columns, so that $T(R^{N+K}) = R^K$ and T is onto. We may assume, without loss of generality, that the last K columns of C are independent. Write C as $C = (A \vdots B)$, where A is a KxN matrix and B is an invertible KxK matrix. Write a vector in R^{N+K} as (x, y), where $x \in R^N$ and $y \in R^K$. The equation z = T(x, y) may be written as

$$z = C\begin{pmatrix} x \\ y \end{pmatrix} = (A \vdots B)\begin{pmatrix} x \\ y \end{pmatrix} = Ax + By,$$

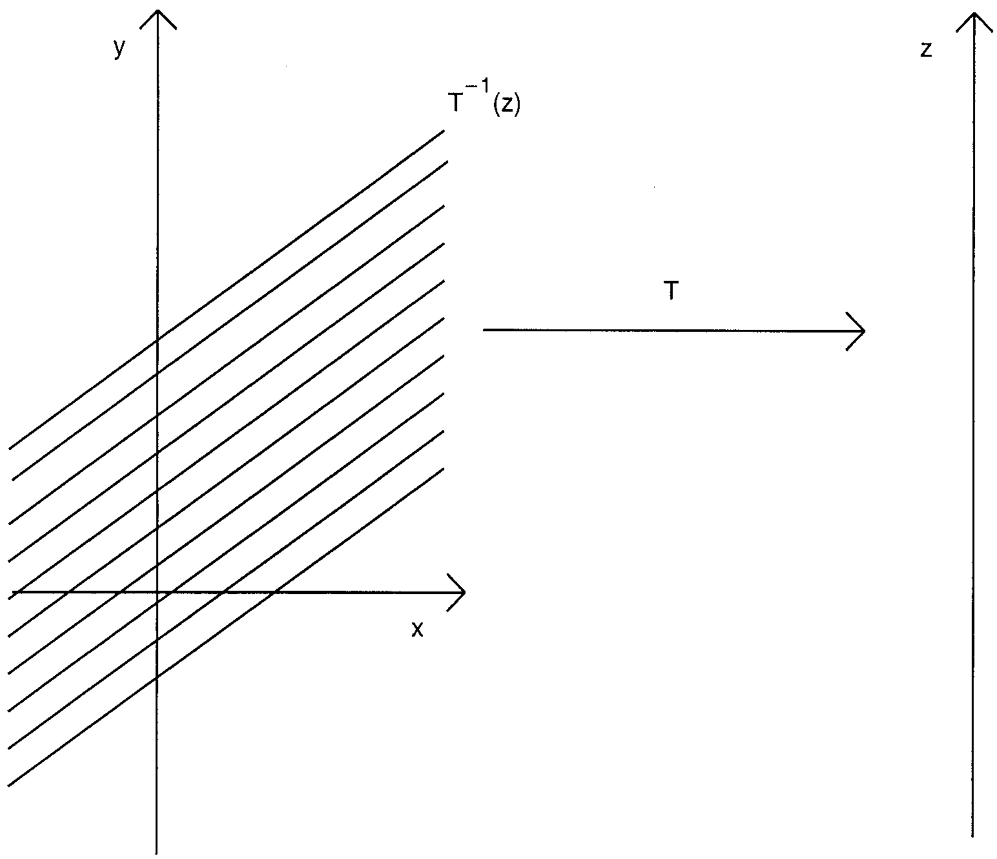
so that By = z – Ax. Solving this last equation for y, we obtain $y = B^{-1}(z - Ax)$. Let $H: R^N \times R^K \to R^K$ be the linear transformation $H(x, z) = B^{-1}(z - Ax)$. Then T(x, H(x, z)) = z. We can double check this by the following calculation.

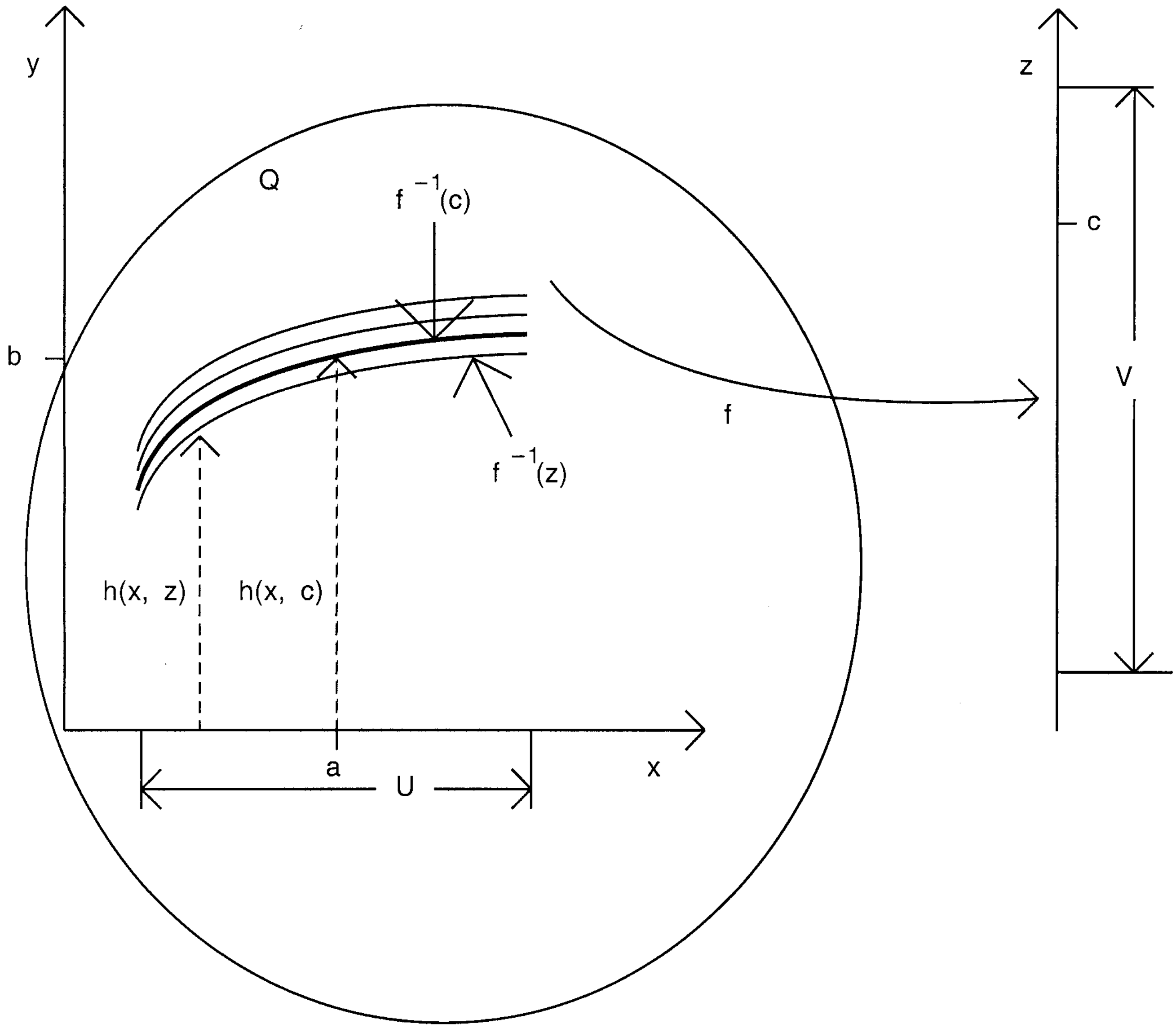
$$T(x, H(x, z)) = (A \vdots B)\begin{pmatrix} x \\ B^{-1}(z - Ax) \end{pmatrix} = Ax + BB^{-1}(z - Ax)$$

$$= Ax + I(z - Ax) = Ax + z - Ax = z.$$

For each $z \in R^K$, the set $T^{-1}(z)$ is the graph of H(x, z) as a function of x with z held fixed. The space R^{N+K} may be seen as $R^N \times R^K$, and the above equations show that the function T is like a projection of R^{N+K} onto R^K followed by an invertible linear transformation, with matrix representation B, from R^K to R^K. In other words, $(x, y) \in R^{N+K}$, equals (x, H(x, z))

$= (x, B^{-1}(z - Ax))$, for some $z \in R^K$, and the point $(x, B^{-1}(z - Ax))$ projects onto $(0, B^{-1}z)$, which in turn is carried to $BB^{-1}z = z$. The figure immediately below may help you visualize these statements. The implicit function theorem says that the same kind of assertions apply locally to a continuously differentiable function.





<u>Implicit Function Theorem</u> 7.5: Let Q be an open set in R^{N+K} and let $f: Q \to R^K$ be continuously differentiable. Let $(a, b) \in Q$, where $a \in R^N$ and $b \in R^K$, and suppose that the last K columns of Df(a, b) are linearly independent. Let x vary over R^N and y and z vary over R^K, so that $f(x, y) = z$, if $(x, y) \in Q$. Then there is an open set U in R^N such that $a \in U$, there is an open set V in R^K such that $c = f(a, b) \in V$, and there is a continuously differentiable function $h: U \times V \to R^K$ such that $f(x, h(x, z)) = z$, for all $x \in U$ and $z \in V$, and $h(a, c) = b$. Also

$$D_x h(a, c) = -(D_y f(a, b))^{-1} D_x f(a, b) \text{ and}$$
$$D_z h(a, c) = (D_y f(a, b))^{-1},$$

where $\left(D_y f(a, b)\right)^{-1}$ exists because the last K columns of Df(a, b) are independent.

The theorem is illustrated in the above figure. The open set Q is indicated by the large oval. The open set U is an interval along the x-axis on the left. The open set V is an interval along the z-axis on the right. It can be seen that f is locally like a projection onto R^K along the sets $f^{-1}(z)$, for $z \in V$, followed by an invertible function from an open subset of R^K to the open subset V of R^K, just as the linear function $T: R^{N+K} \to R^K$ was like a linear projection onto R^K followed by an invertible linear function from R^K to itself.

The equation $f(x, y) = z$ is said to define the function $y = h(x, z)$ implicitly.

The derivative formulas in the implicit function theorem may be obtained by applying the chain rule to the equation $f(x, h(x, z)) = z$ at $(x, y) = (a, b)$. Differentiating this equation with respect to x, we find that $D_x f(a, b) + D_y f(a, b) D_x h(a, c) = 0$, which implies that $D_x h(a, c) = -(D_y f(a, b))^{-1} D_x f(a, b)$. Differentiating with respect to z, we find that $D_y f(a, b) D_z h(a, c) = I$, which implies that $D_z h(a, c) = (D_y f(a, b))^{-1}$.

Example: Let $f(x, y) = x^2 + y^2$. The equation $f(x, y) = r^2$ defines a circle of radius r. If $-\overline{r} < \overline{x} < \overline{r}$ and $\overline{x}^2 + \overline{y}^2 = \overline{r}^2$, then $\overline{y} \neq 0$, so that $\partial f(\overline{x}, \overline{y})/\partial y = 2\overline{y} \neq 0$. For some small positive number ε, there are open intervals $(\overline{x} - \varepsilon, \overline{x} + \varepsilon)$ and $(\overline{r} - \varepsilon, \overline{r} + \varepsilon)$ and there is a continuously differentiable function $h: (\overline{x} - \varepsilon, \overline{x} + \varepsilon) \times (\overline{r} - \varepsilon, \overline{r} + \varepsilon) \to R$ such that $f(x, h(x, r)) = r$, for all $x \in (\overline{x} - \varepsilon, \overline{x} + \varepsilon)$ and $r \in (\overline{r} - \varepsilon, \overline{r} + \varepsilon)$. This function may be chosen to be either $\sqrt{r^2 - x^2}$ or $-\sqrt{r^2 - x^2}$.

The Envelope Theorem

The envelope theorem may be explained intuitively as follows. Consider an optimization problem where the objective function, f, depends on a vector of endogenous variables x and a vector of exogenous parameters b. The objective function is maximized with respect to x, and the question is how do we calculate the derivative of the maximized value of f with respect to b. You might imagine that to calculate this dependence, you would have to take account of the dependence on b of the maximizing value of x. This is not so, however, because the derivative of f with respect to x is zero at the maximum. Therefore, the derivative of the maximized value of f equals the derivative of f with respect to b alone, where the derivative is calculated at the maximizing value of x.

In order to explain this assertion formally, let $f: W \to R$, where W is an open subset of $R^{N+K} = R^N \times R^K$. Write a vector in R^{N+K} as (x, b), where $x \in R^N$ and $b \in R^K$. Consider the problem

$$\max_{x \in R^N} f(x, b) \quad \text{s.t. } (x, b) \in W. \tag{7.3}$$

The vector of endogenous choice variables is x and b is the vector of exogenous parameters. Let the maximized value be

$$F(b) = \max_{x \in R^N}\{f(x, b) \mid (x, b) \in W\},$$

assuming that the maximum exists. Assume that f is twice continuously differentiable and that $D_x^2 f(x, b)$ is negative definite, for all $(x, b) \in W$. If $x = \bar{x}$ solves problem 7.3 at $b = \bar{b}$, then $D_x f(\bar{x}, \bar{b}) = 0$. Because $D_x^2 f(\bar{x}, \bar{b})$ is negative definite, it is invertible, so that by the implicit function theorem applied to the equation $D_x f(x, \bar{b}) = 0$, there is an open set U in R^K such that $\bar{b} \in U$ and there is a continuously differentiable function $h: U \to R^N$ such that $h(\bar{b}) = \bar{x}$ and $D_x f(h(b), b) = 0$, for all $b \in U$. Since $D_x^2 f(x, b)$ is negative definite, for all (a, b), it follows that f(x, b) has a local maximum as a function of x at x = h(b). We will see later that this maximum is in fact global, so that F(b) = f(h(b), b). By the chain rule,

$$DF(b) = Df(h(b), b)$$

$$= D_x f(h(b), b) Dh(b) + D_b f(h(b), b)$$

$$= 0 + D_b f(h(b), b)$$

$$= D_b f(h(b), b),$$

where by $D_b f(h(b), b)$ I mean the derivative at b of f(h(c), b) with respect to b, with c held fixed at the value of b at which the derivative is evaluated. The expression $D_b f(h(b), b)$ is like the partial derivative with respect to the second b, with h(b) held fixed. The equation

$$DF(b) = D_b f(h(b), b)$$

is called the <u>envelope theorem</u>.

<u>Example</u>: Consider the following simple profit maximization problem. A firm uses labor L to produce output y according to the production function y = f(L). Let p be the price of output and w the wage of labor. Profit is

$$\pi(p, w) = \max_{L \geq 0} [pf(L) - wL].$$

Assume that f is twice continuously differentiable and that, for all L, $df(L)/dL > 0$ and $d^2f(L)/dL^2 < 0$. Then there is a continuously differentiable function $L(p, w)$ such that $\pi(p, w) = pf(L(p, w)) - wL(p, w)$. The envelope theorem asserts that

$$\frac{\partial \pi(p, w)}{\partial w} = -L(p, w).$$

Constrained Optimization

Consider the problem

$$\max_{x \in U} f(x)$$
$$\text{s.t. } g_k(x) = \overline{a}_k, \tag{7.4}$$
$$\text{for } k = 1, \ldots, K,$$

where U is an open subset of R^N, f and each of the g_k are functions from U to R, and the $\overline{a}_k$ are numbers. The next theorem states a necessary condition for a solution of this problem and defines what are known as Lagrange multipliers.

Theorem 7.6: Assume that f and $g_1, \ldots, g_K$ are continuously differentiable and that f achieves at $\overline{x}$ a local maximum (or minimum) on $\{x \in U \mid g_k(x) = \overline{a}_k, \text{ for } k = 1, \ldots, K\}$. Assume that the vectors $Dg_1(\overline{x}), \ldots, Dg_K(\overline{x})$ are linearly independent. Then there exist numbers $\lambda_1, \ldots, \lambda_K$ such that

$$Df(\overline{x}) = \sum_{k=1}^{K} \lambda_k Dg_k(\overline{x}).$$

Terminology: The numbers $\lambda_1, \ldots, \lambda_K$ are called Lagrange multipliers, and the requirement that the vectors $Dg_1(\overline{x}), \ldots, Dg_K(\overline{x})$ be linearly independent is called the constraint qualification.

Proof of theorem 7.6: Let $g: U \to R^K$ be the function

$$g(x) = \begin{pmatrix} g_1(x) \\ . \\ . \\ . \\ g_K(x) \end{pmatrix}.$$

The constraint qualification implies that $Dg(\bar{x})$ has rank K. Without loss of generality, we may assume that the last K columns of the KxN matrix $Dg(\bar{x})$ are independent. Write any vector $x \in U$ as $x = (y, z)$, where $y \in R^{N-K}$ and $z \in R^K$. Let $\bar{x} = (\bar{y}, \bar{z})$. Then $g(\bar{y}, \bar{z}) = (\bar{a}_1, ..., \bar{a}_K)$.

The objective is to show that there is a K-vector λ such that $Df(\bar{x}) = \lambda^T Dg(\bar{x})$, where λ^T is the transpose of the 1xK vector λ. This equation is the same as

$$Df(\bar{x}) = \sum_{k=1}^{K} \lambda_k Dg_k(\bar{x}),$$

which implies that

$$D_z f(\bar{y}, \bar{z}) = \lambda^T D_z g(\bar{y}, \bar{z}), \tag{7.5}$$

Because $D_z g(\bar{y}, \bar{z})$ is invertible by assumption, we may define λ^T by the equation

$$\lambda^T = D_z f(\bar{y}, \bar{z}) \left(D_z g(\bar{y}, \bar{z})\right)^{-1}. \tag{7.6}$$

Since $D_z g(\bar{y}, \bar{z})$ is invertible, the implicit function theorem 7.5 implies that there is an open set W in R^{N-K} that contains $\bar{y}$ and there exists a continuously differentiable function $h: W \to R^K$ such that $h(\bar{y}) = \bar{z}$ and $g(y, h(y)) = \bar{a}$, for all $y \in W$. Also

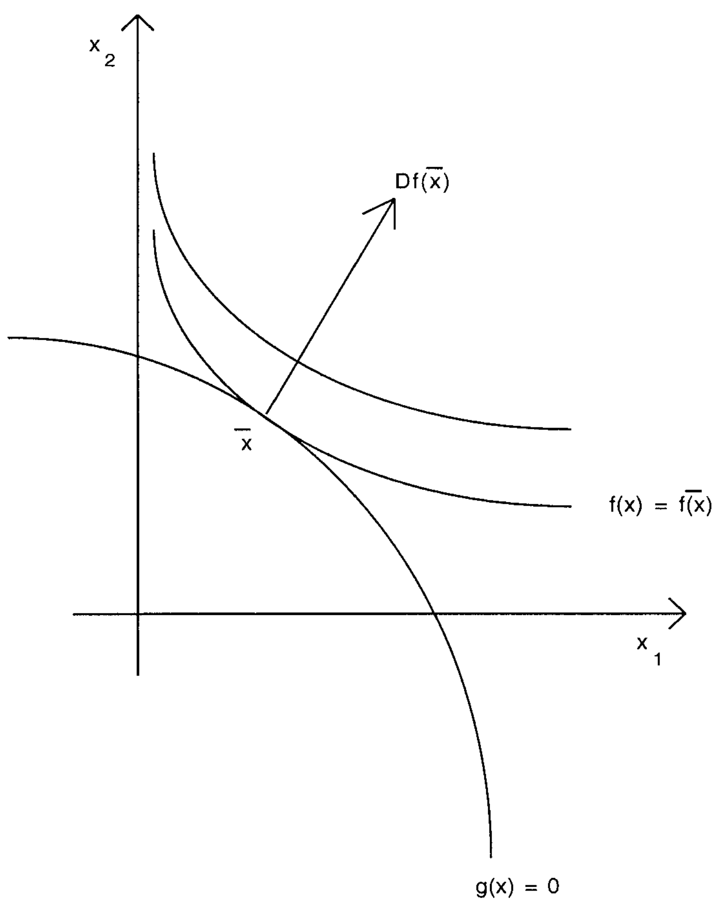
$$Dh(\bar{y}) = -\left(D_z g(\bar{y}, \bar{z})\right)^{-1} D_y g(\bar{y}, \bar{z}). \tag{7.7}$$

Let $F: W \rightarrow R$ be defined by $F(y) = f(y, h(y))$. The function F has a local maximum (or minimum) at $\bar{y}$, so that

$$\begin{aligned} 0 &= DF(\bar{y}) = D_y f(\bar{y}, \bar{z}) + D_z f(\bar{y}, \bar{z}) Dh(\bar{y}) \\ &= D_y f(\bar{y}, \bar{z}) - D_z f(\bar{y}, \bar{z}) \left(D_z g(\bar{y}, \bar{z})\right)^{-1} D_y g(\bar{y}, \bar{z}) \qquad (7.8) \\ &= D_y f(\bar{y}, \bar{z}) - \lambda^T D_y g(\bar{y}, \bar{z}), \end{aligned}$$

where the last two equations follow from equations 7.7 and 7.6. Equations 7.5 and 7.8 imply that $Df(\bar{x}) = \lambda^T Dg(\bar{x})$. ■

I now try to help you visualize the assertion of theorem 7.6.



Example: If $f: R^2 \to R$, $K = 1$, $g: R^2 \to R$, and $\bar{a} = 0$, then the picture is as in the above figure. The vector $Df(\bar{x})$ must lie on the same line as $Dg(\bar{x})$, where $\bar{x}$ is the constrained maximum, the constraint being $g(x) = 0$. This is so because the curve $g(x) = 0$ is tangent at $\bar{x}$ to the curve $f(x) = f(\bar{x})$, the vector $Dg(\bar{x})$ is perpendicular at $\bar{x}$ to the curve $g(x) = 0$ at $\bar{x}$, and the vector $Df(\bar{x})$ is perpendicular at $\bar{x}$ to the curve $f(x) = f(\bar{x})$.

Example: (Necessity of the constraint qualification) Let $N = 2$ and $f(x_1, x_2) = x_2$. Let $g_1(x_1, x_2) = -x_1 + x_2^2$ and $g_2(x_1, x_2) = x_1 + x_2^2$. Let $a_1 = a_2 = 0$. The only point satisfying the constraints $g_1(x_1, x_2) = 0 = g_2(x_1, x_2)$ is $(x_1, x_2) = (0, 0)$, because the equations $-x_1 + x_2^2 = 0 = x_1 + x_2^2$ imply that $x_1 = x_2 = 0$. The point $(x_1, x_2) = (0, 0)$ therefore maximizes $f(x_1, x_2)$ subject to the constraints $g_1(x_1, x_2) = g_2(x_1, x_2) = 0$. Since $Df(0, 0) = (0, 1)$ and $Dg_1(0, 0) = (-1, 0)$ and $Dg_2(0, 0) = (1, 0)$, it is clear that $Df(0, 0)$ is not a linear combination of $Dg_1(0, 0)$ and $Dg_2(0, 0)$, so that the assertion of theorem 7.6 does not apply.

Remark: The function

$$\mathcal{L}(x, \lambda, \bar{a}) = f(x) - \lambda.[g(x) - \bar{a}] = f(x_1, \ldots, x_N) - \sum_{k=1}^{K} \lambda_k [g_k(x_1, \ldots, x_N) - \bar{a}_k]$$

is called the Lagrangian. The equation $D_{(x, \lambda)} \mathcal{L}(\bar{x}, \lambda, \bar{a}) = 0$ yields the equations

$$\frac{\partial}{\partial x_n} \mathcal{L}(\bar{x}, \lambda, \bar{a}) = 0, \tag{7.9}$$

for $n = 1, \ldots, N$, and

$$D_\lambda \mathcal{L}(\bar{x}, \lambda, \bar{a}) = 0, \tag{7.10}$$

for $k = 1, \ldots, K$. Equations 7.9 imply that

$$\frac{\partial f(\bar{x}_1, \ldots, \bar{x}_N)}{\partial x_n} = \sum_{k=1}^{K} \lambda_k \frac{\partial g_k(\bar{x}_1, \ldots, \bar{x}_N)}{\partial x_n}, \tag{7.11}$$

for $n = 1, \ldots, N$, and equation 7.10 implies that

$$g_k(\bar{x}_1, \ldots, \bar{x}_N) = \bar{a}_k, \tag{7.12}$$

for k = 1, ... , K. Equations 7.11 are called the <u>first order conditions,</u> and equations 7.12 are called the <u>constraints</u>. These are necessary conditions for optimality.

The next theorem gives conditions that are sufficient for local optimality in problem 7.4.

<u>Theorem</u> 7.7: Let U be an open subset of R^N, let $f: U \to R$ and $g: U \to R^K$ be twice continuously differentiable and let $\bar{a} \in R^K$. Suppose that $\bar{x} \in U$ is such that $g(\bar{x}) = \bar{a}$ and $Dg(\bar{x})$ has rank K. Let $\lambda \in R^K$ be such that $Df(\bar{x}) = \lambda^T Dg(\bar{x}) = \sum_{k=1}^{K} \lambda_k Dg_k(\bar{x})$. Let

$$\mathcal{L}(x, \lambda, \bar{a}) = f(x) - \sum_{k=1}^{K} \lambda_k [g_k(x) - \bar{a}_k] \text{ and let}$$

$$Z(\bar{x}) = \{v \in R^N \mid Dg(\bar{x})(v) = 0\} = \{v \in R^N \mid Dg_k(\bar{x})(v) = 0, \text{ for } k = 1, \dots, K\}.$$

a) If $v^T D_x^2 \mathcal{L}(\bar{x}, \lambda, \bar{a}) v < 0$, for all $v \in Z(\bar{x})$ such that $v \neq 0$, then f achieves a local maximum at $\bar{x}$ on $\{x \in U \mid g(x) = \bar{a}\}$.

b) If $v^T D_x^2 \mathcal{L}(\bar{x}, \lambda, \bar{a}) v > 0$, for all $v \in Z(\bar{x})$ such that $v \neq 0$, then f achieves a local minimum at $\bar{x}$ on $\{x \in U \mid g(x) = \bar{a}\}$.

<u>Proof</u>: Since $Dg(\bar{x}) : R^N \to R^K$ has rank K, we may assume that the last K columns of the KxN matrix $Dg(\bar{x})$ are independent. Write $x \in R^N$ as $x = (y, z)$, where $y \in R^{N-K}$ and $z \in R^K$. Let $\bar{x} = (\bar{y}, \bar{z})$. Thus the KxK matrix $D_z g(\bar{y}, \bar{z})$ is invertible. By the implicit function theorem 7.5, there is an open set W in R^{N-K} that contains $\bar{y}$ and there is a continuously differentiable function $h: W \to R^K$ such that $h(\bar{y}) = \bar{z}$ and $g(y, h(y)) = \bar{a}$, for all $y \in W$. Also

$$Dh(\bar{y}) = -\left(D_z g(\bar{y}, \bar{z})\right)^{-1} D_y g(\bar{y}, \bar{z}). \qquad (7.13)$$

Since g is twice continuously differentiable, it follows that h is so as well.

I prove assertion (a). The proof of assertion (b) is similar. Let $F: W \to R$ be defined by the equation $F(y) = f(y, h(y))$. It is sufficient to show that F achieves a local maximum at $\bar{y}$. Since f and h are twice continuously differentiable, F is so as well. By theorem 7.3, it is sufficient to show that $DF(\bar{y}) = 0$ and $D^2F(\bar{y})$ is negative definite.

If we differentiate the equation $F(y) = f(y, h(y))$ and substitute equations 7.6 and 7.13, we find that

$$
\begin{aligned}
DF(\bar{y}) &= D_y f(\bar{y}, \bar{z}) + D_z f(\bar{y}, \bar{z}) Dh(\bar{y}) \\
&= D_y f(\bar{y}, \bar{z}) - D_z f(\bar{y}, \bar{z}) \left(D_z g(\bar{y}, \bar{z})\right)^{-1} D_y g(\bar{y}, \bar{z}) \\
&= D_y f(\bar{y}, \bar{z}) - \lambda^T D_y g(\bar{y}, \bar{z}) = 0 .
\end{aligned}
$$

It remains to be shown that $w^T D^2F(\bar{y}) w < 0$, for all $w \in R^{N-K}$ such that $w \neq 0$. Fix such a w and let $\sigma(t) = (\bar{y} + tw, h(\bar{y} + tw))$, for t in so small an interval about 0 that $\bar{y} + tw \in W$, for all t in the interval. Then $\sigma(0) = \bar{x}$. Also since $g(\sigma(t)) = \bar{a}$, for all t, the chain rule implies that

$$0 = \left.\frac{d}{dt}\right|_{t=0} g(\sigma(t)) = Dg(\bar{x}) D\sigma(0) .$$

Hence $D\sigma(0) \in Z(\bar{x})$. The equation $g(\sigma(t)) = \bar{a}$, for all t, also implies that

$$Dg(\sigma(t))D\sigma(t) = 0,$$

for all t, so that

$$Dg_k(\sigma(t)) D\sigma(t) = 0,$$

for $k = 1, \ldots, K$ and for all t. Differentiating this last equation with respect to t and applying the chain rule and Leibniz's rule, we see that

$$\left(D\sigma(t)\right)^T D^2 g_k(\sigma(t)) D\sigma(t) + Dg_k(\sigma(t)) D^2\sigma(t) = 0, \tag{7.14}$$

for all t, where

$$D^2\sigma(t) = \begin{pmatrix} d^2\sigma_1(t)/dt^2 \\ \cdot \\ \cdot \\ \cdot \\ d^2\sigma_N(t)/dt^2 \end{pmatrix}.$$

At t = 0, equation 7.14 becomes

$$Dg_k(\bar{x})\,D^2\sigma(0) = -v^T D^2 g_k(\bar{x})\,v, \tag{7.15}$$

where $v = D\sigma(0) \in Z(\bar{x})$. Since

$$D\sigma(0) = \begin{pmatrix} w \\ Dh(\bar{y})\,w \end{pmatrix}$$

and $w \neq 0$, it follows that $v \neq 0$. By applying the chain rule again, we see that

$$\frac{df(\sigma(t))}{dt} = Df(\sigma(t))\,D\sigma(t)$$

and

$$\frac{d^2f(\sigma(t))}{dt^2} = D\sigma(t)^T D^2 f(\sigma(t))\,D\sigma(t) + Df(\sigma(t))\,D^2\sigma(t)\,.$$

At t = 0, the last equation implies

$$\frac{d^2f(\sigma(0))}{dt^2} = D\sigma(0)^T D^2 f(\bar{x})\,D\sigma(0) + Df(\bar{x})\,D^2\sigma(0)$$

$$= v^T D^2 f(\bar{x})\,v + \sum_{k=1}^{K} \lambda_k Dg_k(\bar{x})\,D^2\sigma(0)$$

$$= v^T D^2 f(\bar{x})\,v - \sum_{k=1}^{K} \lambda_k v^T D^2 g_k(\bar{x})\,v \tag{7.16}$$

$$= v^T \left[D^2 f(\bar{x}) - \sum_{k=1}^{K} \lambda_k D^2 g_k(\bar{x}) \right] v$$

$$= v^T D_x^2 \mathcal{L}(\bar{x}, \lambda, \bar{a})\,v < 0,$$

where the third equation follows from equation 7.15 and the inequality is valid by assumption, since $v \neq 0$ and $v \in Z(\bar{x})$. Since

$$f(\sigma(t)) = f(\bar{y} + tw, h(\bar{y} + tw)) = F(\bar{y} + tw),$$

it follows that

$$\frac{d^2 f(\sigma(0))}{dt^2} = w^T D^2 F(\bar{y})w.$$

Hence inequality 7.16 implies that

$$w^T D^2 F(\bar{y})w < 0,$$

for all $w \in R^{N-K}$ such that $w \neq 0$. This completes the proof that F achieves a local maximum at $\bar{y}$. ■

Remarks:

1) The condition $g(\bar{x}) = \bar{a}$ is the constraint.

2) The condition $Df(\bar{x}) = \lambda^T Dg(\bar{x})$ is a first order condition for optimality.

3) The constraint and first order conditions are necessary for optimality.

4) The condition $v^T D_x^2 \mathcal{L}(\bar{x}, \lambda, \bar{a})v < 0$, for all non-zero vectors $v \in Z(\bar{x})$ is a second order condition for a local maximum.

5) The condition $v^T D_x^2 \mathcal{L}(\bar{x}, \lambda, \bar{a})v > 0$, for all non-zero vectors $v \in Z(\bar{x})$ is a second order condition for a local minimum.

6) The constraint and the first and second order conditions together are sufficient for local optimality.