ECONOMICS 241

5.1 Minimization of Norms

The typical optimization problem is to maximize (or minimize) a function $f:X\to R$ subject to a constraint $x \in C$. We write

Definition 5.1: x^* maximizes $f:X\to R$ subject to $x \in C \subseteq X$ if $x^* \in C$ and

$$(\forall y \in C) f(y) \le f(x^*)$$

x^* minimizes f subject to $x \in C$ if $x^* \in C$ and

$$(\forall y \in C) f(x^*) \le f(y).$$

EXAMPLE 5.1: A consumer can purchase quantity $x_i \ge 0$ of good i at price p_i per unit. $x = (x_1,\ldots,x_n)$ and $p = (p_1,\ldots,p_n)$. The value of the goods bundle to the consumer is $U(x)$. The feasible purchase set is

$$C = \{x \in R^n / x_i \ge 0 \text{ and } p\cdot x < b\}$$

where b is the consumer's budget. The consumer's problem, then, is

$$\text{maximize } U(x) \text{ s.t. } x \in C.$$

EXAMPLE 5.2: A firm's feasible production set can be described by a constraint on the goods it uses $(y_1,\ldots,y_n)=y$ where $y_i > 0$ means i is an output and $y_i < 0$ means i is an input, of the form $y \in C$ (typically, increasing y_i will cause some other y_j to fall). If $p_1,\ldots,p_n$ are the prices of the n commodities, the firm's problem is

$$\max_{y\in C} p\cdot y$$

which means maximize $p\cdot y$ subject to $y \in C$. Typically, we will not distinguish minimization from maximization, since

$$\min_{x\in C} f(x) = -\max_{x\in C} -f(x) \tag{5.1}$$

Theorem 5.1: Let C be a closed, convex nonempty set. Then $\exists! z \in C$ satisfying $(\forall x \in C)\ \|z\| \le \|x\|$.

Proof: Let $\alpha = \inf_{x\in C} \|x\|$. α exists since $\|\ \|$ is bounded below by zero. By Theorem 2.11, there is a sequence of points $x_n \in C$, $\|x_n\| \to \alpha$.

By the parallelogram law:

$$\begin{aligned}\|½(x_n - x_m)\|^2 &= 2(\|½x_n\|^2 + \|½x_n\|^2) - \|½(x_n + x_m)\|^2 \\ &= 2(¼\|x_n\|^2 + ¼\|x_m\|^2) - \|½(x_n+x_m)\|^2.\end{aligned}$$

Since x_n, $x_m \in C$ and C is convex,

$$\|½(x_n+x_m)\| \ge \alpha$$

or $$\|½(x_n + x_m)\|^2 \ge \alpha^2.$$

Thus

$$0 \le \lim_{n,m\to\infty} \|½(x_n-x_m)\|^2 = \lim_{n,m\to\infty} ½(\|x_n\|^2 + \|x_m\|^2 - \|½(x_n+x_m)\|^2 \le$$

$$\lim_{n,m\to\infty} ½(\|x_n\|^2 + \|x_m\|^2) = \alpha^2 = ½(\alpha^2 + \alpha^2) - \alpha^2 = 0.$$

Therefore, x_n is a cauchy sequence and has a limit point z. Since C is closed, $z \in C$. Since $\|\ \|$ is continuous

$$\|z\| = \lim_{n\to\infty} \|x_n\| = \alpha \le \|x\| \text{ for all } x \in C.$$

It remains to prove uniqueness. Suppose $\|z_1\| = \|z_2\| = \alpha$. Then, by the parallelogram law

$$\begin{aligned}\alpha^2 \le \|½(z_1+z_2)\|^2 &= 2(\|½z_1\| + \|½z_2^2\|^2) - \|½(z_1-z_2)^2\|^2 \\ &= ½(\|z_1\|^2 + \|z_2\|^2) - \|½(z_1-z_2)\|^2 \\ &= ½(\alpha^2+\alpha^2) - \|½(z_1-z_2)\|^2.\end{aligned}$$

Thus, since $\|\frac{1}{2}(z_1-z_2)\| \geq 0$:

$\alpha^2 \leq \alpha^2 - \|\frac{1}{2}(z_1-z_2)\|^2 \leq \alpha^2$

forcing $\|\frac{1}{2}(z_1-z_2)\| = 0$, or $z_1=z_2$ Q.E.D.

EXAMPLE 5.3 Let $x,y \in R^n$, and define $1 = (1,1,\ldots,1)$,

$$C = \{y-\alpha 1-\beta x \mid \alpha,\beta \in R\}.$$

Then, for $z \in C$, $\|z\|^2 = \sum_{i=1}^{n} (y_i - \alpha 1 - \beta x_i)^2$. C is easily shown to be convex. Thus, Theorem 5.1 provides the element $z = y-\hat{\alpha}1-\hat{\beta}x$ that has the least norm. If we think of $\alpha + \beta X_i$ as an estimate of the value y_i, then $y_i - \alpha - \beta x_i$ is the estimation error. Theorem 5.1 proves there is a unique vector $z = y-\alpha 1-\beta x$ minimizing the sum of squared errors.

In the same way, the set

$$C = \{y - \alpha 1 - \beta x / 0 \leq \beta \leq 1, \alpha \geq 0\}$$

is also convex, and there is a unique $z \in C$ minimizing the norm. Thus, the addition of these constraints on the parameters does not interfere with the application of the theorem.

When $\hat{\alpha}$ and $\hat{\beta}$ are uniquely defined (x not parallel to 1), they are the "ordinary least squares" estimates of the linear equation $y = \alpha+\beta x+\varepsilon$.

Theorem 5.2: Suppose C is a closed, convex set, and x is any point. Then $\exists! z \in C$ satisfying

$$(\forall y \in C)\ \|x-z\| \leq \|x-y\|.$$

In addition

$$\forall y \in C\ (x-z)\cdot(y-z) \leq 0.$$

Proof: Let $\hat{C} = \{x-y \mid y \in C\}$.

Then there is a unique $V \in \hat{C}$ minimizing $\|v\|$ by Theorem 5.1, since $\hat{C}$ is closed and convex. Let $z = x-v$. Clearly $z \in C$, and $\forall y \in C$:

$$\|x-z\| = \|v\| \leq \|x-y\|$$

which establishes existence and uniqueness.

Define, for any $y \in C$

$$g(\lambda) = \|x - ((1-\lambda)z + \lambda y)\|^2.$$

Since, by convexity, $(1-\lambda)z + \lambda y \in C$ and thus for $1 \geq \lambda \geq 0$:

$$g(0) = \|x-z\|^2 \leq \|x-((1-\lambda)z+\lambda y)\|^2 = g(\lambda).$$

Therefore

$$0 \leq g'(0) = \frac{\partial}{\partial\lambda}(x-z)\cdot(x-z) + 2\lambda(x-z)\cdot(z-y) + \lambda^2(z-y)\cdot(z-y)\Big|_{\lambda=0}$$

$$= 2(x-z)\cdot(z-y)$$

or $(x-z)\cdot(y-z) = -(x-z)\cdot(z-y) \leq 0.$ Q.E.D.

From theorem 5.2 there is a unique point $z \in C$ that is closest to the point x. In addition, the angle between $x-z$ and $y-z$ is at least 90^0 for all $y \in C$. This is illustrated in figure 5.1.

Theorem 5.3: Suppose C is a closed, convex set. Then, for $z \in C$:

$$(\forall y \in C)\|x-z\| \leq \|x-y\| \text{ if and only if } (\forall y \in C)(x-z)\cdot(y-z) \leq 0.$$

Proof: $(\Rightarrow)$ was proved in Theorem 5.2 $(\Leftarrow)$. Again consider

$$g(\lambda) = \|x - ((1-\lambda)z + \lambda y)\|^2.$$

Then $g''(\lambda) = 2(z-y)\cdot(z-y) \geq 0$ and g is convex. Thus, by Theorem 2.24,

$$g(1) \geq g(0) + g'(0)(1-0), \text{ or}$$

$\|x-y\| \geq \|x-z\| + 2(x-z)\cdot(z-y) \geq \|x-z\|.$ Q.E.D.

5.2: Maximizaton of Functions

The treatment of this section will be given at an abstract vector space level, precisely because it is costless to do so: the proofs are the same in an abstract inner product space as in R^n. Nonetheless, we have developed

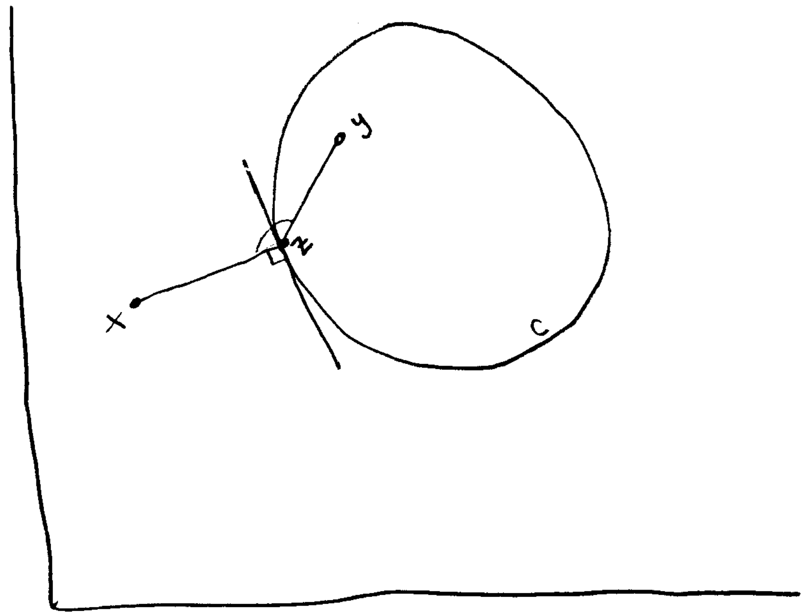

Figure 5.1 For any $y \in C$, the angle between $x - z$ and $y - z$ is at least $90°$.

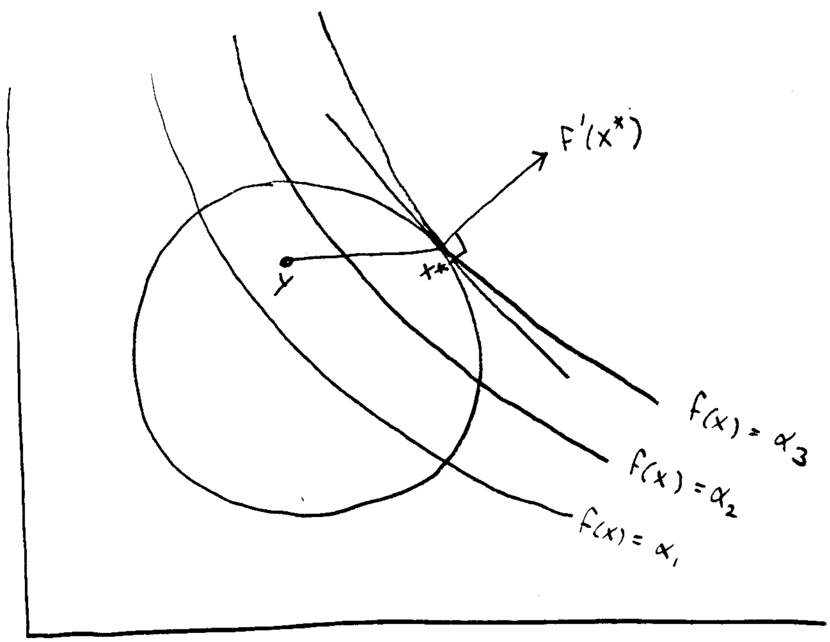

Figure 5.2 The angle between $f'(x^*)$ and $y - x^*$ is at least $90°$.

derivatives only for R^n, and the reader is warned against applications of these theorems outside of R^n will require the development of derivatives in the appropriate inner product space.

Since maximization of a function requires the comparison of values of the function ($f(x) \geq f(y)$), we let $f:X \to R$ be a real valued function. X will be an innder product space with inner product $x \cdot y$. In addition, we shall presume f is continously differentiable.

Theorem 5.4: Suppose x* maximizes $f:X \to R$ subject to $x \in C$, where C is a closed, convex set. Then $\forall y \in C$

$$f'(x^*) \cdot (y - x^*) \leq 0.$$

Proof: Let $y \in C$. Since C is convex, $\lambda y + (1-\lambda)x^* \in C$. Thus, since x* maximizes f,

$$g(\lambda) = f(x^*) - f(x^* + \lambda(y - x^*)) \geq 0.$$

Since $g(0) = 0$, we have

$$0 \leq \lim_{\lambda \to 0} \frac{1}{\lambda}(g(\lambda) - g(0)) = g'(0) = -f'(x^*) \cdot (y - x^*).$$ Q.E.D.

Thus, if x* maximizes f(x) over $x \in C$, we have that $f'(x^*)$ points away from $y-x^*$, that is, the angle formed by $y-x^*$ and $f'(x^*)$ is at least 90^o. This is illustrated in Figure 5.2.

When $C = \{x | g(x) \leq b\}$ for some scalar b, a somewhat simpler treatment can be made. Recall that C is convex if g is a convex function (Theorem 4.23). In this case, we say x* maximizes f subject to $g(x) \leq b$. If $g(x) \leq b$, x is said to be feasible.

Theorem 5.5: Suppose x* maximizes f subject to $g(x) \leq b$, and f,g are continuously differentiable.

Then (i) if $g(x^*) < b$, $f'(x^*) = \bar{0}$

(ii) if $g(x^*) = b$, $\exists\lambda \geq 0$ $f'(x^*) = \lambda g'(x^*)$.

Proof: Case (i). Fix a vector z. Then, since g is continuous, there is a scalar α_0

$$g(x^* + \alpha_0 z) = b$$

by the mean value theorem. Thus, for sufficiently small α, $x^* + \alpha z$ is feasible. Therefore, since x^* maximizes f

$$f(x^*) \geq f(x^* + \alpha z)$$

$$\text{or } \frac{d}{d\alpha} f(x^* + \alpha z)\bigg|_{\alpha=0} \leq 0$$

Consequently, $f'(x^*)\cdot z \leq 0$. Now let $z = f'(x^*)$, and we have

$$\|f'(x^*)\|^2 = f'(x^*)\cdot f'(x^*) \leq 0.$$

This implies $\|f'(x^*)\| = 0$, and hence $f'(x^*) = \bar{0}$, as desired.

Proof of (ii). $g(x^*+\alpha z)$ will be feasible for very small α whenever

$$\frac{d}{dx} g(x^* + \alpha z)\bigg|_{\alpha=0} \leq 0.$$

That is, $g'(x^*)\cdot z \leq 0$. Thus, if $x^* + \alpha z$ is feasible, it cannot increase f, since $f(x^*) \geq f(x^* + \alpha z)$ for feasible $x^* + \alpha z$. Therefore

$$g'(x^*)\cdot z \leq 0 \Rightarrow f'(x^*)\cdot z \leq 0.$$

But, by Lemma 3.7, $\exists\lambda \geq 0$

$$f'(x^*) = \lambda g'(x^*)$$

Q.E.D.

Theorem 5.5 shows that, if x^* maximises f subject to $g(x) \leq b$, then $f'(x^*)$ and $g'(x^*)$ are parallel, or $f'(x^*) = 0$. We may unify this treatment by letting $\lambda=0$ in the latter case, i.e., there will exist a $\lambda \geq 0$ so that

$f'(x^*) - \lambda g'(x^*) = \bar{0}$. This is illustrated in Figure 5.3. Recall that the $g(x) = b$ surface is perpendicular to g', and $f(x) = \alpha$ surface is perpendicular to f'. Thus, if f' and g' are not parallel, neither are the $f =$ constant and $g(x) = b$ surfaces, and we can slip in between them, illustrated by the vector z in Figure 5.~~4~~3.

Theorem 5.5 provides necessary conditions for x to maximize f subject to $g(x) \leq b$. In addition, we have

Theorem 5.6: Consider the unconstrained problem: max $f(x)$ over $x \in R^n$, and suppose f is twice continuously differentiable.

Then if x^* solves this problem: $f'(x^*) = \vec{0}$ and $f''(x^*)$ is nsd.

Proof: The first claim, $f'(x^*) = 0$, follows immediately from the "constraint" which sets $g(x)=b$ for all x, so $g'(x^*) = \vec{0}$. Then note, for $y = \theta^* + (1-\theta)x$, some $0 \leq \theta \leq 1$:

$$f(x) = f(x^*) + f'(x^*)\cdot(x-x^*) + \tfrac{1}{2}(x-x^*)^{\tau} f''(y)(x-x^*)$$

Thus

$$0 \geq \frac{f(x) - f(x^*)}{\|x-x^*\|^2} = \tfrac{1}{2} z^{\tau} f''(y) z$$

where $z = \dfrac{x-x^*}{\|x-x^*\|}$.

Now send x to x^*, and we have

$0 \geq z^{\tau} f''(x^*) z$. Q.E.D.

Theorem 5.7: Suppose f is concave and g is convex, and both are continuously differentiable. Then x^* maximizes $f(x)$ subject to $g(x) \leq b$ if and only if

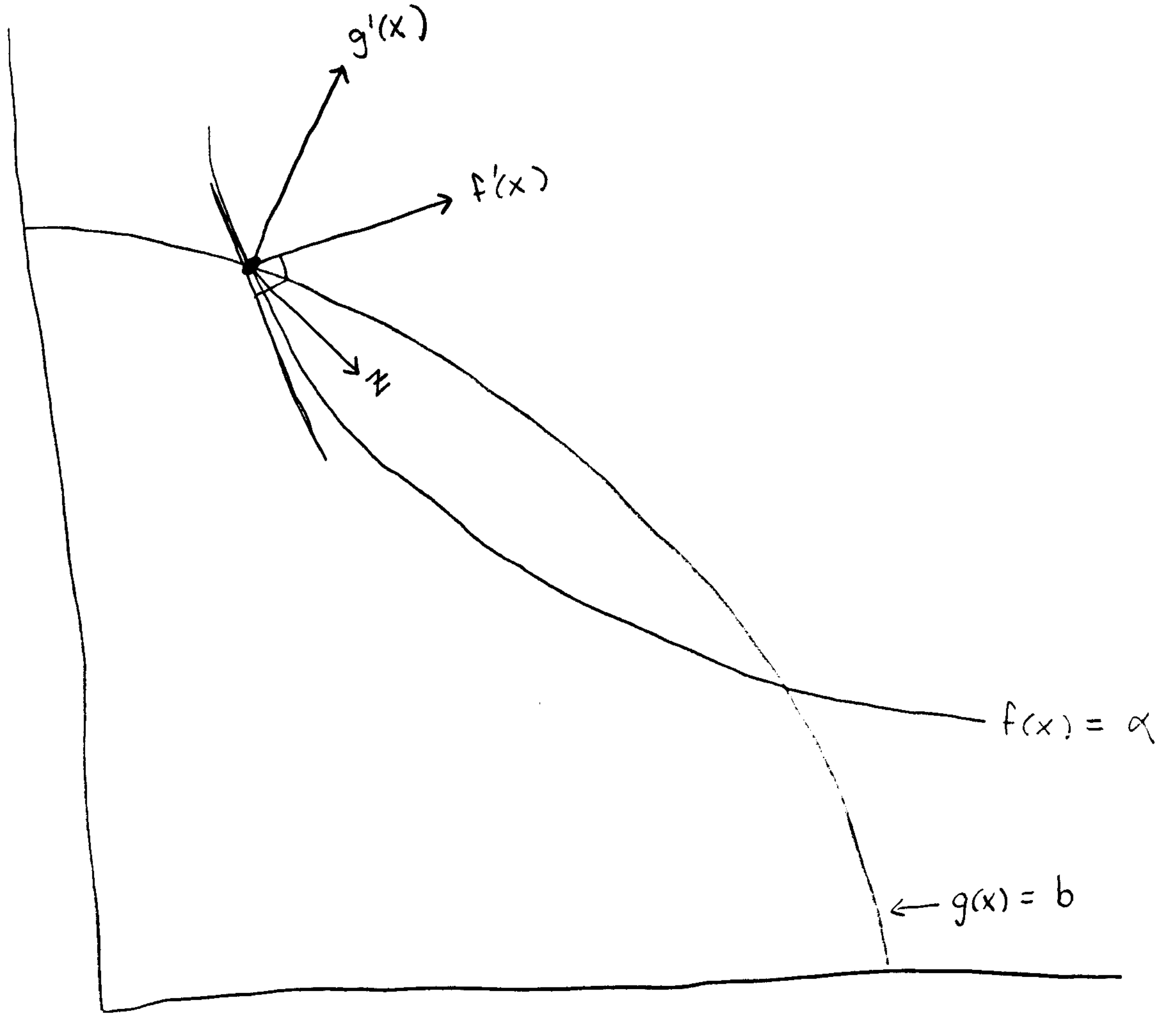

Figure 5.3 : If $f'(x)$ is not parallel to $g'(x)$, a movement in the direction z decreases g (and hence is feasible) and increases f (and hence is desirable), so x does not solve the problem.

$\exists\lambda \geq 0$, $f'(x^*) = \lambda g'(x^*)$, $\lambda(b-g(x^*)) = 0$, and $g(x^*) \leq b$.

Proof: (⇒) follows from Theorem 5.5.

(⇐) Note that the function $f(x) + \lambda(b-g(x))$ is a concave function. By Theorem 4.21:

$$f(x) \leq f(x) + \lambda(b-g(x)) \leq$$

$$f(x^*) + \lambda(b-g(x^*)) + [f'(x^*) - \lambda g'(x^*)](x-x^*) =$$

$$f(x^*) + \lambda(b-g(x^*)) = f(x^*).$$

Thus, if $g(x) \leq b$, $f(x) \leq f(x^*)$ Q.E.D.

One application of this result concerns the direction of fastest increase of a function. That is, if we visualize $f:R^n \to R$ as describing altitude over R^n, in what direction z is f the steepest? Recall the directional derivative of f at x in the direction z may be expressed as

$$f_z(x) = f'(x)\cdot z.$$

To approximate the slope, we should keep the "step size", $\|z\|$ constant. Thus, to find the direction of steepest ascent of f, we should like to find the z solving

$$\max f'(x)\cdot z$$

$$\text{s.t. } \|z\| \leq 1$$

or, equivalently,

$$\max f'(x)\cdot z$$

$$\text{s.t. } \sum_{i=1}^{n} z_i^2 \leq 1.$$

Thus, there is a $\lambda \geq 0$ (since the constraint is convex, and the objective function linear and hence concave) so that:

$$f'(x) = 2\lambda z.$$

Moreover

$$1 = z\cdot z = (2\lambda)^{-2}(f'(x)\cdot f'(x))$$

and $2\lambda = \pm \|f'(x)\|$

Since $\lambda \geq 0$, we desire $2\lambda = \|f'(x)\|$ and this gives the solution

$$z = \frac{f'(x)}{\|f'(x)\|},$$

provided $f'(x) \neq \vec{0}$. That is, the direction of steepest ascent of f at x is f'(x), and the derivative measures not only the slope of the function, but the direction in which the function is increasing fastest.

If $f'(x) = 0$, no direction yields ascent or descent. Finally, if we wanted the direction of steepest descent, we would have used the other solution $2\lambda = \|f'(x)\|$.

5.3 The Value Function

Define

$$V(b) = \max_{\{x|g(x) \leq b\}} f(x).$$

V(b) is the value of the function f, when maximized subject to $g(x) \leq b$. Thus, if f is utility and $g(x) = p \cdot x$ is expenditure, V(b) is the utility the agent achieves for a given expenditure level.

Theorem 5.8: If f is concave and g is convex, V is a nondecreasing concave function and $V'(b) = \lambda$, the Lagrangian multiplier, whenever V' exists.

Proof: First, we show V is nondecreasing. Let x* maximize V for $g(x) \leq b_1$. Then, if $b_1 < b_2$. $g(x^*) \leq b_2$. Thus $V(b_2) \geq f(x^*) = V(b_1)$, since x* is feasible. That is, V is nondecreasing.

Now let $b_1 < b_2$ and x_i^* maximize f subject to $g(x) \leq b_i$. Let $0 \leq \lambda \leq 1$.

Then, by the convexity of g:

$$g(\lambda x_1^* + (1-\lambda)x_2^*) \leq \lambda g(x_1^*) + (1-\lambda)g(x_2^*) \leq \lambda b_1 + (1-\lambda)b_2$$ and thus

$\lambda x_1^* + (1-\lambda)x_2^*$ is feasible for $b = \lambda b_1 + (1-\lambda)b_2$. Thus, by the concavity of f

$$V(\lambda b_1 + (1-\lambda)b_2) \geq f(\lambda x_1^* + (1-\lambda)x_2^*) \geq \lambda f(x_1^*) + (1-\lambda)f(x_2^*) =$$

$$\lambda V(b_1)+(1-\lambda)V(b_2)$$

and hence V is concave.

Now let $x^*(b)$ maximize $f(x)$ subject to $g(x) \leq b$. Note that, if $g(x^*((b)) = b$, $g'(x^*(b))\cdot x^{*\prime}(b) = 1$, and

$$V'(b) = \frac{d}{db}f(x^*(b)) = f'(x^*(b))\cdot x^{*\prime}(b) = \lambda g'(x^*(b))\cdot x^{*\prime}(b) = \lambda.$$

If $g(x^*(b)) < b$, then $\lambda=0$ and

$$V'(b) = f'(x^*(b))\cdot x^{*\prime}(b) = \lambda g'(x^*(b))\cdot x^{*\prime}(b) = 0 = \lambda$$ Q.E.D.

Theorem 5.8 shows that λ receives the interpretation of a shadow value: λ is the value (increase in f) of a slight increase in b. Effectively, λ is the price one would be willing to pay to increase b slightly: λ is the implicit value of weakening the constraint. For f concave and g convex, we now know that λ is nonnegative (Theorem 5.7), and nonincreasing in b, since V is concave forces $0 \geq V''(b) = \lambda'(b)$.

One useful way of expressing the equation

$$f'(x^*) = \lambda g'(x^*)$$

is $$\frac{\partial f/\partial x_1}{\partial b/2x_1} = \frac{\partial f/\partial x_2}{\partial g/\partial x_2} = \ldots = \frac{\partial f/\partial x_n}{\partial g/\partial x_n} = \lambda$$

since this eliminates λ from the first n-1 equalities.

Now suppose the functions f and g have a parameter α as an argument, and our problem is

$$\max_x f(x,\alpha) \text{ s.t. } g(x,\alpha) \leq b$$

Let $x^*(b,\alpha)$ accomplish this, and $V(b,\alpha) = f(x^*(b,\alpha),\alpha)$.

Then

$$\frac{dV(b,\alpha)}{d\alpha} = \frac{\partial f}{\partial x}(x^*(b,\alpha))\cdot\frac{\partial x^*}{\partial \alpha} + \frac{\partial f}{\partial \alpha}$$

$$= \lambda\frac{\partial g}{\partial x}\cdot\frac{\partial x^*}{\partial \alpha} + \frac{\partial f}{\partial \alpha}.$$

But, if $\lambda > 0$, $g(x^*,\alpha) = b$ and

$$\frac{\partial g}{\partial x}\cdot\frac{\partial x^*}{\partial \alpha} + \frac{\partial g}{\partial \alpha} = \frac{\partial b}{\partial \alpha} = 0.$$

$$\text{so } \frac{dV}{d\alpha} = \frac{\partial f(x,\alpha)}{\partial \alpha} - \lambda\frac{\partial g(x,\alpha)}{\partial x}\Bigg|_{x=x^*(b,\alpha)}$$

This is called the Envelope Theorem:

$$\frac{df}{d\alpha}(x^*(b,\alpha)\alpha) = \frac{\partial}{\partial \alpha}[f(x,\alpha)-\lambda g(x,\alpha)]\Bigg|_{x=x^*(b,\alpha)}$$

One way of remembering all these results uses the "Lagrangian":

$$L(x,\lambda,b) = f(x^*) + \lambda(b-g(x^*))$$

$$\frac{\partial L}{\partial b} = \lambda$$

$$0 = \frac{\partial L}{\partial x^*} = f'(x^*) - \lambda g'(x^*)$$

$$\lambda(b-g(x^*)) = 0$$