Reprinted from JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS Vol. 19, No 3, September 1967

Convex Programming and Systems of Elementary Monotonic Relations

R. T. ROCKAFELLAR

Department of Mathematics, University of Washington Seattle, Washington 98105

Submitted by Richard Bellman

1. INTRODUCTION

A subset Γ of the real plane $R \times R$ is said to be *totally ordered* if, for each $(x, x^*) \in \Gamma$ and each $(\bar{x}, \bar{x}^*) \in \Gamma$, either one has $x \leqslant \bar{x}$ and $x^* \leqslant \bar{x}^*$, or one has $x \geqslant \bar{x}$ and $x^* \geqslant \bar{x}^*$. It is called a *complete increasing curve*, if it is totally ordered and is not contained in any properly larger totally ordered subset. Such a subset can be described geometrically as an infinite continuous curve which crosses each of the lines with slope -1 exactly once. A complete increasing curve Γ thus resembles the graph of an increasing (i.e., non-decreasing) function on a real interval, except that it may have vertical as well as horizontal segments, some perhaps infinite in length. One is naturally led to treat Γ as a multivalued function by defining

$$\Gamma(x) = \{x^* \mid (x, x^*) \in \Gamma\} \qquad \text{for each} \qquad x. \tag{1.1}$$

The real interval

$$I = \{x \mid \Gamma(x) \neq \phi\} \tag{1.2}$$

is thus called the *domain* of Γ. (For each $x \in I$, $\Gamma(x)$ is a closed real interval; one could show by a classical argument that the interval is trivial, i.e., consists of a single x^*, except for at most countably many values of x.)

Physical situations often arise in which the relationship between two real variables x and x^* is described by a complete increasing curve Γ. In Minty's elegant theory of monotone networks [1], for example, each branch of a given directed linear graph is assigned such a Γ as its *characteristic curve*, i.e., the set of compatible pairs (x, x^*), where x is the current or flow in the branch and x^* is the tension or potential drop across the branch. A notable feature of Minty's theory is that it is applicable to transportation networks as well as to nonlinear electrical and hydraulic networks. This is mainly because it does not insist that the characteristic curves represent functional relations.

The following correspondence between complete increasing curves and certain convex functions is used extensively by Minty in [1]. Given a complete increasing curve Γ with domain I, form a function γ on I by choosing $\gamma(x)$ to be some particular $x^* \in \Gamma(x)$ for each x. Fix any $\bar{x} \in I$ and $c \in R$, and define

$$f(x) = \int_{\bar{x}}^{x} \gamma(t)\, dt + c \qquad \text{for each} \qquad x \in I. \tag{1.3}$$

The definite integral exists in the sense of Riemann because γ is monotone, and it actually does not depend on which particular γ is selected for Γ. (This will be elaborated in Section 2.) We may therefore speak of f as the indefinite *integral* of Γ, symbolically

$$f = \int \Gamma + \text{const.} \tag{1.4}$$

The function f so defined on I is *convex*. Minty does not raise the question of how to reverse this construction, an important question as we shall see in a moment.

There is also the *inverse* Γ^* of Γ, which is defined by

$$\Gamma^* = \{(x^*, x) \mid (x, x^*) \in \Gamma\}. \tag{1.5}$$

(If Γ were the graph of a strictly increasing function, Γ^* would be the graph of the inverse function.) Clearly Γ^* is again a complete increasing curve, and its domain is

$$I^* = \{x^* \mid \Gamma^*(x^*) \neq \phi\} = \{x^* \mid x^* \in \Gamma(x) \text{ for some } x\}, \tag{1.6}$$

the *range* of Γ. Carrying out the above construction for Γ^*, we get a convex function f^* on I^*,

$$f^* = \int \Gamma^* + \text{const.} \tag{1.7}$$

The arbitrary constant of integration can be chosen so that

$$(x, x^*) \in \Gamma \qquad \text{if and only if} \qquad x \in I, \qquad x^* \in I^*, \tag{1.8}$$

and

$$f(x) + f^*(x^*) = xx^*.$$

(See Section 2.)

Henceforth suppose that we are given a family of complete increasing curves Γ_i, with domains I_i and integrals f_i, for $i = 1,..., N$. Let f_i^* on I_i^* be the integral of the inverse Γ_i^* of Γ_i, with constant of integration chosen so that (1.8) holds. Let K and K^* be subspaces of R^N orthogonally complementary to each other. Consider the three problems:

(A) Find vectors $(x_1, \ldots, x_N) \in K$ and $(x_1^*, \ldots, x_N^*) \in K^*$ such that $(x_i, x_i^*) \in \Gamma_i$ for $i = 1, \ldots, N$.

(B) Minimize $f_1(x_1) + \cdots + f_N(x_N)$ subject to $x_i \in I_i$ for $i = 1, \ldots, N$ and $(x_1, \ldots, x_N) \in K$.

(B*) Minimize $f_1^*(x_1^*) + \cdots + f_N^*(x_N^*)$ subject to $x_i^* \in I_i^*$ for $i = 1, \ldots, N$ and $(x_1^*, \ldots, x_N^*) \in K^*$.

Problem (A) involves solving a certain nonlinear system of monotonic relations, while (B) and (B*) are certain convex programs with linear constraints. These problems were studied by Minty in [1] in the case where Γ_i is the characteristic curve of the ith branch of a monotone network, K is the space of flows (circulations) in the network and K^* is the corresponding space of tensions. (If E is the node-versus-branch signed incidence matrix of the given directed graph with N branches, K is the subspace of R^N consisting of the vectors orthogonal to the rows of E, whereas K^* is the subspace spanned by the rows of E). In applications to electrical networks, say, interest centers on solving (A), and results about (B) and (B*) correspond to well-known variational principles. In applications to networks of the sort arising in operations research, one usually starts from an extremum problem like (B) and works with a certain duality between (B) and (B*).

In the general case of a subspace K not necessarily arising from a network, (B) represents quite a broad class of problems. The class includes, for instance, all linear programs and quadratic programs (with linear constraints), as will be explained in detail in Section 4. It will also be shown in Section 4 that (B) and (B*) can be reformulated as a pair of dual convex programs which fit into the duality scheme we have set forth in [2].

The following facts about (A), (B), and (B*) will be corollaries of deeper results proved in Section 3.

THEOREM 1. (Characterization Theorem.) *A pair of vectors* $(x_1, \ldots, x_N)$ *and* $(x_1^*, \ldots, x_N^*)$ *solves* (A) *if and only if* $(x_1, \ldots, x_N)$ *solves* (B) *and* $(x_1^*, \ldots, x_N^*)$ *solves* (B*).

THEOREM 2. (Duality Theorem.) *Problem* (B) *has a solution if and only if problem* (B*) *has a solution, in which case the minimum value in* (B) *and the minimum value in* (B^*) *have the same magnitude but the opposite sign.*

THEOREM 3. (Existence Theorem.) *If there are vectors* $(x_1, \ldots, x_N) \in K$ *and* $(x_1^*, \ldots, x_N^*) \in K^*$ *such that* $x_i \in I_i$ *and* $x_i^* \in I_i^*$ *for* $i = 1, \ldots, N$, *then* (A), (B) *and* (B*) *all have solutions.*

Observe from Theorem 1 and Theorem 2 that neither (B) nor (B*) can have a solution unless (A) has a solution, in which case the condition in Theorem 3 is certainly satisfied. Thus the condition in Theorem 3 is also necessary for the existence of solutions to any of the three problems. Here is another immediate consequence of Theorem 1 and Theorem 2.

THEOREM 1'. *A vector* $(x_1, \ldots, x_N)$ *solves* (B) *if and only if there exists a vector* $(x_1^*, \ldots, x_N^*)$ *such that* $(x_1, \ldots, x_N)$ *and* $(x_1^*, \ldots, x_N^*)$ *solve* (A).

In view of the symmetry between (B) and (B*), Theorem 1' can be used in turn to derive Theorem 1 and the first part of Theorem 2.

Minty first proved Theorem 3 in [1] under the assumption that the subspaces K and K^* arise from a network as above. He also showed that, if $(x_1, \ldots, x_N)$ and $(x_1^*, \ldots, x_N^*)$ solve (A), then these vectors solve (B) and (B*), respectively, and the minima in (B) and (B*) have the same magnitude and the opposite sign. (This is a weaker version of Theorems 1 and 2.) The proofs are very graph-theoretic, but constructive. They are valid without change in the case of subspaces corresponding to a "digraphoid" rather than a directed graph, as was pointed out by Minty in an appendix to [3].

A weaker form of Theorem 1', is stated by Berge [4, Chap. 2], in terms of monotone networks only, but the proof does not involve graph theory in any essential way. It is not altogether clear from the hypothesis, but it seems that Berge requires the domain intervals I_i to be closed. At least some such assumption must have been in mind, since Berge applies to the minimand in (B) (which is given only on the product of the intervals I_i) a version of the Kuhn-Tucker theorem in which the functions are supposed to be defined on all of R^N. It could be proved that, if the intervals I_i are closed, the minimand in (B) can be extended to be a convex function on all of R^N, so that this Kuhn-Tucker application is justified. Assuming that the intervals I_i^* are closed too, one can get Theorems 1 and 2 this way. To have I_i and I_i^* closed, however, means that Γ_i has neither a vertical nor a horizontal asymptote. This excludes many obviously important curves Γ, such as those yielding $f(x) = 1/x$ on $I = R_+ = \{x \mid x > 0\}$, or $f(x) = -\sqrt{x}$ on $I = R_+$, or $f(x) = -\log x$ on $I = R_+$, or $f(x) = e^x$ on $I = R$.

Berge does not consider any existence theorems like Theorem 3 in [4].

The first proof of Theorem 3 in the general (non-network) case was given by the present author in his dissertation [5, Chap. 5]. Theorems 1 and 2 were also proved in [5] along with more general results along the lines of Theorem 4 (to be introduced in Section 3). The results in [5] are stated in the convex programming form described below in Section 4.

More recently, Minty [6] has given a general proof for a weaker form of Theorem 3 in which x_i and x_i^* are required to lie in the interiors of I_i and I_i^*. Camion [7] has proved that the condition in Theorem 3 is necessary and sufficient for (B) to have a solution. The sufficiency proof is an extension of Minty's original constructive proof. It furnishes an algorithm for solving (B) approximately, starting from any pair of vectors satisfying the condition in Theorem 3. To establish the necessity, Camion invokes Theorem 1′ as proved by Berge, but he does not demonstrate that the Kuhn-Tucker theorem can be extended to cover Berge's argument in this general case.

Theorems 1, 2, and 3 will be deduced below from two theorems about conjugate convex functions, Theorems 4 and 5, which will be given a joint inductive proof. Theorem 4 describes the duality between two extensions of (B) and (B*).

The elementary theory of convex functions of one variable is recounted in Section 2 for background. This theory in particular characterizes the pairs f_i, I_i, which are admissible in (B). It shows how to construct problems (B*) and (A) starting from (B) (rather than from given curves Γ_i).

2. Convex Functions on the Real Line

Many of the facts about convex functions on R can be deduced at once from the established multi-dimensional theory of convex sets and functions [8]. The concepts and arguments are usually simpler, however, in the one-dimensional case. There are also some results not generally true in R^N, which are not well-known and yet will be especially important to us. Since one of our objectives here is to extend linear programming without relying on a "nonelementary" technical background, it makes sense for us to give a self-contained outline here of the theory of convex functions on the real line. Proofs will be omitted where they are easy exercises, depending perhaps on some elementary classical trick.

In harmony with our terminology elsewhere, we define a *proper convex function* on R to be an everywhere-defined function f with values $-\infty < f(x) \leqslant +\infty$, not identically $+\infty$, such that

$$f(\lambda x + (1-\lambda) y) \leqslant \lambda f(x) + (1-\lambda) f(y) \qquad \text{when} \qquad 0 < \lambda < 1. \tag{2.1}$$

(Improper convex functions, not to be discussed here, can have the value $-\infty$.) The set of points where such a function f is finite is evidently a non-empty interval (i.e., a connected set of real numbers); we call it the *effective domain* of f.

A proper convex function f on R is always continuous, except possibly at end-points of its effective domain. (By an *end-point* of an interval, we mean a finite end-point of the closure of the interval.) It is said to be a *closed* function if each of the intervals $\{x \mid f(x) \leqslant \mu\}$, $\mu \in R$, is closed. Equivalently, f is closed if and only if it is actually continuous relative to the closure of its effective domain. This is a constructive property. If f is not already closed, it can be made so in a unique way by lowering its values suitably at the end-points of its effective domain. (The resulting function may have a slightly larger effective domain.) All the closed proper convex functions on R can thus be constructed as follows. On the one hand, there is the trivial case where f has a specified finite value at a certain point $\bar{x}$, but has the value $+\infty$ at every point other than $\bar{x}$. In the more interesting case, we take an arbitrary nonempty open interval I_0 and any finite convex function f on I_0, extend f to the closure of I_0 by taking limits, and give f the value $+\infty$ outside the closure of I_0. (The limit of $f(x)$ at the end-points of I_0 always exists, and it is either finite or $+\infty$.) Then I_0 is the interior of the effective domain of f.

Let γ be an increasing function from R to $[-\infty, +\infty]$, and let $\bar{x}$ be a point where γ is finite. The formula

$$f(x) = \int_{\bar{x}}^{x} \gamma(t)\, dt + \text{const.} \tag{2.2}$$

then defines a closed proper convex function on R. Namely, $f(x)$ exists as a Riemann integral in the interval where γ is finite, and as an improper Riemann integral (limit of ordinary integrals) at the remaining end-points of this interval. The natural interpretation of the integral elsewhere is $+\infty$. The closedness and convexity of f are easy consequences of the monotonocity of γ and the continuity and additivity of Riemann integrals.

We can also construct a complete increasing curve Γ from any such γ by taking

$$\Gamma(x) = \{x^* \in R \mid \lim_{z \uparrow x} \gamma(z) \leqslant x^* \leqslant \lim_{z \downarrow x} \gamma(z)\}. \tag{2.3}$$

Evidently the converse is true: each complete increasing curve Γ arises this way from a somewhere-finite increasing function γ on R. Thus a closed proper convex function f on R can be constructed from any complete increasing curve Γ, via (2.2) and some γ representing Γ as in (2.3). Of course γ need not be unique, since $\gamma(x)$ can be any number in $\Gamma(x)$. However, two increasing functions representing the same Γ have the same points of continuity, and they must agree at those points. Since an increasing function is continuous except at countably many points, it follows that f depends only on Γ and the constant of integration, and not on the particular γ used in the construction. (This is not an "elementary" argument. An easier way of establishing the uniqueness will appear below.)

Observe that the above construction yields an f defined on all of R, not just on the domain I of Γ as in the introduction. Outside the closure of I, $f(x)$ is $+\infty$, but it might be finite at end-points not belonging to I itself.

In order to reverse a generalized kind of integration, one needs a generalized differentiation. Let f be any proper convex function on R. At each x in the effective domain of f, the left and right derivatives

$$f'_-(x) = \lim_{z \uparrow x} (f(z) - f(x))/(z - x)$$

$$f'_+(x) = \lim_{z \downarrow x} (f(z) - f(x))/(z - x) \tag{2.4}$$

exist (although they may be infinite), as is well-known. Both f'_- and f'_+ are finite at interior points of the effective domain. It is convenient to define them both as $-\infty$ to the left of the effective domain, and both as $+\infty$ to the right of the effective domain. Then

$$f'_+(z_1) \leqslant f'_-(x) \leqslant f'_+(x) \leqslant f'_-(z_2) \qquad \text{when} \qquad z_1 < x < z_2 . \tag{2.5}$$

In particular, f'_- and f'_+ are increasing functions on R.

We define the *generalized derivative* of a proper convex function f to be the *multivalued* function f', where

$$f'(x) = \{x^* \in R \mid f'_-(x) \leqslant x^* \leqslant f'_+(x)\} \qquad \text{for each} \qquad x. \tag{2.6}$$

At any point x where f is differentiable in the ordinary sense, $f'(x)$ reduces to a single number, the ordinary derivative of f at x. The *domain* of f',

$$I = \{x \mid f'(x) \neq \phi\} = \{x \mid f'_-(x) < +\infty \text{ and } f'_+(x) > -\infty\}, \tag{2.7}$$

is a nonempty interval contained in the effective domain of f and containing in turn the interior of the effective domain.

The main result about generalized derivatives is the following version of the "fundamental theorem of the calculus." The reader may be interested to know that an analogous result has also been proved in [9] for convex functions on R^N.

LEMMA 1. *If f is a closed proper convex function on R, its generalized derivative f' is a complete increasing curve. Conversely, for each complete increasing curve Γ one has*

$$\Gamma = f'$$

for some closed proper convex function f on R, and f is unique up to an additive constant. In fact

$$f = \int \Gamma + \text{const.}$$

in the sense that formula (2.2) holds for any somewhere-finite increasing function γ representing Γ as in (2.3).

PROOF. When f is a closed proper convex function, the formulas

$$\lim_{z\uparrow x} f'_-(z) = f'_-(x), \qquad \lim_{z\downarrow x} f'_-(z) = f'_+(z),$$
$$\lim_{z\uparrow x} f'_+(z) = f'_-(x), \qquad \lim_{z\downarrow x} f'_+(z) = f'_+(x), \tag{2.8}$$

hold for every $x \in R$, as is easily deduced from (2.5) and the definition of "closed." Thus, if we take any somewhere-finite function γ between f'_- and f'_+, (which will automatically be increasing according to (2.5)), f' will coincide with the complete increasing curve Γ defined by (2.3). Conversely, given Γ and any γ representing it as in (2.3), let f be the closed proper convex function defined by (2.2). Then evidently

$$f'_-(x) = \lim_{z\uparrow x} \gamma(z) \qquad \text{and} \qquad f'_+(x) = \lim_{z\downarrow x} \gamma(z), \tag{2.9}$$

and hence $f' = \Gamma$. It remains now to show that, if f_1 and f_2 are closed proper convex functions on R such that $f'_1(x) = f'_2(x)$ for every x, then $f_2 = f_1 + \text{const.}$ This is trivially true when the effective domain of f_1 consists of only one point. Otherwise the effective domain of f_1 has a nonempty interior I_0, and this consists of the points x for which the interval $f'_1(x)$ is bounded. Then I_0 has to be the interior of the effective domain of f_2, too. On I_0, $f_2 - f_1$ is actually differentiable, indeed its left and right derivatives both equal zero at each point. Therefore $f_2 = f_1 + \text{const.}$ on I_0. This must hold on the closure of I_0 as well, since f_1 and f_2 are closed, and hence it holds throughout R.

COROLLARY. *Let I be a nonempty interval and let f be a finite convex function given on I. In order that there exist a complete increasing curve Γ, such that I is the domain of Γ and f is the restriction to I of $\int \Gamma$, it is necessary and sufficient that f be the restriction of a closed proper convex function on R to the domain I of its generalized derivative. In other words, aside from the trivial case where I consists of a single point, the condition is that f has to have a finite one-sided derivative at any end-point included in I, but the one-sided derivatives must become infinite as one nears an end-point not included in I.*

We shall now develop the one-dimensional case of Fenchel's notion of *conjugacy* [7]. The novel feature of our approach is that we avoid having to use separation theorems for convex sets in R^2. Instead, we rely on Lemma 1 and the following elementary fact: if f is a proper convex function on R, then

$$f \text{ achieves its minimum at } x \text{ if and only if}$$
$$0 \in f'(x), \qquad \text{i.e.,} \qquad f'_-(x) \leqslant 0 \qquad \text{and} \qquad f'_+(x) \geqslant 0. \tag{2.10}$$

Notice that this implies f always achieves a minimum somewhere, unless $f'_+(x) < 0$ for all x (f is "strictly decreasing") or $f'_-(x) > 0$ for all x (f is "strictly increasing"). In particular, no outright compactness argument is needed in proving that a continuous *convex* function on a closed interval achieves its minimum.

LEMMA 2. *Let f be any closed proper convex function on R, and let f^* be the conjugate function defined by*

$$f^*(x^*) = -\inf_x \{f(x) - xx^*\} \qquad \textit{for each} \qquad x^*. \tag{2.11}$$

Then f^* is again a closed proper convex function on R, and its conjugate is in turn f, i.e.,

$$f(x) = -\inf_{x^*} \{f^*(x^*) - xx^*\} \qquad \textit{for each} \qquad x. \tag{2.12}$$

Furthermore, the three conditions

$$x^* \in f'(x), \qquad x \in f^{*\prime}(x^*), \qquad \textit{and} \qquad f(x) + f^*(x^*) = xx^*,$$

are equivalent. Thus

$$f^{*\prime} = f'^*, \tag{2.13}$$

PROOF. Applying (2.10) to $h(x) = f(x) - xx^*$ in place of f, we see in (2.11) that

$$-f^*(x^*) = f(x) - xx^* \qquad \text{if and only if} \qquad x^* \in f'(x). \tag{2.14}$$

In particular, f^* is not identically $+\infty$. Trivially, $f^*(x^*) > -\infty$ for all x, because f itself is not identically $+\infty$. Furthermore, (2.11) expresses f^* as a supremum of affine functions (one for each x in the effective domain of f), so f^* satisfies the convexity and closure conditions. Hence f^* is a closed proper convex function. By definition of f^*, for a fixed x we have

$$f^*(x^*) - xx^* \geqslant -f(x) \qquad \text{for every} \qquad x^*.$$

Thus $f^*(x^*) - xx^*$ is sure to be at its minimum when the left half of (2.14) holds. Consequently $x \in f^{*\prime}(x^*)$ whenever $x^* \in f'(x)$, in other words $f^{*\prime}$ is an extension of f'^*, the complete increasing curve obtained by reflecting f' across the line $x^* = x$. Since complete increasing curves by definition cannot be properly extended, it follows that actually $f^{*\prime} = f'^*$, so that the three cited conditions are equivalent as asserted. Observe that f^* is the *only* closed proper convex function for which the equivalence holds, because f^*

is determined up to a constant as the integral of f'^*, and the constant is fixed by (2.14). The uniqueness and symmetry imply that f is in turn the conjugate of f^*.

COROLLARY. *Given any complete increasing curve Γ and $f = \int \Gamma$, let f^* be the conjugate of f. Then $f^* = \int \Gamma^*$, and*

$$(x, x^*) \in \Gamma \qquad \text{if and only if} \qquad f(x) + f^*(x^*) = xx^*. \tag{2.15}$$

REMARK. Starting from a convex function f on an interval I satisfying the condition in the corollary to Lemma 1, one can evidently construct the corresponding function f^* on an interval I^* directly as follows. For each $x \in I$, let f^* coincide with the affine function $xx^* - f(x)$ on the interval $\{x^* \in R \mid f'_-(x) \leqslant x^* \leqslant f'_+(x)\}$. The union of these intervals will be I^*. This is a simple generalization of the Legendre transform.

Certain combinatorial operations are useful in theory as well as in practice. Suppose f_1 and f_2 are closed proper convex functions on R. Then so is $f_1 + f_2$, unless it is identically $+\infty$. Likewise, if Γ_1 and Γ_2 are complete increasing curves, so is $\Gamma_1 + \Gamma_2$, where

$$\Gamma_1 + \Gamma_2 = \{(x, x_1^* + x_2^*) \mid x_1^* \in \Gamma_1(x) \text{ and } x_2^* \in \Gamma_2(x)\}, \tag{2.16}$$

provided the domains of Γ_1 and Γ_2 have a point in common. This follows from Lemma 1 and the additivity of left and right derivatives:

$$(f_1 + f_2)' = f_1' + f_2' \tag{2.17}$$

if $f_1 + f_2$ is not identically $+\infty$. Hence also

$$\int (\Gamma_1 + \Gamma_2) = \int \Gamma_1 + \int \Gamma_2 \tag{2.18}$$

if $\Gamma_1 + \Gamma_2$ is not empty.

Another interesting operation is the *convolution* defined by

$$(f_1 \mathbin{\Box} f_2)(x) = \inf_z \{f_1(x - z) + f_2(z)\} \qquad \text{for each} \qquad x. \tag{2.19}$$

It is easy to show that $f_1 \mathbin{\Box} f_2$ is again a closed proper convex function, unless it is identically $-\infty$. Convolution and addition are dual to each other with respect to taking conjugates:

$$(f_1 + f_2)^* = f_1^* \mathbin{\Box} f_2^* \qquad \text{and} \qquad (f_1 \mathbin{\Box} f_2)^* = f_1^* + f_2^*. \tag{2.20}$$

(The first formula follows by duality from the second, which can be established by direct computation using definition (2.11).) The related operation for complete monotone curves is given by

$$\Gamma_1 \square \Gamma_2 = \{(x_1 + x_2, x^*) \mid x^* \in \Gamma_1(x_1) \text{ and } x^* \in \Gamma_2(x_2)\}. \tag{2.21}$$

We call this operation *inverse addition*, because trivially

$$(\Gamma_1 + \Gamma_2)^* = \Gamma_1^* \square \Gamma_2^* \qquad \text{and} \qquad (\Gamma_1 \square \Gamma_2)^* = \Gamma_1^* + \Gamma_2^*. \tag{2.22}$$

The fact that $\Gamma_1 \square \Gamma_2$ is another complete increasing curve, unless it is empty, is apparent from (2.22). The formulas

$$(f_1 \square f_2)' = f_1' \square f_2', \tag{2.23}$$

$$\int (\Gamma_1 \square \Gamma_2) = \int \Gamma_1 \square \int \Gamma_2 \tag{2.24}$$

follow from (2.17) and (2.18) by duality.

A function g on R is called a *closed proper concave function* if $-g$ is a closed proper convex function. The theory of such functions can be developed in obvious analogy with the above: $-\infty$, $\geqslant$ and "sup" will play the roles of $+\infty$, $\leqslant$ and "inf". In particular, the *conjugate* relationship for concave functions is given by

$$\begin{aligned} g^*(x^*) &= -\sup_x \{g(x) - xx^*\} \qquad \text{for each} \quad x^*, \\ g(x) &= -\sup_{x^*} \{g^*(x^*) - xx^*\} \qquad \text{for each} \quad x. \end{aligned} \tag{2.25}$$

The *generalized derivatives* of the closed proper concave functions g on R, given by

$$g'(x) = \{x^* \in R \mid g_+'(x) \leqslant x^* \leqslant g_-'(x)\} \tag{2.26}$$

are the *complete decreasing curves* Δ (which are the reflections of the complete increasing curves across the horizontal axis).

A key result involving both convex and concave functions is the following special version of Fenchel's Duality Theorem [8].

LEMMA 3. *Let f be a closed proper convex function on R with conjugate f^*, and let g be a closed proper concave function on R with conjugate g^*. Then*

$$\inf_x \{f(x) - g(x)\} = \sup_{x^*} \{g^*(x^*) - f^*(x^*)\}. \tag{2.27}$$

Furthermore, let I, I^*, J, J^*, denote the domains of the generalized derivatives of f, f^*, g, g^*. The infimum above is attained if and only if $I^* \cap J^* \neq \phi$, while the supremum is attained if and only if $I \cap J \neq \phi$.

PROOF. Let $h(x) = -g(x)$. Then $h^*(x^*) = -g^*(-x^*)$. One can therefore re-express (2.27) as

$$\inf_x \{(f + h)(x) - x \cdot 0\} = -(f^* \,\square\, h^*)(0).$$

Now, as seen above, either $f + h$ is identically $+\infty$ and $f^* \,\square\, h^*$ is identically $-\infty$, or the two are closed proper convex functions conjugate to each other. In the former case the equation is trivial, while in the latter case it is true by the definition of "conjugate." If $f - g$ is identically $+\infty$, its minimum is trivially attained. Otherwise, $f - g$ is a closed proper convex function and consequently attains its minimum except when $(f - g)'_+(x) < 0$ for all x, or when $(f - g)'_-(x) > 0$ for all x. Altogether, therefore, $f - g$ attains its minimum except when

$$f'_+(x) < g'_+(x) \qquad \text{for every} \qquad x, \tag{2.28}$$

or when

$$g'_-(x) < f'_-(x) \qquad \text{for every} \qquad x. \tag{2.29}$$

The generalized derivatives of f^* and g^* are merely the inverses of those of f and g, so that

$$I^* = \{x^* \in R \mid f'_-(x) \leqslant x^* \leqslant f'_+(x) \text{ for some } x\},$$

$$J^* = \{x^* \in R \mid g'_+(x) \leqslant x^* \leqslant g'_-(x) \text{ for some } x\}.$$

Since f'_- and f'_+ are increasing, and g'_- and g'_+ are decreasing, condition (2.28) means that J^* lies entirely to the right of I^*, while (2.29) means that J^* lies entirely to the left of I^*. Therefore, the infimum in (2.27) is unattained if and only if I^* and J^* fail to overlap. The assertion about the supremum in (2.27) follows by duality.

3. THE BASIC THEOREMS

The elements introduced in Section 1 can be viewed in a better light, now that the elementary facts and definitions in Section 2 are at our disposal. We assume for $i = 1, \ldots, N$ that Γ_i is a complete increasing curve and f_i is a closed proper convex function on R, such that $f'_i = \Gamma_i$, $\int \Gamma_i = f_i$. We let f_i^* be conjugate of f_i, so that $f_i^{*\prime} = \Gamma_i^*$, $\int \Gamma_i^* = f_i^*$, and

$$(x_i, x_i^*) \in \Gamma_i \qquad \text{if and only if} \qquad f_i(x_i) + f_i^*(x_i^*) = x_i x_i^*. \tag{3.1}$$

The domain of Γ_i is denoted by I_i, while the domain of the inverse curve

Γ_i^* (the range of Γ_i) is denoted by I_i^*. Finally, we assume that K is an arbitrary subspace of R^N, and that K^* is its orthogonal complement. This notation will be in effect throughout this section.

Theorems 1, 2, and 3 will be proved below using a further theorem, which treats the slightly more general problem where the functions f_i are not restricted to the domains I_i of their generalized derivatives.

THEOREM 4. *If*

$$\inf\{f_1(x_1) + \cdots + f_N(x_N) \mid (x_1, \ldots, x_N) \in K\} \tag{3.2}$$

is not $+\infty$, *then it is the negative of*

$$\inf\{f_1^*(x_1^*) + \cdots + f_N^*(x_N^*) \mid (x_1^*, \ldots, x_N^*) \in K^*\}, \tag{3.3}$$

and it is attained if and only if there exists some $(x_1^*, \ldots, x_N^*) \in K^*$ *such that* $x_i^* \in I_i^*$ *for* $i = 1, \ldots, N$. *In order that* $(x_1, \ldots, x_N) \in K$ *and* $(x_1^*, \ldots, x_N^*) \in K^*$ *be points where the respective infima are finitely attained, it is necessary and sufficient that* $(x_i, x^*) \in \Gamma_i$ *for* $i = 1, \ldots, N$.

In establishing this theorem we shall also automatically establish the result below, which has an interesting corollary. (Here we use the convention that the infimum of an empty set of numbers is $+\infty$.)

THEOREM 5. *Assume*

$$g(x_1) = -\inf(f_2(x_2) + \cdots + f_N(x_N) \mid (x_1, x_2, \ldots, x_N) \in K\} \tag{3.4}$$

is finite for at least one x_1. *Then* g *is a closed proper concave function on* R *whose conjugate is given by*

$$g^*(x_1^*) = -\inf\{f_2^*(x_2^*) + \cdots + f_N^*(x_N^*) \mid (x_1^*, x_2^*, \ldots, x_N^*) \in K^*\}. \tag{3.5}$$

Furthermore, the domain J *of the generalized derivative of* g *is*

$$\{x_1 \mid (x_1, x_2, \ldots, x_N) \in K \text{ for some } x_2 \in I_2, \ldots, x_N \in I_N\} \tag{3.6}$$

it this set is nonempty. In the nonempty case the infimum in (3.5) is attained for every x^*, *whereas in the empty case the infimum is attained only trivially when it is* $+\infty$. *Finally, if the set in (3.6) and the set*

$$\{x_1^* \mid (x_1^*, x_2^*, \ldots, x_N^*) \in K^* \textit{ for some } x_2^* \in I_2^*, \ldots, x_N^* \in I_N^*\} \tag{3.7}$$

are both nonempty, g *is indeed finite somewhere and the generalized derivative of* g *is*

$$\Delta = \{(x_1, x_1^*) \mid \textit{there exists some } (x_1, x_2, \ldots, x_N) \in K \textit{ and } (x_1^*, x_2^*, \ldots, x_N^*) \in K^* \\ \textit{such that } (x_2, x_2^*) \in \Gamma_2, \ldots, (x_N, x_N^*) \in \Gamma_N\}. \tag{3.8}$$

COROLLARY. *If the set Δ in (3.8) is not empty, it is a complete decreasing curve.*

REMARK. Minty [1] has already proved this corollary in the network-theoretic case of K and K^*, where it has an important interpretation. Suppose one is given a monotone network with two distinguished nodes, the "input node" and the "output node." Construct a new network by adding a "return branch" (labeled as branch 1 for convenience) from the output node back to the input node. Each circulation $(x_1, x_2, \ldots, x_N) \in K$ in the augmented network corresponds to a flow of x_1 from the input to the output of the original network. Similarly, each tension $(x_1^*, x_2^*, \ldots, x_N^*) \in K^*$ corresponds to a potential drop of $-x_1^*$ from input to output. Thus, if we want to lump the original network together, its characteristic curve as a whole will be

$$\{(x_1, -x_1^*) \mid (x_1, x_1^*) \in \Delta\}.$$

This is another *complete increasing curve* according to the Corollary.

JOINT PROOF OF THEOREMS 4 AND 5. First of all, Theorem 4 is true when $N = 1$, where it is the special case of Lemma 3 with $g(x) = 0$ for $x \in K$ and $g(x) = -\infty$ for $x \notin K$. (The attainment conditions then reduce to (2.10) or its dual.) Theorem 5 is vacuous when $N = 1$.

Assume now that the given value of N is bigger than 1, and that Theorem 4 has been verified for all smaller values of N. We shall prove that then Theorem 5 is true for the given N. This will be shown to imply in turn that Theorem 4 is true for the given N. Theorems 4 and 5 will thus be true for all N by induction.

Let $h(x_1^*)$ denote the right side of (3.5). We begin by demonstrating that

$$g(x_1) = -\sup_{x_1^*} \{h(x_1^*) - x_1 x_1^*\}, \tag{3.9}$$

except possibly when $g(x_1)$ is $-\infty$ and the right side is $+\infty$. Fix any x_1. The right side of (3.9) can be expressed as

$$\inf \{x_1 x_1^* + f_2^*(x_2^*) + \cdots + f_N^*(x_N^*) \mid (x_1^*, \ldots, x_N^*) \in K^*\}. \tag{3.10}$$

We can suppose that there exist constants $c_2, \ldots, c_N$ such that $(x_1, c_2, \ldots, c_N) \in K$, since otherwise $g(x_1)$ is $-\infty$ while (3.10) is $-\infty$ or $+\infty$. Then

$$x_1 x_1^* = -c_2 x_2^* - \cdots - c_N x_N^* \quad \text{when} \quad (x_1^*, \ldots, x_N^*) \in K^* \tag{3.11}$$

by the orthogonality of K and K^*.

Consider the closed proper convex functions h_i and the subspace K_1 of R^{N-1} defined by

$$h_i(x_i) = f_i(x_i + c_i) \qquad \text{for} \qquad i = 2,..., N,$$
$$K_1 = \{(x_2 ,..., x_N) \mid (0, x_2 ,..., x_N) \in K\}.$$

Obviously

$$g(x_1) = - \inf \{h_2(x_2) + \cdots + h_N(x_N) \mid (x_2 ,..., x_N) \in K_1\}. \qquad (3.12)$$

On the other hand, the conjugate h_i^* of h_i and the orthogonal complement K_1^* of K_1 in R^{N-1} are given by

$$h_i^*(x_i^*) = f_i^*(x_i^*) - c_i x_i^* \qquad \text{for} \qquad i = 2,..., N,$$

$$K_1^* = \{(x_2^*,..., x_N^*) \mid (x_1^*, x_2^*,..., x_N^*) \in K^* \text{ for some } x_1 \}.$$

Hence, on applying (3.11) to (3.10), we get

$$- \sup_{x_1^*} \{h(x_1^*) - x_1 x_1^*\} = \inf \{h_2^*(x_2^*) + \cdots + h_N^*(x_N^*) \mid (x_2^*,..., x_N^*) \in K_1^*\}. \qquad (3.13)$$

Theorem 4 (and its dual) are valid by hypothesis in R^{N-1}, so the right sides of (3.12) and (3.13) are equal except when the first is $-\infty$ and the second is $+\infty$. This is what we wanted first to demonstrate.

The significant thing about the fact just proved is that the right side of (3.9) cannot be $+\infty$ unless h is identically $-\infty$, in which case the right side is $+\infty$ for *every* x_1. Inasmuch as $g(x_1)$ is finite for at least one x_1 by the hypothesis of Theorem 5, it follows that for every x_1 (3.9) holds and $g(x_1) \neq +\infty$. This implies further that $h(x_1^*)$ is finite for at least one x_1^* and that $h(x_1^*) < +\infty$ for every x_1^*. The function $-g$, which is not identically $+\infty$ and which nowhere has the value $-\infty$, is expressed by (3.9) as a supremum of affine functions on R (one for each x_1^* such that $h(x_1^*)$ is finite). Hence $-g$ is a closed proper convex function, i.e., g is a closed proper concave function. Since h is finite somewhere too, what we have proved for g can now be applied to h. Hence h is a closed proper concave function. By (3.9), h and g are conjugate to each other, i.e., (3.5) holds.

When we applied Theorem 4 above to the infimum in (3.12), which is just a re-expression of the one in (3.4), we skipped over the part about whether the infimum would be attained. Actually, Theorem 4 also yields the conclusion that, when $g(x_1) \neq -\infty$, the infimum in (3.4) is attained if and only if the set in (3.7) is nonempty. Dually, then, when $g^*(x_1^*) \neq -\infty$, the infimum in (3.5) is attained if and only if the set in (3.6) is nonempty. This condition does not involve the particular x_1^* one is looking at, so the infimum is attained for every x_1^* if it is finite and attained for one x_1^*. Now x_1 belongs to the domain J of the generalized derivative of g if and only if $g(x_1) + g^*(x_1^*) = x_1x_1^*$ for some x_1^*. This is equivalent to the supremum on the right of (3.9) being attained ($h = g^*$). If the set in (3.6) is nonempty, the infimum in (3.5) is always attained (as we have just seen), so attainment in (3.9) is equivalent to attainment in (3.10), which is the same as the right side of (3.13). Once again we have a situation where Theorem 4, or rather its dual, can be invoked in R^{N-1}. The resulting condition for attainment is that there exist some $(0, x_2, ..., x_N) \in K$ such that $x_i + c_i \in I_i$ for $i = 2, ..., N$, in other words (recalling the meaning of $c_2, ..., c_N$) that x_1 belong to the set in (3.6). This proves that (3.6) gives J in the nonempty case.

Finally, assume that the sets in (3.6) and (3.7) are both nonempty. Let $(x_1, ..., x_N) \in K$ and $(x_1^*, ..., x_N^*) \in K^*$. By the orthogonality of K and K^*,

$$\sum_{i=2}^{N} (f_i(x_i) + f_i^*(x_i^*) - x_ix_i^*) = \sum_{i=2}^{N} f_i(x_i) + \sum_{i=2}^{N} f_i^*(x_i^*) + x_1x_1^*. \tag{3.14}$$

Each term on the left side is non-negative by definition of the conjugate function, and is zero if and only if $(x_i, x_1^*) \in \Gamma_i$. Therefore

$$\sum_{i=2}^{N} f_i(x_i) + \sum_{i=2}^{N} f_i^*(x_i^*) \geqslant - x_1x_1^*, \tag{3.15}$$

with equality if and only if

$$(x_i, x_i^*) \in \Gamma_i \qquad \text{for} \qquad i = 2, ..., N. \tag{3.16}$$

When x_1 and x_1^* belong to the sets in (3.6) and (3.7), respectively, the right sides of (3.4) and (3.5) are not $-\infty$, and it follows from (3.15) that they cannot be $+\infty$ either. In particular, therefore, in the present situation g is finite somewhere and the part of Theorem 5 already established can be brought to bear. For instance, we have the fact that the extrema in (3.4) and (3.5) will always be attained. Thus by (3.15), (x_1, x_1^*) belongs to Δ if and only if $-g(x_1) - g^*(x_1^*) = -x_1x_1^*$, which means that $x_1^* \in g'(x_1)$. This proves Δ is the generalized derivative of g.

Next we shall employ Theorem 5 for N to prove Theorem 4 for N. Suppose first that g is finite at least somewhere, so that g is a closed proper concave function and g^* is given by (3.5). Then (3.2) is $\inf (f_1 - g)$, while (3.3) is $\inf (f_1^* - g^*)$. These are the negatives of each other by Lemma 3. Infimum (3.2) is attained if and only if $\inf (f_1 - g)$ is attained at some x_1 at which (3.4) is finite and attained, in other words if and only if $\inf (f_1 - g)$ is attained and the set in (3.7) is nonempty. This set is J^* when nonempty by the dual of Theorem 5, whereas by Lemma 3 $\inf (f_1 - g)$ is attained if and only if I_1^* and J^* have some x_1^* in common. That establishes the first attainment condition in Theorem 4.

If g were not finite at least somewhere, the infimum in (3.2) would have to be $-\infty$ (the $+\infty$ case being excluded by hypothesis). This is because

$$\sum_{i=1}^{N} f_i(x_i) + \sum_{i=1}^{N} f_i^*(x_i^*) = \sum_{i=1}^{N} (f_i(x_i) + f_i^*(x_i^*) - x_i x_i^*) \geq 0, \qquad (3.17)$$

when $(x_1, ..., x_N) \in K$ and $(x_1^*, ..., x_N^*) \in K^*$. We can also deduce the last assertion of Theorem 4 from (3.17). The infima add up to zero when they are finitely attained, so they are attained precisely at the points where equality holds in (3.17). That means that each of the non-negative terms on the right of (3.17) is actually zero, i.e., that $(x_i, x_i^*) \in \Gamma_i$ for $i = 1, ..., N$, by (3.1).

PROOF OF THEOREMS 1, 2 AND 3. The key to everything is the fact that, *if the constraints of problem* (B) *can be satisfied, the solutions to* (B) *will be the same as the points minimizing the extended function* $f_1 + \cdots + f_N$ *on* K. (And dually for (B*).) We prove this as follows. Let $(x_1, ..., x_N)$ be any point of K where $f_1(x_1) + \cdots + f_N(x_N)$ is finite, and let $(z_1, ..., z_N)$ be a point of K such that $z_i \in I_i$ for $i = 1, ..., N$. All the points of form

$$(\lambda z_1 + (1 - \lambda) x_1, ..., \lambda z_N + (1 - \lambda) x_N)$$

belong to K. Moreover

$$h(\lambda) = \sum_{i=1}^{N} f_i(\lambda z_i + (1 - \lambda) x_i)$$

has a right derivative at $\lambda = 0$, namely,

$$h'_+(0) = \sum_{i=1}^{N} \lim_{\lambda \downarrow 0} \frac{[f_i(x_i + \lambda(z_i - x_i)) - f_i(x_i)]}{\lambda}. \qquad (3.18)$$

(We saw in Section 2 that each of the limits in the sum exists and is either finite or $-\infty$.) If $x_i \notin I_i$ for some i, x_i is an end-point of the effective domain of f_i at which the corresponding limit in (3.18) is $-\infty$. Then $h'_+(0) = -\infty$ and the infimum of $f_1 + \cdots + f_N$ on K cannot be attained at $(x_1, ..., x_N)$.

This fact implies that, if a pair of vectors solves (B) and (B*), then it is a pair where the respective infima in (3.2) and (3.3) are finitely attained. The converse is also true, since by Theorem 4 the constraints in (B*) can be satisfied when the infimum in (3.2) is finitely attained (and dually the constraints in (B) can be satisfied when the infimum in (3.3) is finitely attained). Theorem 1 is now immediate from the last statement of Theorem 4. Moreover, it follows from the first assertion of Theorem 4 that the minima in (B) and (B*) are the negatives of each other when they exist. The key fact also implies via Theorem 4 that, if the constraints in (B) can be satisfied, (B) has a solution if and only if the constraints in (B*) can be satisfied. By duality, therefore, when the constraints in both (B) and (B*) can be satisfied both (B) and (B*) have solutions, whereas when the constraints in either (B) or (B*) cannot be satisfied neither (B) nor (B*) has a solution. This proves Theorem 2, and in view of Theorem 1 (which we have already verified) it proves Theorem 3,

4. Convex Programming Reformulation

We shall now deduce, from the results stated in the introduction, a rather general theory of convex programs with linear constraints.

The notation is the following. For $i = 1, \ldots, m$, let f_i be a closed proper convex function on R, with generalized derivative Γ_i having domain I_i and range I_i^*, and with conjugate function f_i^*. For $j = 1, \ldots, n$, let g_j be a closed proper concave function on R, with generalized derivative Δ_j having domain J_j and range J_j^*, and with conjugate function g_j^*. Let $((a_{ij}))$ be an $m \times n$ real matrix with transpose $((a_{ji}^*))$. Here are the corresponding problems.

(P) Minimize $\sum_{i=1}^m f_i(x_i) - \sum_{j=1}^n g_j(\sum_{i=1}^m x_i a_{ij})$ subject to $x_i \in I_i$ for $i = 1, \ldots, m$, and $\sum_{i=1}^m x_i a_{ij} \in J_j$ for $j = 1, \ldots, n$.

(P*) Maximize $\sum_{j=1}^n g_j^*(y_j^*) - \sum_{i=1}^m f_i^*(\sum_{j=1}^n y_j^* a_{ji}^*)$ subject to $y_j^* \in J_j^*$ for $j = 1, \ldots, n$, and $\sum_{j=1}^n y_j^* a_{ji}^* \in I_i^*$ for $i = 1, \ldots, m$.

(R) Find $(x_1, \ldots, x_m)$ and $(y_1^*, \ldots, y_n^*)$ satisfying $(x_i, \sum_{j=1}^n y_j^* a_{ji}^*) \in \Gamma_i$ for $i = 1, \ldots, m$, $(\sum_{i=1}^m x_i a_{ij}, y_j^*) \in \Delta_j$ for $j = 1, \ldots, n$.

The nature of (P) is brought out very clearly if we set $N = m + n$ and

$$f_{m+j} = -g_j \quad \text{and} \quad I_{m+j} = J_j \quad \text{for} \quad j = 1, \ldots, n. \tag{4.1}$$

The constraints in (P) are of the form

$$L_i(X) \in I_i \quad \text{for} \quad i = 1, \ldots, N, \tag{4.2}$$

where $X \in R^m$, each L_i is a linear function, and each I_i is a certain interval of R (not necessarily a closed interval, and possibly consisting of all of R or degenerating to a single point). The feasible solutions to (P) thus constitute a convex set C in R^m which is polyhedral, except that some of its faces might be missing. The problem is to minimize on C the convex function

$$F(X) = \sum_{i=1}^N f_i(L_i(X)), \tag{4.3}$$

where each f_i is a convex function given on I_i and satisfying the constructive regularity condition in the Corollary to Lemma 1 in Section 2. (Here $L_1, \ldots, L_m$ correspond to the canonical coordinate system on R^m. In general, one could consider a problem of minimizing the F in (4.3) subject to (4.2), where the L_i are linear functions on a certain real vector space. This problem could be reformulated as (P) by choosing $L_1, \ldots, L_m$ to be a maximal linearly independent set among the L_i and introducing coordinates $x_i = L_i(X)$.) This is a linear programming problem when every f_i is linear on I_i, which implies the I_i are closed. Quadratic programming subject to linear constraints is also included, for example, because any positive semi-definite quadratic function can be expressed (nonuniquely) as a sum of squares of linear functions. (Recall, incidentally, that finding such an expression is a very simple matter involving congruence of matrices and not eigenvalues.)

The dual problem (P*) is just like (P), except that it is a concave program instead of a convex program. We shall see below that problem (R) correponds in the linear programming case to the complementary slackness conditions.

THEOREM 6. (Characterization Theorem.) *A pair of vectors* $(x_1, \ldots, x_m)$ *and* $(y_1^*, \ldots, y_n^*)$ *solves* (R) *if and only if* $(x_1, \ldots, x_m)$ *solves* (P) *and* $(y_1^*, \ldots, y_n^*)$ *solves* (P*).

THEOREM 7. (Duality Theorem.) *Program* (P) *has a solution if and only if program* (P*) *has a solution, in which case the minimum in* (P) *equals the maximum in* (P*).

THEOREM 8. (Existence Theorem.) *If the constraints can be satisfied in both* (P) *and* (P*), *then* (P), (P*) *and* (R) *have solutions.*

PROOF OF THEOREMS 6, 7 AND 8. Let $N = m + n$, and

$$f_{m+j} = -g_j \qquad \text{for} \qquad j = 1, \ldots, n,$$

$$K = \left\{ (x_1, \ldots, x_N) \mid x_{m+j} = \sum_{i=1}^{m} x_i a_{ij} \text{ for } j = 1, \ldots, n \right\}.$$

Then f_i is a closed proper convex function on R for $i = 1, \ldots, N$, and K is a subspace of R^N. This puts problem (P) in the form of problem (B). The orthogonal complement of K is

$$K^* = \left\{ (x_1^*, \ldots, x_N^*) \mid x_i^* = -\sum_{j=1}^{n} x_{m+j}^* a_{ij}^* \text{ for } i = 1, \ldots, m \right\}. \qquad (4.5)$$

In addition, for $j = 1, \ldots, n$, the conjugate of f_{m+j} is

$$f^*_{m+j}(x^*_{m+j}) = -g^*_j(-x^*_{m+j}), \tag{4.6}$$

by Definition (2.11). The generalized derivative of f_{m+j} is

$$\Gamma_{m+j} = \{(x_{m+j}, x^*_{m+j}) \mid (x_{m+j}, -x^*_{m+j}) \in \Delta_j\}, \tag{4.7}$$

which has domain and range given by

$$I_{m+j} = J_j, \qquad I^*_{m+j} = \{x^*_{m+j} \mid -x^*_{m+j} \in J^*_j\}. \tag{4.8}$$

Thus problems (P*) and (R) are just (B*) and (A), with $y^*_{m+j} = -x^*_{m+j}$, under choices (4.1) and (4.2). Theorems 6, 7, and 8 are therefore corollaries of Theorems 1, 2, and 3.

REMARK. We have just seen how problems (R), (P), (P*) can be reformulated as (A), (B), (B*). As a matter of fact, this reformulation also works in the opposite direction. If K is an m-dimensional subspace of R^N, we can always arrange a permutation of $1, \ldots, N$, so as to get K represented as in (4.4) for a certain $((a_{ij}))$. (Of course, the representation is not uniquely determined; the set of $m \times n$ matrices one gets by considering the various suitable permutations forms a combinatorial equivalence class in the sense of Tucker.) Defining g_j, g^*_j, Δ_j, J_j and J^*_j from (4.4), (4.6), (4.7), and (4.8), one can express (A), (B) and (B*) as (R), (P) and (P*). Thus Theorems 6, 7, and 8 are really *equivalent* to Theorems 1, 2, and 3.

If one replaces (P) by the problem of minimizing the given function on all of R^m (i.e. if one replaces each (f_i, I_i) or (g_j, J_j) pair by the corresponding infinite-valued closed proper convex function f_i or concave function g_j defined on all of R), one has a program specializing the model handled in [10] and [2]. The theorem below says it is a *normal* program in the sense of [2, Section 6].

THEOREM 9. *For the extended functions, one has*

$$\inf_{x_1, \ldots, x_m} \left\{ \sum_{i=1}^{m} f_i(x_i) - \sum_{j=1}^{n} g_j \left(\sum_{i=1}^{m} x_i a_{ij} \right) \right\}$$

$$= \sup_{y^*_1, \ldots, y^*_n} \left\{ \sum_{j=1}^{n} g^*_j(y^*_j) - \sum_{i=1}^{m} f^*_i \left(\sum_{j=1}^{n} y^*_j a^*_{ji} \right) \right\},$$

except for the trivial case where the infimum is $+\infty$ and the supremum is $-\infty$. When the infimum is not $+\infty$, it is attained if and only if the constraints *in* (P*) *can be satisfied. When the supremum is not* $-\infty$, *it is attained if and only if the constraints in* (P) *can be satisfied.*

PROOF. This is a corollary of Theorem 4, via the reformulation already used in the preceding proof.

The usual dual linear programs correspond to the case where

$$f_i(x_i) = b_i^* x_i \quad \text{if} \quad x_i \geqslant 0, \qquad f_i(x_i) = +\infty \quad \text{if} \quad x_i < 0,$$
$$g_j(y_j) = 0 \quad \text{if} \quad y_j \geqslant b_j\,, \qquad g_j(y_j) = -\infty \quad \text{if} \quad y_j < b_j\,. \tag{4.9}$$

The generalized derivatives are then

$$\Gamma_i(x_i) = \begin{cases} b_i^* & \text{if} \quad x_i > 0, \\ \{x_i^* \mid x_i^* \leqslant b_i^*\} & \text{if} \quad x_i = 0, \\ \phi & \text{if} \quad x_i < 0, \end{cases}$$

$$\Delta_j(y_j) = \begin{cases} 0 & \text{if} \quad y_j > b_j\,, \\ \{y_j^* \mid y_j^* \geqslant 0\} & \text{if} \quad y_j = b_j\,, \\ \phi & \text{if} \quad y_j < b_j\,. \end{cases}$$

Thus the conjugate functions are

$$f_i^*(x_i^*) = 0 \quad \text{if} \quad x_i^* \leqslant b_i^*, \qquad f_i^*(x_i^*) = +\infty \quad \text{if} \quad x_i^* > b_i^*$$
$$g_j^*(y_j^*) = b_j y_j^* \quad \text{if} \quad y_j^* \geqslant 0, \qquad g_j^*(y_j^*) = -\infty \quad \text{if} \quad y_j^* < 0,$$

and one has

$$I_i = \{x_i \mid x_i \geqslant 0\} \qquad I_i^* = \{x_i^* \mid x_i^* \leqslant b_i^*\},$$
$$J_j = \{y_j \mid y_j \geqslant b_j\}, \qquad J_j^* = \{y_j^* \mid y_j^* \geqslant 0\}$$

The problems become:

(P) Minimize $\sum_{i=1}^m b_i^* x_i$ subject to $x_i \geqslant 0$ for $i = 1,\ldots, m$, and $\sum_{i=1}^m x_i a_{ij} \geqslant b_j$ for $j = 1,\ldots, n$.

(P*) Maximize $\sum_{j=1}^n b_j y_j^*$ subject to $y_j^* \geqslant 0$ for $j = 1,\ldots, n$, and $\sum_{j=1}^n y_j^* a_{ji}^* \leqslant b_i^*$ for $i = 1,\ldots, m$.

(R) Find $(x_1 ,\ldots, x_m)$ and $(y_1^*,\ldots, y_n^*)$ such that, for $x_i^* = \sum_{j=1}^n y_j^* a_{ji}^*$ and $y_i = \sum_{i=1}^m x_i a_{ji}$, one has $x_i \geqslant 0$, $x_i^* \leqslant b_i^*$ and $x_i(b_i^* - x_i^*) = 0$ for $i = 1,\ldots,m$, $y_j \geqslant b_j$, $y_j^* \geqslant 0$ and $(y_j - b_j)\, y_j^* = 0$ for $j = 1,\ldots, n$.

The familiar linear programming theorems of Gale, Kuhn and Tucker result when the theorems above are applied to this case. Observe that we have in fact provided an independent proof of these facts without using arguments from N-dimensional topology or convexity, and in particular without invoking the Minkowski-Farkas Lemma.

It is interesting to view each g_j in (4.9) as a *penalty function* in the sense of its contribution to the extended minimand in Theorem 9. There is no penalty if the constraint $\sum_{i=1}^{m} x_i a_{ij} \geqslant b_j$ is satisfied, but infinite penalty if it is not. In many situations, it ought to be more realistic to have the constraints correspond instead to penalty functions g_j which grow rapidly but continuously from zero to infinity rather than making an abrupt jump. The theory set forth here handles such functions as easily as it handles the all-or-nothing ones of linear programming.

References

1. G. J. Minty. Monotone networks. *Proc. Roy. Soc. London Ser. A* 257 (1960), 194-212.
2. R. T. Rockafellar. Duality and stability in extremum problems involving convex functions. *Pacific J. Math.* **21** (1967), 167-187.
3. G. J. Minty. On the axiomatic foundations of the theories of directed linear graphs, electrical networks and network programming. *J. Math. Mech.* **15** (1966), 485-520.
4. C. Berge and A. Gbouila-Houri. "Programmes, Jeux et Réseaux de Transport." Dunod, Paris, 1962.
5. R. T. Rockafellar. Convex functions and dual extremum problems. Dissertation. Harvard, 1963.
6. G. J. Minty. A theorem on maximal monotonic sets in Hilbert space. *J. Math. Anal. Appl.* **11** (1965), 434-439.
7. P. Camion. Application d'une généralisation du lemme de Minty a une problème d'infimum de fonction convexe. *Cahiers Centre Etudes Recherche Oper.* **7** (1965), 230-247.
8. W. Fenchel. Convex cones, sets and functions (mimeographed lecture notes). Princeton University, 1953.
9. R. T. Rockafellar. Characterization of the subdifferentials of convex functions. *Pacific J. Math.* **17** (1966), 497-510.
10. R. T. Rockafellar. Duality theorems for convex functions. *Bull. Amer. Math. Soc.* **70** (1964), 189-192.
11. W. Fenchel. On conjugate convex functions. *Canad. J. Math.* **1** (1949), 73-77.
12. R. T. Rockafellar. The elementary vectors of a subspace of R^N. To appear.

PRINTED IN BRUGES, BELGIUM, BY THE ST CATHERINE PRESS, LTD.