Notes on Convexity

Roy Radner
Stern School, NYU

September 11, 2006

Abstract

These notes are intended to complement the material in an intermediate microeconomic theory course. In particular, they provide a rigorous discussion of optimality conditions for functions of one and several variables, including minimization of a convex (or maximization of a concave) function subject to a linear constraint. In the case of a function of a single variable, the notes should help the reader go somewhat beyond the standard treatment, which typically deals only with interior solutions and functions that are twice differentiable. The notes should be accessible to somone with a good basic background in univariate and multivariate differential and integral calculus.

[**Preliminary Draft - Do Not Quote**]

Preamble. The mathematical concept of convexity (along with the mirror concept of concavity) plays an important role in economic theory. These notes are intended to complement the material in an intermediate microeconomic theory course. The standard calculus treatment of optimization ("first- and second-order conditions") is justified in Corollary 14 and Propositions 16 and 17. The "method of Lagrange" for optimization subject to a linear constraint is justified in Proposition 19 and the subsequent remarks. However, the notes should help the reader go somewhat beyond the standard treatment, which typically deals only with interior solutions and functions that are twice differentiable.

The notes should be accessible to someone with a good basic background in univariate and multivariate differential and integral calculus. Proofs of propositions that are very easy are omitted. On the other hand, the proofs of some of the propositions require a more advanced background, such as is usually covered in a course in "real analysis;" those propositions are indicated with an asterisk, and the proofs are also omitted. For a more advanced treatment of convexity see, for example, H. L. Royden, *Real Analysis*, Prentice Hall, Englewood Cliffs, NJ, 1988, 2nd ed.

1 Functions of One Variable

1.1 Properties of Convex and Concave Functions

This section deals with real-valued functions of a real variable, defined on a finite interval, a half-line, or the entire real line. A function is *convex* if for any numbers x, y, t, such that $0 \leq t \leq 1$,

$$f[ty + (1-t)x] \leq tf(y) + (1-t)f(x). \tag{1}$$

(See Fig. 1a.) The point $ty + (1-t)x$ is said to be a *convex combination of the points x and y*. A function is said to be *strictly convex* if the inequality in (1) is strict for $0 < t < 1$. (Fig. 1b.) A function f is (strictly) concave if $(-f)$ is (strictly) convex. (Fig. 1c.) In what follows, properties of concave functions that follow immediately from properties of convex functions will be omitted. A function that is both convex and concave is linear.

Proposition 1 *If $f_1, ..., f_N$ are convex, and $b_1, ..., b_N$ are nonnegative numbers, then $f = \sum_n b_n f_n$ is convex.*

Proposition 2 *If f is convex, and c is any number, then the set of points x such that $f(x) \leq c$ is either empty, a point, an interval, a half-line or the whole line.*

Proposition 3 *If f is convex, $x_1, ..., x_N$, are any numbers, and $t_1, ..., t_N$, are any nonegative numbers such that $\sum_n t_n = 1$, then*

$$f\left(\sum_n t_n x_n\right) \leq \sum_n t_n f(x_n). \tag{2}$$

Proof. The proof is by induction. The proposition is trivially true for $N = 1$, and by the definition of convexity it is true for $N = 2$. Suppose it is true for $N \leq K$. I shall show that it is then true for $N = K + 1$. Let

$$c = \sum_1^K t_n, \quad b_n = \frac{t_n}{c};$$

then, noting that $1 - c = t_{K+1}$, and using the convexity of f,

$$\begin{aligned} f\left(\sum_1^{K+1} t_n x_n\right) &= f\left(c\sum_1^K b_n x_n + t_{K+1}x_{K+1}\right). \\ &\leq cf\left(\sum_1^K b_n x_n\right) + t_{K+1} f\left(x_{K+1}\right). \end{aligned}$$

By the induction hypothesis,

$$f\left(\sum_1^K b_n x_n\right) \leq \sum_1^K b_n f\left(x_n\right),$$

and hence

$$\begin{aligned} cf\left(\sum_1^K b_n x_n\right) + t_{K+1} f\left(x_{K+1}\right) &\leq c\sum_1^K b_n f\left(x_n\right) + t_{K+1} f\left(x_{K+1}\right) \\ &= \sum_1^{K+1} t_n f(x_n), \end{aligned}$$

which completes the proof.

For a function f, define the *left-hand derivative*, $f^-(x)$, at a point x by

$$f^-(x) = \lim_{\substack{h\to 0 \\ h\leq 0}} \frac{f(x+h)-f(x)}{h}. \tag{3}$$

Similarly, define the *right-hand derivative*, $f^+(x)$, at a point x by

$$f^+(x) = \lim_{\substack{h\to 0 \\ h\geq 0}} \frac{f(x+h)-f(x)}{h}. \tag{4}$$

If the left- and right-hand derivatives are equal to each other at a point, then their common value is the *derivative of* f at the point, and one says that the function is *differentiable* there. Recall that a set is *countable* if it can be put into one-to-one correspondence with the integers.

Proposition 4 *(*) If a function is convex on an open interval* (a, b)*, then it is continuous there. Its left- and right-hand derivatives exist at each point of* (a, b)*, and are equal to each other except on a countable set.*

For an example of a convex function that is not differentiable at every point, consider N linear functions, f_n, with different slopes, and let

$$f(x) = \max_n \{f_n(x)\};$$

then f is convex and piece-wise linear, and is not differentiable at the "kinks" where different "pieces" join. (Fig. 2.)

Proposition 5 *If a function is convex on an open interval* (a, b)*, then (1) its left- and right-hand derivatives are each monotone nondecreasing functions in that interval, (2) at each point the left-hand derivative does not exceed the right-hand derivative, and (3)* $x < z$ *implies* $f^+(x) \leq f^-(z)$*.*

Proof. The proof is based on the following lemma.

Lemma 6 *Let* $x < y < z$ *be three points in* (a, b)*; then*

$$\frac{f(y)-f(x)}{y-x} \leq \frac{f(z)-f(x)}{z-x} \leq \frac{f(z)-f(y)}{z-y}. \tag{5}$$

Proof of Lemma. We may write y as a convex combination of x and z as follows:

$$y = \left(\frac{y-x}{z-x}\right) z + \left(\frac{z-y}{z-x}\right) x. \tag{6}$$

By the convexity of f,

$$f(y) \leq \left(\frac{y-x}{z-x}\right) f(z) + \left(\frac{z-y}{z-x}\right) f(x) = \frac{(y-x)f(z) + (z-y)f(x)}{z-x},$$

so

$$\begin{aligned} f(y) - f(x) &\leq \frac{(y-x)[f(z) - f(x)]}{z-x}, \\ \frac{f(y) - f(x)}{y-x} &\leq \frac{f(z) - f(x)}{z-x}, \end{aligned}$$

which verifies the first inequality. A symmetric calculation verifies the second inequality, and completes the proof of the Lemma.

To complete the proof of the proposition, first let $y \to x$; then the first inequality in the Lemma implies that

$$f^+(x) \leq \frac{f(z) - f(x)}{z-x}.$$

Similarly, letting $y \to z$ we get

$$f^-(z) \geq \frac{f(z) - f(x)}{z-x}.$$

[Cf. the definitions (3) and (4).] Hence

$$f^+(x) \leq f^-(z), \tag{7}$$

which proves part (3) of the conclusion. Next, letting $x \to y$ and $z \to y$ we get

$$f^-(y) \leq f^+(y).$$

Since this last holds for any point in (a, b), it also holds at z, so that

$$f^-(z) \leq f^+(z),$$

which verifies part (2) of the conclusion of the theorem. Putting this last together with (7), we have

$$f^+(x) \leq f^+(z).$$

A symmetric argument shows that

$$f^-(x) \leq f^-(z),$$

which completes the proof of the theorem.

The last proposition shows that if a differentiable function is convex on an interval, then its derivative is nondecreasing. The converse is also true.

Proposition 7 *If a function f is differentiable on the interior of an interval, and its derivative is nondecreasing, then it is convex there.*

Corollary 8 *If a function is twice differentiable on the interior of an interval, then it is convex there if and only if its second derivative is nonnegative. Furthermore, if its second derivative is strictly positive on the interior of the interval, then the function is strictly convex there.*

Proof of the Proposition. Let $x < z$ be two points in the interior of the interval, let t be a number such that $0 < t < 1$, and let

$$y = tz + (1-t)x.$$

As in (6) we may write

$$y = \left(\frac{y-x}{z-x}\right) z + \left(\frac{z-y}{z-x}\right) x. \tag{8}$$

Since f is differentiable and nondecreasing,

$$f(y) - f(x) = \int_x^y f'(s)ds \le \int_x^y f'(y)ds = f'(y)(y-x),$$

so

$$\frac{f(y)-f(x)}{y-x} \le f'(y). \tag{9}$$

Similarly,

$$f(z) - f(y) = \int_y^z f'(s)ds \ge \int_y^z f'(y)ds = f'(y)(z-y),$$

and so

$$\frac{f(z)-f(y)}{z-y} \ge f'(y). \tag{10}$$

Hence, by (9) and (10),

$$\frac{f(y)-f(x)}{y-x} \le \frac{f(z)-f(y)}{z-y}.$$

Multiplying the last inequality by $(y-x)(z-y)$, and simplifying the resulting inequality, we get

$$\begin{aligned} (z-x)f(y) &\le (y-x)f(z) + (z-y)f(x), \\ f(y) &\le \left(\frac{y-x}{z-x}\right) f(z) + \left(\frac{z-y}{z-x}\right) f(x), \end{aligned}$$

which last, by (6), can be rewritten in the form,

$$f[tz + (1-t)x] \le tf(z) + (1-t)f(x),$$

thus completing the proof of the proposition.

The proof of the corollary is left as an exercise for the reader. Note, however, that the second derivative of a function can be zero at a point in the interior of an interval, and yet the function can be strictly convex in the interval. An example is provided by the function defined by $f(x) = x^4$, which is strictly convex on the whole real line, and yet $f''(0) = 0$.

1.2 Application to Risk Aversion

Define a *lottery* to be a random variable whose outcomes ("payoffs") are amounts of money. A lottery, say X, is determined by

$$X = [x_1, ..., x_N; p_1, ..., p_N],$$

where N is a positive integer, $x_1, ..., x_N$ are the possible payoffs, and

$$p_n = \Pr\{X = x_n\}, \quad n = 1, ..., N.$$

The *expected payoff* is

$$EX = \sum_{n=1}^{N} p_n x_n. \tag{11}$$

According to a well-known theory of decision-making under uncertainty, if a person is "rational," then her preferences among lotteries can be represented (scaled) by a function u, such that she prefers lottery X to lottery X' if and only if

$$Eu(X) = \sum_{n=1}^{N} p_n u(x_n) > \sum_{n=1}^{N} p'_n u(x'_n) = Eu(X').$$

The mathematical expectation, $Eu(X)$, is called the *expected utility of the lottery* X.

A special case of a lottery is one in which the person receives a fixed payoff, say y, for sure. The person is said to be *averse to risk* (or risk-averse) if she would prefer receiving the amount of money EX for sure to receiving the actual lottery X. This will happen if

$$u(EX) > Eu(X),$$

or

$$u\left(\sum_{n=1}^{N} p_n x_n\right) > \sum_{n=1}^{N} p_n u(x_n). \tag{12}$$

Since the probabilities p_n are nonnegative and their sum is 1, we see that risk-aversion is equivalent to the condition that the utility function u be strictly concave.

[*Questions*: If the person's utility function is linear, then one says that the person is *risk-neutral*. Why is this an apt description? What behavior is exhibited by a person whose utility function is strictly convex?]

1.3 Application to Optimization

In this subsection we characterize where a convex function attains a minimum on a closed interval, $S = [a, b]$, with a and b finite, and $a < b$. Some of the results will carry over immediately to the case of maximizing a concave function. It will also be easy to extend the results to the cases in which S is a half-line or the whole real line. These extensions will be left to the reader. Henceforth, $S = [a, b]$.

A fundamental proposition of real analysis implies that a convex function does attain a minimum on S.

Proposition 9 *(*). If a function is continuous on S, then it attains a minimum at a point in S.*

This Proposition , together with Proposition 4, imply the following corollary.

Corollary 10 *If a function is convex on S, and continuous at the endpoints of S, then it attains a minimum at a point in S.*

We now characterize those points at which a convex function attains a minimum on S. In general, a convex function may attain a minimum at more than one point. However, it follows from Proposition 2 that the set of all such points forms an interval (possibly consisting of a single point.)

Proposition 11 *If f is convex on S, and continuous at the endpoints of S, then it attains a minimum at a point y in S if and only if y satisfies one of the following three conditions:*

$$\begin{aligned} (i) \quad a &< y < b \text{ and } f^-(y) \le 0 \le f^+(y); \\ (ii) \quad y &= a \text{ and } 0 \le f^+(y); \\ (iii) \quad y &= b \text{ and } f^-(y) \le 0. \end{aligned}$$

Proposition 11 is illustrated in Figures 3(i), 3(ii), and 3(iii).

The following corollaries are an immediate consequence of the Proposition and Proposition 5, and are stated without proof.

Corollary 12 *If, in addtion to the hypotheses of the Propositon, f is differentiable, then (i) reduces to*

$$(i') \quad a < y < b \text{ and } f'(y) = 0.$$

Corollary 13 *If, in addtion to the hypotheses of Propositon 11, f is strictly convex, then the minimizing point y is unique.*

Proof of the Proposition. First consider the case in which $a < y < b$. If y satisfies (i), then by Proposition 5,

$$\text{for all } x < y, \quad f^-(x) \le 0 \text{ and } f^+(x) \le 0;$$

hence, $f(x)$ is nonincreasing in x and $f(x) \geq f(y)$ for $x < y$. A symmetric argument shows that $f(z) \geq f(y)$ for $z > y$. Hence y minimizes f on S. Conversely, suppose that y minimizes f on S, and $a < y < b$. If $f^{+}(y) < 0$, then there exists a z such that:

$$\begin{aligned} z &< b, \\ \frac{f(z) - f(y)}{z - y} &< 0, \end{aligned}$$

and hence

$$f(z) < f(y),$$

and so y does not minimize f on S. A symmetric argument shows that if $f^{-}(y) > 0$, then y does not minimize f on S, either. This completes the proof for case (i).

Similar, but one-sided, arguments can be applied to cases (ii) and (iii), completing the proof of Proposition 11.

A further corollary deals with the case in which f is twice-differentiable, i.e., its second derivative is continuous on the interior of S. This corollary provides the justification for the so-called "first-order" and "second-order" conditions for a minimum.

Corollary 14 *Suppose that, in addition to the hypotheses of Proposition 11, f satisfies the following four conditions:*

$$(i) \quad f'' \text{ is continuous on } (a, b),$$

$$(ii) \quad a < y < b,$$

$$\begin{aligned} (iii) \quad f'(y) &= 0, \\ (iv) \quad f''(y) &> 0; \end{aligned}$$

then there exist numbers c, d such that

$$a < c < y < d < b,$$

$$y \text{ is the unique minimizer of } f \text{ on } [c, d].$$

If, in addition, $f''(x) > 0$ on all of (a, b), then y is the unique minimizer of f on $[a, b]$.

Proof of Corollary. Since $f''(y) > 0$ and f'' is continuous, there exist numbers c, d such that

$$\begin{aligned} a &< c < y < d < b, \\ f''(x) &> 0 \text{ for } c < x < d. \end{aligned}$$

Hence, by Corollary 13, y is the unique minimizer of f on $[c, d]$. If $f''(x) > 0$ on all of (a, b), then take $c = a, d = b$, which completes the proof of the Corollary.

In the preceding corollary, $f(y)$ is sometimes called a *local minimum.* Of course, if f is *not* convex everywhere in S, there may be more than one local minimum, and they need not have the same value.

2 Functions of Severable Variables

2.1 Convex Sets and Functions

In this section, some of the results of the previous section are extended to the case of functions of several variables. These results are applied to some problems in the theory of consumer preferences and the theory of cost minimization subject to constraints. In addition to the mathematics used in the previous section, we shall use some very elementary ideas from linear algebra.

Recall that N-dimensional Euclidean space, denoted by $\mathbb{R}^N$, is the set of vectors $x = (x_1, ..., x_N)$, where the coordinates, x_n, are real numbers. In this notation, $\mathbb{R} = \mathbb{R}^1$ is the real line. What follows is a summary of some elementary properties of $\mathbb{R}^N$. If x and y are in $\mathbb{R}^N$, and c is in $\mathbb{R}$, then define $x + y$ to be the vector with coordinates $(x_n + y_n)$, and cx to be the vector with coordinates cx_n. The *distance between x and y* in $\mathbb{R}^N$ is defined by

$$\|x - y\| = \sqrt{\sum_n (x_n - y_n)^2}.$$

The *(open) ball with center x and radius* $r > 0$ is defined by

$$B(x, r) = \left\{ y | y \in \mathbb{R}^N, \|x - y\| < r \right\}.$$

A subset S of $\mathbb{R}^N$ is said to be *open* if for every point x in S there exists a number $r > 0$ such that $B(x, r)$ is entirely contained in S.

A subset S of $\mathbb{R}^N$ is *convex* if for any vectors x, y, in $\mathbb{R}^N$, and any number t, such that $0 \leq t \leq 1$,

$$ty + (1 - t)x \in S.$$

In the rest of this section, the set S will be assumed to be open and convex, unless something is said to the contrary. Note that in the real line, this amounts to assuming that S is an open interval, finite or infinite.

A (real-valued) function f on S is *convex* if, for any vectors x, y, in S, and any number t, such that $0 \leq t \leq 1$,

$$f[ty + (1 - t)x] \leq tf(y) + (1 - t)f(x). \tag{13}$$

Note that this last inequality looks the same as (1) in Section 1, except that x and y are in $\mathbb{R}^N$, not just $\mathbb{R}^1$. Propositions 1 and 3 of Section 3 remain true in this more general setting, and Proposition 2 takes the form:

Proposition 15 *If f is convex, and c is any number, then the set of points x such that $f(x) \leq c$ is a (possibly empty) convex set.*

The proof is essentially the same as for Proposition 2, and is omitted.

Many of the other propositions in Section 1.1 have analogues in $\mathbb{R}^N$. However, a serious treatment of those topics is beyond the scope of these Notes.

Henceforth, unless otherwise noted, it will be assumed that *the functions being considered have continuous second-order partial derivatives on the set S.*

The following proposition is an analogue of Corollary 8 in Section 1.1. It will not be used in these notes, but it is stated without proof for the sake of completeness, because it has many applications in economic theory and econometrics. A symmetric $N \times N$ matrix $Q = ((q_{mn}))$ is said to be *nonnegative semi-definite* if for any vector $x = (x_1, ..., x_N)$ in $\mathbb{R}^N$,

$$\sum_{m,n} q_{mn} x_m x_n \geq 0.$$

In addition, Q is said to be *positive definite* if the preceding inequality is strict for any $x \neq 0$. For a function f defined on S, define $Q(x)$ to be the matrix of its second partial derivatives evaluated at the vector x, i.e., using a standard (if ambiguous) notation, it is the matrix with elements

$$f_{mn}(x) = \frac{\partial^2 f}{\partial x_m \partial x_n}.$$

Note that "f is twice differentiable on S" means that the functions $f_{mn}(x)$ are continuous on S.

Proposition 16 *(*) If f is twice differentiable on S, then f is convex if and only if the matrix $Q(x)$ is nonnegative semi-definite at every point x of S. Furthermore, if $Q(x)$ is everywhere positive definite then f is strictly convex.*

2.2 Application to Optimization

Our first result is an analogue of Corollary 12 in Section 1.3. Recall that the set S is open and convex, and that we are restricting attention to functions f that are twice differentiable. Denote the partial derivative of f with respect to its mth argument by f_m, i.e.

$$f_m(x) = \frac{\partial f}{\partial x_m}.$$

Proposition 17 *If f is convex, then y minimizes f on S if and only if $f_m(y) = 0$ for each $m = 1, ..., N$.*

Lemma 18 *Fix z in S, and let T denote the set of all t such that $tz + (1-t)y$ is in S. For $t \in T$, define*

$$g(t) = f[tz + (1-t)y]; \tag{14}$$

then g is convex.

Proof of Lemma. Note that, since S is open and convex, T is an open interval (possibly infinite). Let t_1 and t_2 be in T, and let α_1 and α_2 be nonnegative numbers such that their sum is 1. Let

$$x = (\alpha_1 t_1 + \alpha_2 t_2)z + (1 - \alpha_1 t_1 - \alpha_2 t_2)y;$$

then

$$g(\alpha_1 t_1 + \alpha_2 t_2) = f(x).$$

One easily verifies that

$$x = \alpha_1[t_1 z + (1 - t_1)y] + \alpha_2[t_2 z + (1 - t_2)y].$$

Hence, since f is convex,

$$\begin{aligned} f(x) &\leq \alpha_1 f[t_1 z + (1 - t_1)y] + \alpha_2 f[t_2 z + (1 - t_2)y] \\ &= \alpha_1 g(t_1) + \alpha_2 g(t_2), \end{aligned}$$

which completes the proof of the Lemma.

To complete the proof of the Proposition, note that

$$g'(t) = \sum_m f_m[tz + (1 - t)y].$$

The rest of the proof now follows from a straightforward application of Corollary 12.

I now turn to the problem of minimizing a convex function f on S subject to a linear constraint. (Two examples from microeconomic theory are presented in subsequent subsections.) Let L be a linear function on S, given by

$$L(x) = \sum_n b_n x_n, \tag{15}$$

where at least one of the coefficients, b_n, is different from zero. We wish to characterize the vectors y that minimize $f(x)$ subject to the constraint that

$$L(x) = \sum_n b_n x_n = c, \tag{16}$$

where c is any given number such that at least one vector in S satisfies the constraint. Denote by K the set of points in S that satisfy the constraint. Note that K is convex (an easy exercise), and since we have assumed that it is not empty, and S is open, it contains more than one point. A vector that minimizes f on K will be called *optimal*.

Proposition 19 *y is optimal if and only if there is a number μ such that*

$$f_n(y) = \mu b_n \text{ for } n = 1, ..., N. \tag{17}$$

Proof. Let y be in K. If y maximizes f on S (not just on K), then by the previous proposition we may take $\mu = 0$. Conversely, if $\mu = 0$ then (again using the previous proposition) y maximizes f on S, and hence on K. Hence, without loss of generality we may suppose that y does not maximize f on S.

If necessary, renumber the coordinates so that $b_N \neq 0$. Solve the constraint for x_N so that

$$x_N = \left(\frac{1}{b_N}\right)\left(c - \sum_{n=1}^{N-1} b_n x_n\right), \tag{18}$$

and define

$$g(x_1, ..., x_{N-1}) = f\left[x_1, ..., x_{N-1}, \left(\frac{1}{b_N}\right)\left(c - \sum_{n=1}^{N-1} b_n x_n\right).\right]$$

Let

$$S' = \{(x_1, ..., x_{N-1}) : \text{for some } x_N, (x_1, ..., x_{N-1}, x_N) \in S\};$$

then S' is open in $\mathbb{R}^{N-1}$, and g is convex and twice differentiable in S'. The first derivatives of g are given by

$$g_n\left(x_1, ..., x_{N-1}\right) = f_n(x) - \left(\frac{b_n}{b_N}\right) f_N(x), n = 1, ..., N-1,$$

where it is understood that in these last equations, x_N is given by (18). The proof is now completed by applying the previous proposition, and taking

$$\mu = \frac{f_N(x)}{b_N}.$$

Remark 1. The optimal vector y and the number μ both depend, in principle, on the parameter c of the constraint equation. In this context, μ has an interesting interpretation. With a slight abuse of notation, rewrite the optimality condition of Proposition 19 as

$$f_n[y(c)] = \mu(c) b_n \text{ for } n = 1, ..., N.$$

The minimum of the function f on the constraint set K is then

$$F(c) = f[y(c)].$$

If the y_n are differentiable with respect to c, then the derivative of F with respect to c is

$$\begin{aligned} F'(c) &= \sum_n f_n[y(c)] y_n'(c) \\ &= \sum_n \mu(c) b_n y_n'(c) \\ &= \mu(c) \sum_n b_n y_n'(c). \end{aligned}$$

Now note that differentiating both sides of the constraint equation (16) with respect to c yields

$$\sum_n b_n y_n'(c) = 1,$$

and so we have the result:

Corollary 20

$$F'(c) = \mu(c). \tag{19}$$

Thus $\mu(c)$ gives us the approximate rate at which a "small change " in the constant c changes the minimum value of the function f.

Remark 2. Define the function $\mathcal{L}$ by

$$\mathcal{L}(y) = f(y) - \mu[L(y) - c].$$

Setting the derivatives of $\mathcal{L}$ with respect to the coordinates of y equal to zero gives us the optimality conditions of the proposition, and setting the derivative with respect to μ equal to zero gives us the constraint equation. This recipe is called the "method of Lagrange," and in this context μ is called the "Lagrange multiplier." The preceding discussion gives a set of conditions under which the recipe gives the right answer.

Remark 3. Note that b_n is the partial derivative of the constraint function L. This suggest how Proposition 19 can be generalized to cover the case of nonlinear functions L that satisfy certain regularity conditions.

2.3 Application to Cost Minimization

Suppose that a firm has several plants that can produce the same commodity (e.g., electricity). The cost of producing a quantity x_n in plant n is $C(x_n)$, so the total cost of producing the vector of outputs $x = (x_1, ..., x_N)$ is

$$C(x) = \sum_n C_n(x_n).$$

The firm wants to produce a total quantity q at minimum cost. Hence it wants to find a vector x that minimizes $C(x)$ subject to the constraint that

$$\sum_n x_n = q.$$

Of course, the output of each plant must be nonnegative. We consider here the case in which at an optimum the output of every firm is strictly positive. Thus take S to be the set defined by the strict inequalities:

$$x_n > 0, \quad n = 1, ..., N.$$

Assume that each cost function C_n is nonnegatve, twice differentiable, strictly increasing, and strictly convex. It follows that C has the same properties.

Note that, if we assume further that $C_n'(0) = 0$, then it is an easy exercise to show that, at the optimum, the output of every plant will be strictly positive. Applying Proposition 19, one immediately gets the optimality condition:

$$C_n'(y_n) = \mu, \quad n = 1, ..., N.$$

The derivative $C_n'(y_n)$ is called the *marginal cost* for plant n, so the optimality condition can be paraphrased: "The total quantity q should be allocated to the plants so as to equalize their marginal costs." From Corollary 20, the value of the Lagrangean multiplier, μ, at the optimum is equal to the derivative of the minimum total cost with respect to the required quantity of total output.

[The reader should consider how the optimality conditions would be modified if at the optimum some plants have zero output. Note that this could happen without the assumption that, for each plant, $C_n'(0) = 0$. The reader should also consider the case in which each plant has a fixed capacity, i.e., an upper bound on output.]

Example. Let

$$C_n(x_n) = c_n x_n^2;$$

then

$$\begin{aligned} C_n'(x_n) &= 2c_n x_n; \\ C_n'(0) &= 0, \end{aligned}$$

and so at an optimal y,

$$\begin{aligned} 2c_n y_n &= \mu, \\ y_n &= \frac{\mu}{2c_n}, \\ q &= \sum_n y_n = \left(\frac{\mu}{2}\right) \sum_n \frac{1}{c_n}, \\ \mu &= \frac{2q}{\sum_n \frac{1}{c_n}}. \end{aligned}$$

2.4 Application to Consumer Choice

Suppose that a consumer must choose a vector x of consumption (in a given period) subject to an expenditure constraint and nonnegativity constraints,

$$\begin{aligned} \sum_n p_n x_n &\leq y, \\ x_n &\geq 0, \quad n = 1, ..., N, \end{aligned}$$

where p_n is the price of commodity n, and y is the maxiumum feasible expenditure. Let $u(x)$ denote the consumer's utility from consuming the vector x, and suppose that the consumer is not satiated in the constraint set, so that she will spend up to the limit y. Assume that the utility function is concave and twice differentiable, so that $-u$ is convex. Suppose, also, that the maximum will occur at a point at which the consumption of every commodity is strictly positive. (Every commodity is a "necessity.") Thus we can take S to be the set defined by the strict inequalities:

$$x_n > 0, \quad n = 1, ..., N,$$

and the consumer's problem is to minimize $-u(x)$ on S subject to the linear constraint,

$$\sum_n p_n x_n = y.$$

Applying Proposition 19, we get the condition that

$$u_n(x) = \mu p_n \text{ for } n = 1, ..., N,$$

or, for all m, n,

$$\frac{u_m(x)}{u_n(x)} = \frac{p_m}{p_n}.$$

The ratio $u_m(x)/u_n(x)$ is called the "marginal rate of substition," i.e., it is the (infinitesimal) rate at which the consumer can substitute commodity m for commodity n, keeping her utility constant. Hence the optimality condition can be paraphrased: "for any two commodities, the marginal rate of substitution equals the price ratio."

Finally, the reader should consider how the optimality conditions would be modified if at the optimal consumption vector the consumption of one or more commodities would be zero.

Exercise. Analyze the example in which

$$u(x) = \sum_n a_n \ln x_n,$$

where the parameters a_n are strictlly positive.

Acknowledgment. I thank A. Radunskaya for helpful comments on an earlier draft. I am, of course, solely responsible for any errors.

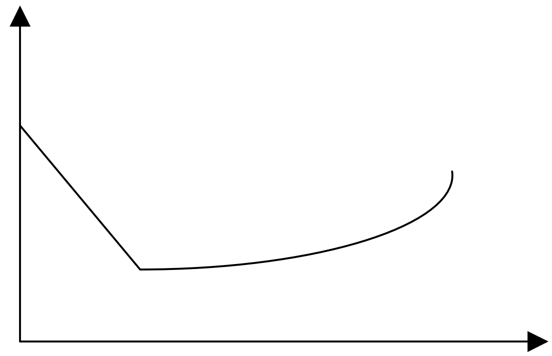

Fig. 1a. A convex function

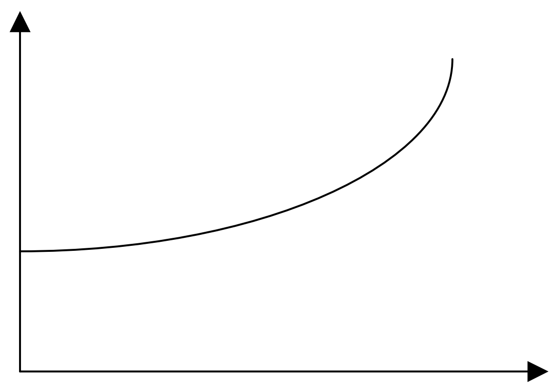

Fig. 1b. A strictly convex function

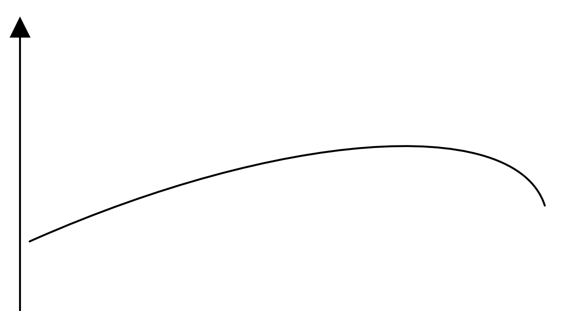

Fig. 1c. A strictly concave function

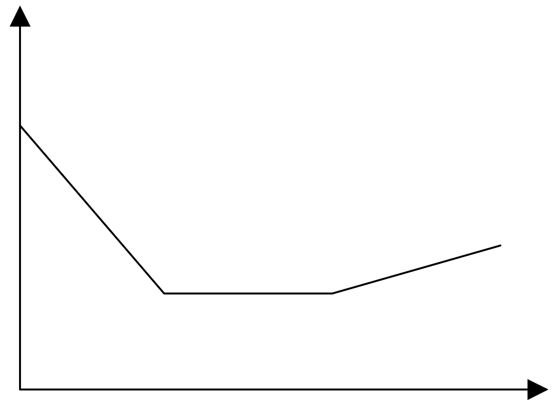

Fig. 2. A piecewise-linear convex function

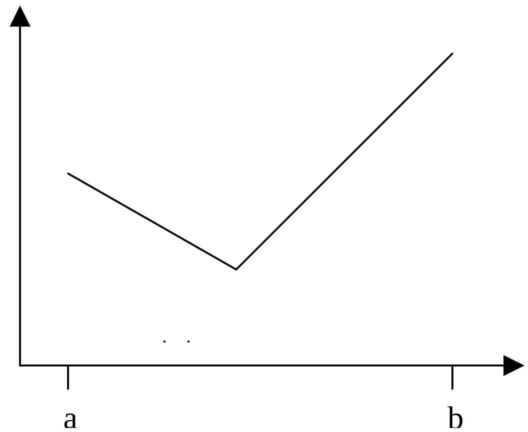

Fig. 3 (i)

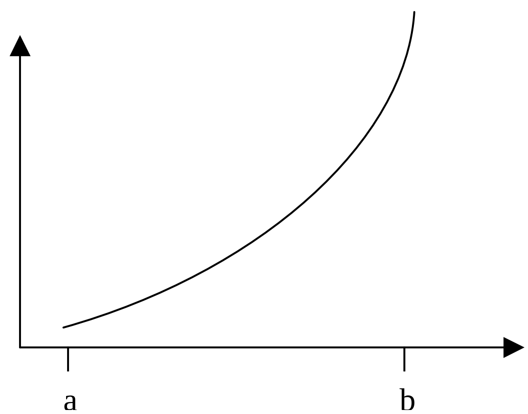

Fig. 3 (ii)

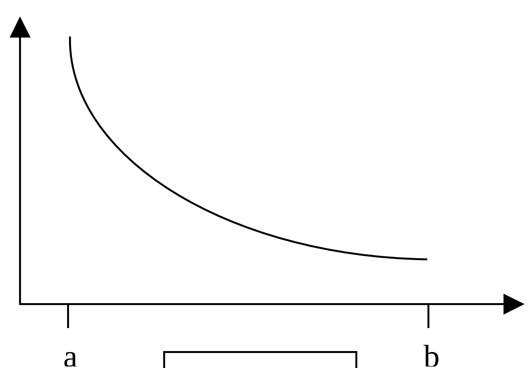

Fig. 3 (iii)