JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 19, 103-124 (1967)

Constrained Minimization Under Vector-Valued Criteria in Finite Dimensional Spaces*

N. O. DA CUNHA** AND E. POLAK

Department of Electrical Engineering
University of California, Berkeley 94720

Submitted by Lotfi Zadeh

INTRODUCTION

In his 1896 treatise "Cours d'Economie Politique" [1], Pareto discussed in the qualitative terms of values and prices the formation of a single, real-valued optimality criterion from a number of essentially noncomparable "elementary" real-valued criteria. It is obvious that these "elementary" criteria can be viewed as the components of a vector-valued optimality criterion. Since then, discussions of vector-valued optimization problems have kept reappearing in the economics literature (see Karlin [2], Debreu [3]), in the literature of nonlinear programming (see Kuhn and Tucker [4]) and, more recently, in the literature of control theory (see Zadeh [5], Chang [6]).

Since a vector-valued criterion usually induces a partial ordering on the set of alternatives, one cannot speak of " optimal " solutions under a vector-valued criterion as one can in the case of a linear ordering induced by a scalar-valued criterion. Thus, in the case of a partial ordering, the notion of an optimal solution is replaced by that of the set of noninferior (efficient, minimal, nondominated) solutions, that is, the set of solutions which are not inferior to any other solution under the partial ordering. The present paper is devoted to developing a broad theory of necessary conditions for characterizing noninferior points and to determining when a vector-valued criterion problem can be treated as a family of problems with scalar-valued criteria. The necessary conditions presented in this paper extend the results of [13] which dealt with ordinary constrained minimization problems.

I. NECESSARY CONDITIONS FOR THE CANONICAL PROBLEM

Let $f : E^n \to E^p$, $r : E^n \to E^m$ be continuously differentiable functions and let Ω be a subset of E^n.

* The research reported herein was supported wholly by the National Aeronautics and Space Administration under Grant NsG-354, Supplement 3.

** On leave from the Instituto Tecnologico de Aeronautica, Sao Jose dos Campos, Brazil.

1. Canonical Problem. Find a point $\hat{x}$ in E^n such that

2. $\hat{x} \in \Omega$ and $r(\hat{x}) = 0$ and

3. for every x in Ω with $r(x) = 0$, the relation $f(x) \leqslant f(\hat{x})$[3] (componentwise) implies that $f(x) = f(\hat{x})$.

It can easily be shown [7] that the solutions of the canonical problem (1) usually constitute an uncountable set of points.

Before we can obtain necessary conditions for a point $\hat{x}$ in E^n to be a solution to the canonical problem, we must introduce an approximation to the set Ω at $\hat{x}$.

4. Definition. *A subset $C(\hat{x}, \Omega)$ of E^n will be called a conical approximation of the set Ω at $\hat{x}$ if*

5. *$C(\hat{x}, \Omega)$ is a convex cone, and*

6. *for any finite collection $\{x_1, x_2, \ldots, x_k\}$ of linearly independent vectors in $C(\hat{x}, \Omega)$, there exist a positive scalar ϵ_0 and a continuous map ζ from co $\{\epsilon x_1, \ldots, \epsilon x_k\}$, the convex hull of $\{\epsilon x_1, \ldots, \epsilon x_k\}$, with $0 < \epsilon \leqslant \epsilon_0$, into $\Omega - \{\hat{x}\}$ of the form:*

$$\zeta(\delta x) = \delta x + o(\delta x) \textit{ for all } \delta x \in \textit{co } \{\epsilon x_1, \ldots, \epsilon x_k\},\ 0 < \epsilon \leqslant \epsilon_0$$

where the function $o(\cdot)$ is such that

$$\lim_{\|y\| \to 0} \frac{\| o(y) \|}{\| y \|} = 0.$$

An important special case of a conical approximation is one where the map ζ is the identity map, i.e., co $\{\epsilon x_1, \ldots, \epsilon x_k\}$ is contained in $\Omega - \{\hat{x}\}$ for $0 < \epsilon \leqslant \epsilon_0$. We call this special case a *conical approximation the first kind.*

7. Theorem. *Let $\hat{x}$ be a solution to the canonical problem and let $C(\hat{x}, \Omega)$ be a conical approximation for Ω at $\hat{x}$. Then, there exist a vector μ in E^p and a vector η in E^m such that*

8. $\mu^i \leqslant 0,\ i = 1, 2, \ldots, p,$

9. $(\mu, \eta) \neq 0$, *and*

10. $$\left\langle \mu, \frac{\partial f(\hat{x})}{\partial x} x \right\rangle + \left\langle \eta, \frac{\partial r(\hat{x})}{\partial x} x \right\rangle \leqslant 0 \qquad \textit{for all} \quad x \in \overline{C(\hat{x}, \Omega)},$$

where $\overline{C(\hat{x}, \Omega)}$ is the closure of $C(\hat{x}, \Omega)$.

---

[3] We use the following notation. For any vectors y_1, y_2 in E^p, $y_1 \leqslant y_2$ if and only if $y_1^i \leqslant y_2^i$ for $i = 1, 2, \ldots, p$; $y_1 \leq y_2$ if and only if $y_1 \neq y_2$ and $y_1 \leqslant y_2$; $y_1 < y_2$ if and only if $y_1^i < y_2^i$ for $i = 1, 2, \ldots, p$.

PROOF. Let

11. $$A(\hat{x}) = \left\{ y \in E^p \mid y = \frac{\partial f(\hat{x})}{\partial x} x, \quad x \in C(\hat{x}, \Omega) \right\},$$

12. $$B(\hat{x}) = \left\{ z \in E^m \mid z = \frac{\partial r(\hat{x})}{\partial x} x, \quad x \in C(\hat{x}, \Omega) \right\},$$

13. $$K(\hat{x}) = \left\{ u \in E^p \times E^m \mid u = \left(\frac{\partial f(\hat{x})}{\partial x} x, \frac{\partial r(\hat{x})}{\partial x} x \right), \quad x \in C(\hat{x}, \Omega) \right\}.$$

Since the Jacobian matrices $\partial f(\hat{x})/\partial x$ and $\partial r(\hat{x})/\partial x$ define linear maps, $A(\hat{x})$, $B(\hat{x})$, and $K(\hat{x})$ are convex cones in E^p, E^m, and $E^p \times E^m$, respectively. Clearly, $K(\hat{x}) \subset A(\hat{x}) \times B(\hat{x})$.

Let C and R be the convex cones in E^p and $E^p \times E^m$, respectively, defined by

14. $C = \{y = (y^1, \ldots, y^p) \in E^p \mid y^i < 0, i = 1, 2, \ldots, p\}$, and

15. $R = \{(y, 0) \in E^p \times E^m \mid y \in C, 0 \in E^m\}$.

Examining (9) and (10), we observe that the claim of the theorem is that the sets $K(\hat{x})$ and R are separated in $E^p \times E^m$. We now construct a proof by contradiction.

Suppose that $K(\hat{x})$ and R are not separated in $E^p \times E^m$. We then find that the following two statements must be true.

16. The smallest linear variety containing the union of R and $K(\hat{x})$ is the entire space $E^p \times E^m$, and $R \cap K(\hat{x}) \neq \phi$, the empty set.

17. The convex cone $B(\hat{x})$ in E^m, contains the origin as an interior point and, since $B(\hat{x})$ is a convex cone, $B(\hat{x}) = E^m$.

This follows from the fact that if 0 is not an interior point of the convex set $B(\hat{x})$, then by the *separation theorem*,[4] it can be separated from $B(\hat{x})$ by a hyperplane in E^m, i.e., there exists a nonzero vector η_0 in E^m such that

$$\langle \eta_0, z \rangle \leqslant 0 \text{ for all } z \in B(\hat{x}).$$

Clearly, the vector $(0, \eta_0)$ in $E^p \times E^m$ separates R from $A(\hat{x}) \times B(\hat{x})$ and hence from $K(\hat{x})$ contradicting our assumption that they are not separated.

We now proceed to utilize facts (16) and (17). Since the origin in E^m belongs to the nonvoid interior of $B(\hat{x}) = E^m$ (see (17)), let us construct a simplex Σ in $B(\hat{x})$, with vertices $z_1, z_2, \ldots, z_{m+1}$ such that

18. 0 is in the interior of Σ;

[4] See [8] p. 118, 2.22. Corollary to the Hahn-Banach Theorem.

19. there exists a set of linearly independent vectors $\{x_1, x_2, \ldots, x_{m+1}\}$ in $C(\hat{x}, \Omega)$ satisfying

20. $$z_i = \frac{\partial r(\hat{x})}{\partial x} x_i \quad \text{for} \quad i = 1, 2, \ldots, m+1,$$

21. $\zeta(x) = x + o(x) \in \{\Omega - \{\hat{x}\}\}$ for all $x \in \mathrm{co}\{x_1, x_2, \ldots, x_{m+1}\}$,

where ζ is the map entering the definition of $C(x, \Omega)$, see (4), and

22. the points $y_i = \dfrac{\partial f(\hat{x})}{\partial x} x_i$ are in C for $i = 1, 2, \ldots, m+1$.[5]

The existence of such a simplex is easily established. First, we construct any simplex Σ' in $B(\hat{x})$ with vertices $z_1', z_2', \ldots, z_{m+1}'$, which contains the origin in its interior. This is clearly possible since $B(\hat{x}) = E^m$ by (17). Let $x_1', x_2', \ldots, x_{m+1}'$ be any set of points in $C(\hat{x}, \Omega)$ which satisfy (20), i.e., $z_i' = [\partial r(\hat{x})/\partial x]\, x_i'$, $i = 1, 2, \ldots, m+1$. If $[\partial f(\hat{x})/\partial x] x_i' < 0$ for $i = 1, 2, \ldots, m+1$, then (22) is satisfied and we can satisfy (21) by letting $x_i = \epsilon x_i'$, for some $\epsilon > 0$, and still satisfy (18), (20), and (22). But suppose, without loss of generality, that $[\partial f(\hat{x})/\partial x]\, x_1' \geqslant 0$ and $[\partial f(\hat{x})/\partial x]\, x_i' < 0$ for $i = 2, 3, \ldots, m+1$. Since by (16) $K(\hat{x}) \cap R \neq \varnothing$, there exists a point

$$u = \left(\frac{\partial f(\hat{x})}{\partial x} \tilde{x}, 0\right) \in K(\hat{x}) \cap R,$$

i.e., $[\partial f(\hat{x})/\partial x]\tilde{x} < 0$ and $[\partial r(\hat{x})/\partial x]\tilde{x} = 0$. Choose any scalar $\lambda > 0$ such that $[\partial f(\hat{x})/\partial x](\lambda x_1' + (1-\lambda)\tilde{x}) < 0$, and let $x_1 = \lambda x_1' + (1-\lambda)\tilde{x}$. Then the simplex Σ with vertices $\epsilon\lambda z_1', \epsilon z_2', \ldots, \epsilon z_{m+1}'$, satisfies conditions (18), (19), (20), (21), and (22) for the corresponding vectors $\lambda x', x_2', x_3', \ldots, x_{m+1}'$ and some $\epsilon > 0$.

It is easy to show that (18) implies that the vectors $(z_1 - z_{m+1}), (z_2 - z_{m+1}), \ldots, (z_m - z_{m+1})$ are linearly independent. Consequently, since $[\partial r(\hat{x})/\partial x]$ is a linear map, the vectors $(x_1 - x_{m+1}), (x_2 - x_{m+1}), \ldots, (x_m - x_{m+1})$ are also linearly independent. Let Z be the nonsingular $m \times m$ matrix whose columns are $(z_1 - z_{m+1}), (z_2 - z_{m+1}), \ldots, (z_m - z_{m+1})$ and let X be the $n \times m$ matrix whose columns are $(x_1 - x_{m+1}), (x_2 - x_{m+1}), \ldots, (x_m - x_{m+1})$. Then $z \to XZ^{-1}(z - z_{m+1}) + x_{m+1}$ is a continuous map from Σ into $\mathrm{co}\{x_1, x_2, \ldots, x_{m+1}\}$.

Now, for $0 < \alpha \leqslant 1$, let S_α be a sphere in E^m with radius $\alpha\rho$ (where $\rho > 0$), center at the origin, and contained in the interior of the simplex Σ.

[5] It is easy to show that (18), (20), and (22) imply that the vectors $x_1, x_2, \ldots, x_{m+1}$, are linearly independent.

Next we define a continuous map G_α from the sphere S_α into E^m by

23. $$G_\alpha(\alpha z) = r(\hat{x} + \zeta(\alpha XZ^{-1}(z - z_{m+1}) + \alpha x_{m+1}))$$
$$= r(\hat{x} + \alpha XZ^{-1}(z - z_{m+1}) + \alpha x_{m+1} + o(\alpha XZ^{-1}(z - z_{m+1}) + \alpha x_{m+1})),$$

where $\| z \| \leqslant \rho$, $\alpha z \in S_\alpha$, and ζ is the map associated with the conical approximation $C(\hat{x}, \Omega)$. Since r is continuously differentiable, we can expand the right-hand side of (23) about $\hat{x}$ to obtain

24. $$G_\alpha(\alpha z) = r(\hat{x}) + \alpha \frac{\partial r(\hat{x})}{\partial x} (XZ^{-1}(z - z_{m+1}) + x_{m+1}) + o(\alpha XZ^{-1}(z - z_{m+1}) + \alpha x_{m+1}).$$

But $r(\hat{x}) = 0$, $[\partial r(\hat{x})/\partial x]\, X = Z$, and $[\partial r(\hat{x})/\partial x]\, x_{m+1} = z_{m+1}$. Hence, (24) becomes

25. $$G_\alpha(\alpha z) = \alpha z + o(\alpha XZ^{-1}(z - z_{m+1}) + \alpha x_{m+1}).$$

Now, since

$$\lim_{\alpha \to 0} \frac{\| o(\alpha XZ^{-1}(z - z_{m+1}) + \alpha x_{m+1})\|}{\alpha} = 0,$$

there exists for $\| z \| = \rho$, an $\bar{\alpha}_0$, $0 < \bar{\alpha}_0 \leqslant 1$, such that

26. $\| o(\alpha XZ^{-1}(z - z_{m+1}) + \alpha x_{m+1}) \| < \alpha\rho$, for all $0 < \alpha \leqslant \bar{\alpha}_0$ and $\| z \| = \rho$.

By assumption, f is differentiable, hence we can expand each component of f about $\hat{x}$ as follows:

27. $$f^i(\hat{x} + \zeta(\alpha XZ^{-1}(z - z_{m+1}) + \alpha x_{m+1}))$$
$$= f^i(\hat{x}) + \alpha \frac{\partial f^i(\hat{x})}{\partial x} [XZ^{-1}(z - z_{m+1}) + x_{m+1}] + o(\alpha XZ^{-1}(z - z_{m+1}) + \alpha x_{m+1}). \qquad i = 1, 2, \ldots, p.$$

Since by construction, (see (22)), $[\partial f^i(\hat{x})/\partial x]x_j < 0$, for $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, m+1$, and the point $XZ^{-1}(z - z_{m+1}) + x_{m+1}$ is in $\mathrm{co}\{x_1, x_2, \ldots, x_{m+1}\}$, we have $[\partial f^i(\hat{x})/\partial x][XZ^{-1}(z - z_{m+1}) + x_{m+1}] < 0$, with $i = 1, 2, \ldots, p$. Hence there exist $\bar{\alpha}_i$, $i = 1, 2, \ldots, p$, such that

28. $f^i(\hat{x} + \alpha(XZ^{-1}(z - z_{m+1}) + x_{m+1})) < f^i(\hat{x})$ for all $0 < \alpha \leqslant \bar{\alpha}_i$, $\| z \| = \rho$ and $i = 1, 2, \ldots p$.

Let α^* be the minimum of $\{\bar{\alpha}_0, \bar{\alpha}_1, \ldots, \bar{\alpha}_p\}$. It now follows from Brouwer's fixed point theorem [9] that there exists a point $\alpha^* z^*$ such that $G_{\alpha^*}(\alpha^* z^*) = 0$.

Now, let $x^* = \hat{x} + \zeta(\alpha^* X Z^{-1}(z^* - z_{m+1}) + \alpha^* x_{m+1})$, then

29. $r(x^*) = 0$ (since $r(x^*) = G_{\alpha^*}(\alpha^* z^*) = 0$), and

30. $x^* \in \Omega$, since $(x^* - \hat{x}) \in \zeta(\text{co}\{\alpha^* x_1, \alpha^* x_2, ..., \alpha^* x_{m+1}\}) \subset \Omega - \{\hat{x}\}$ by construction.

But (28), (29), and (30) contradict the assumption that x is a solution to the canonical problem (1). Therefore, the convex cones $K(\hat{x})$ and R are separated in $E^p \times E^m$, i.e., there exists a nonzero vector (μ, η) in $E^p \times E^m$ such that

31. $$\left\langle \mu, \frac{\partial f(\hat{x})}{\partial x} x \right\rangle + \left\langle \eta, \frac{\partial r(\hat{x})}{\partial x} x \right\rangle \leqslant 0 \qquad \text{for all} \quad x \in C(\hat{x}, \Omega),$$

and

32. $$\langle \mu, y \rangle + \langle \eta, 0 \rangle \geqslant 0 \qquad \text{for all} \quad y \in C.$$

But (31) implies that

$$\left\langle \mu, \frac{\partial f(\hat{x})}{\partial x} x \right\rangle + \left\langle \eta, \frac{\partial r(\hat{x})}{\partial x} x \right\rangle \leqslant 0 \qquad \text{for all} \quad x \in \overline{C(\hat{x}, \Omega)}$$

and (32) and (14) implies that $\mu^i \leqslant 0$, $i = 1, 2, ..., p$. Q.E.D.

II. Reduction of a Vector-Valued Criterion to a Family of Scalar-Valued Criteria

An examination of (9) and (10) indicates that if we had used the scalar-valued criterion $\langle -\mu, f(x) \rangle$ instead of the vector-valued criterion $f(x)$ in the definition of the canonical problem (1), with $\mu \in E^p$ specified by theorem (7) for the vector-valued criterion, we would have obtained from theorem (7) exactly the same set of necessary conditions. This observation leads us to the following important question: can wc obtain the solutions to the canonical problem (1) by solving a family of scalar-valued criterion problems? A partial answer to this question is given below by theorems (38) and (41).

To simplify our exposition, we lump the constraint set Ω with the set $\{x \in E^n \mid r(x) = 0\}$. We shall therefore consider a subset A of E^n, a continuous mapping f from E^n into E^p and introduce the following definitions.

33. Definition. *We shall denote by P the problem of finding a point $\hat{x}$ in A such that for every x in A, the relation $f(x) \leqslant f(\hat{x})$ (component-wise) implies that $f(x) = f(\hat{x})$.*

34. Definition. *Let Λ be the set of all vectors $\lambda = (\lambda^1, \lambda^2, ..., \lambda^p)$ in E^p such that $\sum_{i=1}^{p} \lambda^i = 1$ and $\lambda^i > 0$, $i = 1, 2, ..., p$.*

35. DEFINITION. *Given any vector λ in E^p, we shall denote by $P(\lambda)$ the problem of finding a point $\bar{x}$ in A such that $\langle\lambda, f(\bar{x})\rangle \leqslant \langle\lambda, f(x)\rangle$ for all x in A.*

We shall consider the following subsets of E^n:

36. $$L = \{x \in A \mid x \text{ solves } P\},$$

37. $$M = \{x \in A \mid x \text{ solves } P(\lambda) \text{ for some } \lambda \in \Lambda\}.$$

38. THEOREM. *The set L contains the set M.*

PROOF. Suppose that $\bar{x} \in M$ and $\bar{x} \notin L$. Then there must exist a point x' in A such that $f(x') \leqslant f(\bar{x})$. But for any $\lambda \in \Lambda$, this implies that $\langle\lambda, f(x')\rangle < \langle\lambda, f(\bar{x})\rangle$, and hence $\bar{x}$ is not in M, which is a contradiction.

39. DEFINITION. *We shall say that a solution $\hat{x}$ of the problem P is regular if there exists a closed convex neighborhood U of $\hat{x}$ such that for any $y \in A \cap U$ the relation $f(\hat{x}) = f(y)$ implies $\hat{x} = y$.*

40. DEFINITION. *We shall say that the problem P is regular if every solution of P is a regular solution.*

It is easy to verify that if f is convex and one of its components is strictly convex then P is regular.

41. THEOREM. *Suppose that the problem P is regular, that the performance criterion f is convex (component-wise) and that the constraint set A is closed and convex. Then the set L (36) is contained in the closure of the set M (37).*

PROOF. We shall show that for every $\hat{x} \in L$, there exists a sequence of points in M which converges to $\hat{x}$.

We begin by constructing a sequence which converges to an arbitrary, but fixed, $\hat{x}$ in L. We shall then show that this sequence is in M.

Let $\hat{x}$ be any point in L. Since we can translate the origins of E^n and E^p, we may suppose, without loss of generality, that $\hat{x} = 0$ and that $f(\hat{x}) = 0$.

Let U be a closed convex neighborhood of $\hat{x}$ satisfying the conditions of definition (39), and let $N \subset U$ be a compact convex neighborhood of $\hat{x}$. For any positive scalar ϵ, $0 < \epsilon \leqslant (1/p)$, (where p is the dimension of the space containing the range of $f(\cdot)$), let

42. $$\Lambda(\epsilon) = \left\{\lambda = (\lambda^1, \lambda^2, \dots, \lambda^p) \;\middle|\; \sum_{i=1}^{p} \lambda^i = 1, \lambda^i \geqslant \epsilon, \qquad i = 1, 2, \dots, p\right\}.$$

Let g be the real-valued function with domain $A \cap N \times \Lambda(\epsilon)$, defined by

43. $$g(\lambda, x) = \langle\lambda, f(x)\rangle.$$

Clearly, since f is convex and hence continuous, g is continuous in $A \cap N \times \Lambda(\epsilon)$, furthermore, g is convex in x for fixed λ and linear in λ for fixed x. Since the sets $A \cap N$ and $\Lambda(\epsilon)$ are compact and convex, the sets

44. $$\{x \in A \cap N \mid g(\bar{\lambda}, x) = \min_{\eta \in A \cap N} g(\bar{\lambda}, \eta)\},$$

45. $$\{\lambda \in \Lambda(\epsilon) \mid g(\lambda, \bar{x}) = \max_{\nu \in \Lambda(\epsilon)} g(\nu, \bar{x})\},$$

are well defined for every $\bar{\lambda} \in \Lambda(\epsilon)$ and every $\bar{x} \in A \cap N$, respectively. Obviously, the sets defined in (44) and (45) are convex.

By Ky Fan's theorem [10],[6] there exist a point $\lambda(\epsilon)$ in $\Lambda(\epsilon)$ and a point $x(\epsilon)$ in $A \cap N$ such that

46. $$\langle \lambda(\epsilon), f(x) \rangle \geqslant \langle \lambda(\epsilon), f(x(\epsilon)) \rangle \geqslant \langle \lambda, f(x(\epsilon)) \rangle$$

for every x in $A \cap N$ and λ in $\Lambda(\epsilon)$.

Since $\hat{x} = 0$ is in $A \cap N$ and $f(\hat{x}) = 0$, we have from (46):

47. $$\langle \lambda(\epsilon), f(x(\epsilon)) \rangle \leqslant 0.$$

And from (46) and (47),

48. $$\langle \lambda, f(x(\epsilon)) \rangle \leqslant 0 \text{ for every } \lambda \text{ in } \Lambda(\epsilon).$$

Since $A \cap N$ is compact, we can choose a sequence ϵ_n, $n = 1, 2,\ldots$, with $0 < \epsilon_n \leqslant 1/p$, converging to zero in such a way that the resulting sequence of points $x(\epsilon_n)$, satisfying (46), converges, i.e.,

49. $$\lim_{n\to\infty} x(\epsilon_n) = x^*, \; x^* \in A \cap N.$$

Since $g(\lambda, x)$ is continuous, it follows from (48) and (49) that

$$\langle \lambda, f(x^*) \rangle \leqslant 0 \qquad \text{for all} \quad \lambda \in \Lambda,$$

[6] Ky Fan's Theorem. *Let L_1, L_2 be two separated locally convex, topological linear spaces, and K_1, K_2 be two, compact convex sets in L_1, L_2, respectively. Let g be a real-valued continuous function on $K_1 \times K_2$. If, for any $x_0 \in K_1$, $y_0 \in K_2$ the sets*

$$\{x \in K_1 \mid g(x, y_0) = \max_{\nu \in K_1} g(\nu, y_0)\}$$

and

$$\{y \in K_2 \,.\, g(x_0, y) = \min_{\eta \in K_2} g(x_0, \eta)\}$$

are convex, then

$$\max_{x \in K_1} \min_{y \in K_2} g(x, y) = \min_{y \in K_2} \max_{x \in K_1} g(x, y).$$

which implies that $f(x^*) \leqslant 0$. But $\hat{x}$ is a solution to P, hence $f(x^*) \leqslant 0 = f(\hat{x})$ implies that $f(x^*) = f(\hat{x})$. Consequently, since P is regular, $x^* = \hat{x} = 0$. Thus, we have constructed a sequence, $\{x(\epsilon_n)\}$ which converges to $\hat{x}$.

We shall now show that the sequence $\{x(\epsilon_n)\}$ contains a subsequence $\{x(\epsilon_n)\}$ also converging to $\hat{x}$, which is contained in M.

Since $\hat{x}$ is in the interior of N, there exists a positive integer n_0 such that the points $x(\epsilon_n) \in A \cap N$ belong to the interior of N for $n \geqslant n_0$.

We will show that for $n \geqslant n_0$, $x(\epsilon_n)$ is a solution to $P(\lambda(\epsilon_n))$, i.e., that for $n \geqslant n_0$, $x(\epsilon_n) \in M$. By contradiction, suppose that for $n \geqslant n_0$, $x(\epsilon_n)$ is not a solution to $P(\lambda(\epsilon_n))$. Then there must be a point x' in A such that

51. $$\langle \lambda(\epsilon_n), f(x') \rangle < \langle \lambda(\epsilon_n), f(x(\epsilon_n)) \rangle.$$

Let $x''(\alpha) = (1 - \alpha)x(\epsilon_n) + \alpha x_1'$, $0 < \alpha < 1$; since A is convex, $x''(\alpha)$ is an A for $0 < \alpha < 1$. But for $n \geqslant n_0$, $x(\epsilon_n)$ is in the interior of N and hence there exists an α^*, $0 < \alpha^* < 1$ such that $x''(\alpha^*)$ belongs to N.

Now,

52. $$\langle \lambda(\epsilon_n), f(x''(\alpha^*)) \rangle = \langle \lambda(\epsilon_n), f((1 - \alpha^*)\, x(\epsilon_n) + \alpha x') \rangle.$$

But for $\lambda(\epsilon_n) \in \Lambda(\epsilon_n)$, $\langle \lambda(\epsilon_n), f(x) \rangle$ is convex in x. Hence (51) and (52) imply that

53. $$\langle \lambda(\epsilon_n), f(x''(\alpha^*)) \rangle < \langle \lambda(\epsilon_n), f(x(\epsilon_n)) \rangle,$$

which contradicts (46).

Therefore, for $n \geqslant n_0$, $x(\epsilon_n)$ is a solution to $P(\lambda(\epsilon_n))$, i.e., $x(\epsilon_n)$ is in M. Thus, for any given $\hat{x} \in L$ there exists a sequence $\{x(\epsilon_n)\}$ contained in M such that $x(\epsilon_n) \to \hat{x}$ as $n \to \infty$. This completes our proof.

III. Applications to Nonlinear Programming

In nonlinear programming the set Ω is usually defined by a set of inequalities. Thus, let q^i, $i = 1, 2, \ldots, s$ be continuously differentiable functions from E^n into E^1. Then Ω is defined by

54. $$\Omega = \{x \in E^n \mid q^i(x) \leqslant 0, \qquad i = 1, 2, \ldots, s\}.$$

55. The Nonlinear Programming Problem. We shall refer to the particular case of the canonical problem (1), arising when the constraint set Ω is defined by (54), as the nonlinear programming problem.

At each point x in Ω, the index set of active constraints is defined as

56. $$I(x) = \{i \mid q^i(x) = 0, \qquad i \in \{1, 2,..., s\}\}.$$

Similarly, the index set of inactive constraints is defined as

57. $$\bar{I}(x) = \{i \mid q^i(x) < 0, \qquad i \in \{1, 2,..., s\}\}.$$

Let $\hat{x}$ be a solution to the nonlinear programming problem. In order to bring the additional structure of the nonlinear programming problem into play, it is convenient to begin by allowing the following assumption, which will subsequently be removed.

58. Assumption. *There exists a vector z in E^n such that $[\partial q^i(\hat{x})/\partial x]\, z < 0$ for every $i \in I(\hat{x})$.*

Under this assumption, the nonvoid set

$$C(\hat{x}, \Omega) = \left\{x \in E^n \,\middle|\, \frac{\partial q^i(\hat{x})}{\partial x} x < 0, \qquad i \in I(\hat{x})\right\}$$

is a conical approximation of the first kind for Ω at $\hat{x}$, and

$$\overline{C(\hat{x}, \Omega)} = \left\{x \in E^n \,\middle|\, \frac{\partial q^i(\hat{x})}{\partial x} x \leqslant 0, \qquad i \in I(\hat{x})\right\}.$$

By theorem (7) there exist vectors μ in E^p and η in E^m such that

(i) $\mu^i \leqslant 0, \qquad i = 1, 2,..., p,$

(ii) $(\mu, \eta) \neq 0,$

(iii) $$\sum_{i=1}^{p} \mu^i \frac{\partial f^i(\hat{x})}{\partial x} x + \sum_{i=1}^{m} \eta^i \frac{\partial r^i(\hat{x})}{\partial x} x \leqslant 0$$

for every

$$x \in \left\{x \in E^n \,\middle|\, \frac{\partial q^i(\hat{x})}{\partial x} x \leqslant 0, \qquad i \in I(\hat{x})\right\}.$$

And by Farkas' lemma [11], there exist scalars $\rho^i \leqslant 0$, $i \in I(\hat{x})$ such that

59. $$\sum_{i=1}^{p} \mu^i \frac{\partial f^i(\hat{x})}{\partial x} + \sum_{i=1}^{m} \eta^i \frac{\partial r^i(\hat{x})}{\partial x} + \sum_{i \in I(\hat{x})} \rho^i \frac{\partial q^i(\hat{x})}{\partial x} = 0.$$

Defining $\rho^i = 0$ for $i \in \bar{I}(\hat{x})$, we have just proved

60. THEOREM. *Let $\hat{x}$ be a solution to the nonlinear programming problem* (55). *If assumption* (58) *holds, then there exist scalars μ^i, $i = 1, 2,..., p$, η^j, $j = 1, 2,..., m$ and ρ^k, $k = 1, 2,..., s$ such that*

61. $$\mu^i \leqslant 0, \qquad i = 1, 2,..., p,$$

62. $$\rho^k \leqslant 0, \qquad k = 1, 2,..., s,$$

63. $$(\mu, \eta) \neq 0,$$

64. $$\sum_{i=1}^{p} \mu^i \frac{\partial f^i(\hat{x})}{\partial x} + \sum_{j=1}^{m} \eta^j \frac{\partial r^j(\hat{x})}{\partial x} + \sum_{k=1}^{s} \rho^k \frac{\partial q^k(\hat{x})}{\partial x} = 0,$$

and

65. $$\sum_{k=1}^{s} \rho^k q^k(\hat{x}) = 0.$$

When the additional assumption (58) does not hold, we can use the following lemma to obtain somewhat weaker necessary conditions for the nonlinear programming problem, still involving its entire structure.

65. LEMMA. *Let v_i, $i = 1, 2,..., k$ be any k vectors in E^n. If the system*

66. $$\langle v_i , x\rangle < 0, \qquad i = 1, 2,..., k$$

has no solution x in E^n, then there exists a nonzero vector $\bar{\rho}$ in E^k, with $\bar{\rho}^i \leqslant 0$, $i = 1, 2,..., k$ such that $\sum_{i=1}^{k} \bar{\rho}^i v_i = 0$.

PROOF. Let

$$B = \left\{x \in E^n \,\middle|\, x = \sum_{i=1}^{k} \rho^i v_i , \rho^i \leqslant 0, \quad \text{not all zero}\right\}.$$

We want to prove that the origin belongs to B. By contradiction, suppose that the origin does not belong to B. Then 0 does not belong to the convex hull of $\{-v_1 , -v_2 ,..., -v_k\}$ since $\text{co}\,\{-v_1 , -v_2 ,..., -v_k\}$ is a subset of B. But $\text{co}\,\{-v_1 , -v_2 ,..., -v_k\}$ is a closed convex set in E^n not containing the origin. Hence, by the strong separation theorem,[7] there exists a hyperplane in E^n strictly separating the set $\text{co}\,\{-v_1 , -v_2 ,..., -v_k\}$ from the origin, i.e., there exists a nonzero vector $\bar{x}$ in E^n such that

67. $$\langle \bar{x}, x\rangle > 0 \quad \text{for every} \quad x \in \text{co}\{-v_1 , -v_2 ,..., -v_k\}.$$

Hence,

68. $$\langle \bar{x}, v_i\rangle < 0, \qquad \text{for} \quad i = 1, 2,..., k,$$

[7] See Edwards [8], p. 118, 2.2.3 Corollary to the Hahn-Banach Theorem.

which contradicts the assumption of the lemma. Therefore $0 \in B$, i.e., there exists a nonzero vector $\tilde{\rho}$ in E^k, $\tilde{\rho}_i \leqslant 0$, $i = 1, 2, \ldots, k$, such that $\sum_{i=1}^{k} \tilde{\rho}^i \nu_i = 0$.

Combining theorem (60), assumption (58), and lemma (65), we obtain the following extension of the Fritz John Theorem [12].

69. THEOREM. *Let $\hat{x}$ be a solution to the nonlinear programming problem* (55). *Then, there exist vectors μ in E^p, η in E^m, and ρ in E^s such that*

(i) $\mu^i \leqslant 0, \qquad i = 1, 2, \ldots, p,$

(ii) $\rho^i \leqslant 0, \qquad i = 1, 2, \ldots, s,$

(iii) $(\mu, \eta, \rho) \neq 0,$

(iv) $$\sum_{i=1}^{p} \mu^i \frac{\partial f^i(\hat{x})}{\partial x} + \sum_{i=1}^{m} \eta^i \frac{\partial r^i(\hat{x})}{\partial x} + \sum_{i=1}^{k} \rho^i \frac{\partial q^i(\hat{x})}{\partial x} = 0,$$

(v) $$\sum_{i=1}^{k} \rho^i q^i(\hat{x}) = 0.$$

The following corollaries are immediate consequences of theorem (19):

75. COROLLARY. *If the gradient vectors $[\partial r^1(\hat{x})/\partial x], \ldots, [\partial r^m(\hat{x})/\partial x]$ are linearly independent; then any vectors $\mu \in E^p$, $\eta \in E^m$, $\rho \in E^s$, satisfying the conditions of theorem* (69), *also satisfy* $(\mu, \rho) \neq 0$.

76. COROLLARY. *If the gradient vectors $[\partial r^1(\hat{x})/\partial x]$, $[\partial r^2(\hat{x})/\partial x], \ldots$, $[\partial r^m(\hat{x})/\partial x]$ together with the gradient vectors $[\partial q^i(\hat{x})/\partial x]$, with $i \in I(\hat{x})$, are linearly independent, then any vectors $\mu \in E^p$, $\eta \in E^m$, $\rho \in E^s$ satisfying the conditions of theorem* (69), *also satisfy* $\mu \neq 0$.

77. COROLLARY. *If the set*

$$\left\{ x \in E^n \;\middle|\; \frac{\partial r^j(\hat{x})}{\partial x} x = 0, \quad j = 1, 2, \ldots, m, \frac{\partial q^i(\hat{x})}{\partial x} x < 0, \quad i \in I(\hat{x}) \right\}$$

is nonvoid and the vectors $[\partial r^1(\hat{x})/\partial x]$, $[\partial r^2(\hat{x})/\partial x], \ldots$, $[\partial r^m(\hat{x})/\partial x]$ are linearly independent, then any vectors $\mu \in E^p$, $\eta \in E^m$, $\rho \in E^s$ satisfying the conditions of theorem (69), *also satisfy* $\mu \neq 0$.

78. COROLLARY. *If the system*

$$\frac{\partial f^i(\hat{x})}{\partial x} x < 0, \qquad i \in \{\{1, 2, \ldots, p\} - \{\hat{\imath}\}\},$$

$$\frac{\partial r^j(\hat{x})}{\partial x} x = 0, \qquad j = 1, 2, \ldots, m,$$

$$\frac{\partial q^k(\hat{x})}{\partial x} x < 0, \qquad k \in I(\hat{x}),$$

has a solution for some $i \in \{1, 2,..., p\}$ *and the gradient vectors* $[\partial r^1(\hat{x})/\partial x]$, $[\partial r^2(\hat{x})/\partial x],..., [\partial r^m(\hat{x})/\partial x]$ *are linearly independent, then any vectors* $\mu \in E^p$, $\eta \in E^m$, $\rho \in E^s$ *satisfying the conditions of theorem* (69), *also satisfy* $\mu^i < 0$.

IV. Applications to Optimal Control

79. Definition. *Let P be a convex cone in E^s. A subset S of E^s is said to be P-directionally convex if for every z_1, z_2 in S and $0 \leqslant \lambda \leqslant 1$, there exists a vector $z(\lambda)$ in P such that*

$$\lambda z_1 + (1 - \lambda) z_2 + z(\lambda) \in S.$$

80. Remark. It is very easy to show that a subset S of E^s is P-directionally convex if and only if for any finite subset $\{z_1, z_2,..., z_k\}$ of S and any scalars $\{\lambda_1, \lambda_2,..., \lambda_k\}$ with $\sum_{i=1}^{k} \lambda_j = 1$, $\lambda_i \geqslant 0$, $i = 1, 2,..., k$, there exists a vector $z(\lambda_1, \lambda_2,..., \lambda_k)$ in P such that

$$\sum_{i=1}^{k} \lambda_i z_i + z(\lambda_1, \lambda_2,..., \lambda_k) \in S.$$

On rereading theorem (7), we observe that it may be rephrased in the following equivalent form.

81. Theorem. *Let $\bar{x}$ be any feasible solution to the canonical problem* (1), *i.e.,* $\bar{x} \in \Omega$ *and* $r(\bar{x}) = 0$, *and let* $C(\bar{x}, \Omega)$ *be a conical approximation for* Ω *at* $\bar{x}$. *If the sets*

$$K(\bar{x}) = \left\{u \in E^p \times E^m \;\middle|\; u = \left(\frac{\partial f(\bar{x})}{\partial x} x, \frac{\partial r(\bar{x})}{\partial x} x\right), \quad x \in C(\bar{x}, \Omega)\right\}$$

and

$$R = \{(y, 0) \in E^p \times E^m \mid y^i < 0, \quad i = 1, 2,..., p, \quad 0 \in E^m\}$$

are not separated, then there exists a vector x^* *in* Ω, *with* $r(x^*) = 0$ *and* $f(x^*) < f(\bar{x})$ *(component-wise).*

We now make one more observation.

82. Theorem. *Let* $\Omega' \subset E^n$ *be any set with the property that if* $x' \in \Omega'$, *then there is a vector* x *in* Ω *with* $r(x') = r(x)$ *and* $f(x) \leqslant f(x')$. *If* $\hat{x}$ *is a solution to the canonical problem* (1), *if* $\hat{x} \in \Omega'$ *and if* $C(\hat{x}, \Omega')$ *is a conical approximation*

for Ω' at $\hat{x}$, then there exists a vector μ in E^p and a vector η in E^m such that

83. $$\mu^i \leqslant 0, \qquad i = 1, 2, ..., p,$$
84. $$(\mu, \eta) \neq 0,$$

and

85. $$\left\langle \mu, \frac{\partial f(\hat{x})}{\partial x} x \right\rangle + \left\langle \eta, \frac{\partial r(\hat{x})}{\partial x} x \right\rangle \leqslant 0 \qquad \textit{for all} \quad x \in \overline{C(\hat{x}, \Omega')}.$$

PROOF. The theorem claims that the cones

$$K'(\hat{x}) = \left\{ u \in E^p \times E^m \;\middle|\; u = \left(\frac{\partial f(\hat{x})}{\partial x} x, \frac{\partial r(\hat{x})}{\partial x} x \right), \quad x \in C(\hat{x}, \Omega') \right\}$$

and

$$R = \{(y, 0) \in E^p \times E^m \mid y^i < 0 \qquad \text{for} \quad i = 1, 2, ..., p, \quad 0 \in E^m\}$$

must be separated if $\hat{x}$ is a solution to the canonical problem (1). Suppose that $K'(\hat{x})$ and R are not separated. Then by theorem (81) with Ω' taking the place of Ω, there exists a x^* in Ω' such that $r(x^*) = 0$ and $f(x^*) < f(\hat{x})$. But by assumption, there must exist an $\bar{x}$ in Ω such that $r(\bar{x}) = r(x^*) = 0$ and $f(\bar{x}) \leqslant f(x^*) < f(\hat{x})$, which contradicts the assumption that $\hat{x}$ is a solution to the canonical problem (1).

Now consider a dynamical system described by the difference equation

86. $$x_{i+1} - x_i = f_i(x_i, u_i) \qquad \text{for} \quad i = 0, 1, 2, ..., k - 1,$$

where $x_i \in E^n$ is the system state at time i, $u_i \in E^m$ is the system input at time i, and f_i is a function defined in $E^n \times E^m$ with range in E^n.

The optimal control problem is that of finding a control sequence $\hat{\mathscr{U}} = (\hat{u}_0, \hat{u}_1, ..., \hat{u}_{k-1})$ and a corresponding trajectory $\hat{\mathscr{X}} = (\hat{x}_0, \hat{x}_1, ..., \hat{x}_k)$ determined by (86), such that

87. $$\hat{u}_i \in U_i \subset E^m, \qquad i = 0, 1, 2, ..., k - 1,$$

88. $\hat{x}_i \subset X_0 = X_0' \cap X_0''$, with $X_0' = \{x \in E^n \mid q_0(x) \leqslant 0\}$, and $X_0'' = \{x \in E^n \mid q_0(x) = 0\}$, where g_0 maps E^n into E^{ℓ_0} and q_0 maps E^n into E^{m_0},

89. $\hat{x}_k \in X_k = X_k' \cap X_k''$, with $X_k' = \{x \in E^n \mid q_n(x) \leqslant 0\}$ and $X'' = \{x \in E^n \mid g_k(x) = 0\}$, where g_k maps E^n into E^{l_k} and g maps E^n into E^{m_k},

90. $\hat{x}_i \in X_i = X_i', X_i' = \{x \in E^n \mid q_i(x) \leqslant 0\}, \qquad i = 1, 2, ..., k - 1$

where q_i maps E^n into E^{m_i}, and

91. for every control sequence $\mathscr{U} = (u_0, u_1, \ldots, u_{k-1})$ and corresponding trajectory $\mathscr{X} = (x_0, x_1, \ldots, x_k)$, satisfying the conditions (87), (88), (89), and (90), the relation $\sum_{i=0}^{k-1} c_i(x_i, u_i) \leqslant \sum_{i=0}^{k-1} c_i(\hat{x}_i, \hat{u}_i)$ implies that $\sum_{i=0}^{k-1} c_i(x_i, u_i) = \sum_{i=0}^{k-1} c_i(\hat{x}_i, \hat{u}_i)$, where the c_i map E^n into E^p for $i = 0, 1, 2, \ldots, k-1$.

The following assumptions will be made:

92. for $i = 0, 1, 2, \ldots, k-1$ and for every fixed u_i in U_i, the functions $f_i(x_i, u_i)$ and $c_i(x_i, u_i)$ are continuously differentiable in x_i;

93. let $\bar{R} = \{(y, 0) \in E^p \times E^m \mid y \in E^p,\ y^i \leqslant 0, i = 1, 2, \ldots, p, 0 \in E^m\}$ and let $\mathbf{f}_i(x, u) = (c_i(x, u), f_i(x, u))$; then for each x in E^n, the sets $\mathbf{f}_i(x_i, U_i)$, $i = 0, 1, 2, \ldots, k-1$ are $\bar{R}$-directionally convex;

94. the functions $g_0(x)$ and $g_k(x)$ are continuously differentiable and the corresponding Jacobian matrices $[\partial g_0(x)/\partial x]$, $[\partial g_k(x)/\partial x]$ are of maximum rank for every x in X_0 and every x in X_k, respectively; and

95. for every $x_i \in X_i'$, $i = 0, 1, 2, \ldots, k$,

$$\left\{ \frac{\partial q_i^j(x)}{\partial x} \,\middle|\, j \in \{j \mid q_i^j(x) = 0, \quad j = 1, 2, \ldots, m_i\} \right\}$$

is a set of linearly independent vectors.

In order to transcribe the control problem into the form of the canonical problem, we introduce the following definitions:

96. For $i = 0, 1, 2, \ldots, k-1$, let $\mathbf{v}_i = (a_i, v_i)$ where $a_i \in c_i(x_i, U_i)$ and $v_i \in f_i(x_i, U_i)$, i.e., $\mathbf{v}_i \in \mathbf{f}_i(x_i, U_i)$.

97. Let

$$z = (x_0, x_1, \ldots, x_k, \mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{k-1}).$$

98. Let

$$f(z) = \sum_{i=0}^{k-1} a_i .$$

99. Let $r(z)$ be the function defined by

$$r(z) = \begin{bmatrix} x_1 - x_0 - v_0 \\ \vdots \\ x_k - x_{k-1} - v_{k-1} \\ g_0(x_0) \\ g_k(x_k) \end{bmatrix}.$$

100. Let

$$\Omega = \{z \mid x_i \in X_i', \quad i = 0, 1, 2, ..., k, \qquad \mathbf{v}_i \in \mathbf{f}_i(x_i, U_i), \quad i = 0, 1, ..., k-1\}.$$

Thus, the optimal control problem is equivalent to the canonical problem with z, f, r, and Ω given by (97), (98), (99), and (100), respectively.

Let us define the set Ω' by

101. $$\Omega' = \{z \mid x_i \in X_i', \quad i = 0, 1, 2, ..., k,$$
$$\mathbf{v}_i \in \mathrm{co}\, \mathbf{f}_i(x_i, U_i), \quad i = 0, 1, 2, ..., k-1\}.$$

We now show that the sets Ω and Ω' defined in (100) and (101), respectively, satisfy the conditions stated in theorem (82). Let z^* be any point in Ω'. Then for $i = 0, 1, 2, ..., k$, $x_i^* \in X_i'$ and $\mathbf{v}_i^* = \sum_{j \in J^*} \lambda_i^j \mathbf{v}_i^j$, where $\sum_{j \in J^*} \lambda_i^j = 1$, $\lambda_i^j \geqslant 0$, J^* a finite set and $\mathbf{v}_i^j \in \mathbf{f}_i(x_i, U_i)$. But by assumption (93), the sets $\mathbf{f}_i(x_i, U_i)$, $i = 0, 1, 2, ..., k-1$, are $\bar{R}$-directionally convex and hence there exists a $\tilde{z}$ in Ω such that $\tilde{x}_i = x_i^*$, $\tilde{v}_i = v_i^*$, and $\tilde{a}_i \leqslant a_i^*$.

Now let $\hat{z}$ be a solution to the optimal control problem. Then $\hat{z} \in \Omega$ and, since Ω' contains Ω, $\hat{z} \in \Omega'$.

In the appendix we prove that the set

102. $$C(\hat{z}, \Omega')$$
$$= \Big\{ \delta z = (\delta x_0, \delta x_1, ..., \delta x_k, \delta \mathbf{v}_0, \delta \mathbf{v}_1, ..., \delta \mathbf{v}_{k-1}) \;\Big|\; \frac{\partial q_i^j(\hat{x}_i)}{\partial x_i} \delta x_i < 0$$
$$\text{for all} \quad j \in \{j \mid q_i^j(\hat{x}_i) = 0\} \qquad \text{and} \qquad \delta \mathbf{v}_i \in \left\{ \frac{\partial \mathbf{f}_i(\hat{x}_i, \hat{u}_i)}{\partial x_i} \delta x_i \right\}$$
$$+ RC(\hat{\mathbf{v}}_i, \mathrm{co}\, \mathbf{f}_i(\hat{x}_i, U_i)) \Big\}^{7}$$

is a conical approximation for the set Ω' at $\hat{z}$.

It now follows from theorem (82) that there exists a nonzero vector $\psi = (p^0, \pi)$, $p^0 \in E^{\nu}$, $p^0 \leqslant 0$, $\pi = (-p_1, -p_2, ..., -p_k, \mu_0, \mu_k)$, $p_i \in E^n$, $\mu_0 \in E^{\ell_0}$, $\mu_k \in E^{\ell_k}$ such that

103. $$p^0 \frac{\partial f(\hat{z})}{\partial z} \delta_z + \pi \frac{\partial r(\hat{z})}{\partial z} \delta z \leqslant 0 \qquad \text{for all} \quad \delta z \in \overline{C(\hat{z}, \Omega')}.$$

[7] Definition. *Given a subset A of an Euclidean space, we define the radial cone to A at $\bar{x} \in A$ to be the cone*

$$RC(\bar{x}, A) = \{x \mid (\bar{x} + \alpha x) \in A \quad \textit{for all} \quad 0 < \alpha < \epsilon(\bar{x}, x), \quad \textit{where} \quad \epsilon > 0\}$$

Substituting (98) and (99) into (103) we obtain

104. $$p^0 \sum_{i=0}^{k-1} \delta a_i - \sum_{i=0}^{k-1} p_{i+1}(\delta x_{i+1} - \delta x_i - \delta v_i) + \mu_0 \frac{\partial g_0(\hat{x}_0)}{\partial x} \delta x_0 + \mu_k \frac{\partial g_k(\hat{x}_k)}{\partial x} \delta x_k \leqslant 0$$

for every $\delta z \in \overline{C(\hat{z}, \Omega')}$.

Now, by interpreting (104) we obtain the following theorem:

105. THEOREM. *If the control sequence $\hat{\mathscr{U}} = (\hat{u}_0, \hat{u}_1, ..., \hat{u}_{k-1})$ and the corresponding trajectory $\hat{\mathscr{X}} = (\hat{x}_0, \hat{x}_1, ..., \hat{x}_k)$ constitute a solution to the optimal control problem, then there exists a vector $p^0 \in E^p$, $p^0 \leqslant 0$, vectors $p_0, p_1, ..., p_k$ in E^n, vectors $\lambda_0, \lambda_1, ..., \lambda_k$, $\lambda_i \in E^{n_i}$, $i = 0, 1, ..., k$, vectors $\mu_0 \in E^{\ell_0}$, $\mu_k \in E^{\ell_k}$ such that*[8]

106. (i) $$(p^0, p_0, p_1, ..., p_k, \mu_0, \mu_k) \neq 0,$$

107. (ii) $$p_i - p_{i+1} = p_{i+1} \frac{\partial f_i(\hat{x}_i, \hat{u}_i)}{\partial x} + p^0 \frac{\partial c_i(\hat{x}_i, \hat{u}_i)}{\partial x} + \lambda_i \frac{\partial q_i(\hat{x}_i)}{\partial x}, \quad i = 0, 1, ..., k-1.$$

108. (iii) $$p_0 = -\mu_0 \frac{\partial g_0(\hat{x})}{\partial x},$$

109. (iv) $$p_k = \mu_k \frac{\partial g_k(\hat{x}_k)}{\partial x} + \lambda_k \frac{\partial g_k(\hat{x}_k)}{\partial x},$$

110. (v) $$\lambda_i q_i(\hat{x}_i) = 0, \qquad i = 0, 1, ..., k,$$

and

111. (vi) *the Hamiltonian*
$$H(x, u, p, p^0, i) = \langle p^0, c_i(x, u)\rangle + \langle p, f_i(x, u)\rangle$$
satisfies the maximum principle
$$H(\hat{x}_i, \hat{u}_i, p, p^0, i) \geqslant H(\hat{x}_i, u_i, p, p^0, i) \quad \textit{for all } u_i \in U_i, \quad i = 0, 1, ..., k-1.$$

PROOF.

(i) This was established in theorem (82)

(ii) Let
$$\delta \mathbf{v}_i = \frac{\partial \mathbf{f}_i(\hat{x}_i, \hat{u}_i)}{\partial x} \delta x_i.$$

Then (104) becomes

$$p^0 \frac{\partial c_i(\hat{x}_i, \hat{u}_i)}{\partial x} \delta x_i + p_{i+1} \frac{\partial f_i(\hat{x}_i, \hat{u}_i)}{\partial x} \delta x_i + p_{i+1} \delta x_i - p_i \delta x_i \leqslant 0$$

[8] Note that the p_i and λ_i are row vectors.

for every δx_i satisfying $[\partial q_i^j(\hat{x}_i)/\partial x]\,\delta x_i \leqslant 0$, with $q_i^j(\hat{x}_i) = 0$. Applying Farkas' lemma [11] we obtain (107) and that $\lambda_i q_i(\hat{x}_i) = 0$ for $i = 0, 1,\ldots, k-1$.

(iii) This is seen to be merely an arbitrary but consistent definition.

(iv) and (v). We select $\delta z = (0, 0,\ldots, 0, \delta x_k, 0, 0,\ldots, 0)$, with δx_k such that $[\partial q_k^j/\partial x]\,\delta x_k \leqslant 0$ whenever $q_k^j(\hat{x}_k) = 0$. Again applying Farkas' lemma, we get (109) and $\lambda_k q_k(\hat{x}_k) = 0$.

(vi) For $i = 0, 1, 2,\ldots, k-1$, let $\mathbf{v}_i'$ be an arbitrary point in $\operatorname{co}\mathbf{f}_i(\hat{x}_i, U_i)$, which is convex by construction. Then $\delta\mathbf{v}_i = \mathbf{v}_i' - \mathbf{v}_i$ is in $RC(\hat{v}_i, \operatorname{co}\mathbf{f}_i(\hat{x}_i, U_i))$ and, choosing $\delta z = (0, 0,\ldots, 0, \delta\mathbf{v}_i, 0,\ldots, 0)$, we find that $\delta z \in C(\hat{z}, \Omega')$, and hence we obtain from (104),

112. $$p^0\,\delta a_i + p_{i+1}\,\delta v_i \leqslant 0.$$

Substituting $\mathbf{v}_i' - \mathbf{v}_i$ for $\delta\mathbf{v}_i$ in (112) we obtain

113. $$p^0(a_i' - c_i(\hat{x}_i, \hat{u}_i)) + p_{i+1}(v_i' - f_i(\hat{x}_i, \hat{u}_i)) \leqslant 0.$$

Clearly (113) also holds for every $(a_i', v_i') \in \mathbf{f}_i(\hat{x}_i, U_i)$. Therefore,

$$p^0(c_i(\hat{x}_i, u_i) - c_i(\hat{x}_i, \hat{u}_i)) + p_{i+1}(f_i(\hat{x}_i, u_i) - f_i(\hat{x}_i, \hat{u}_i)) \leqslant 0 \quad \text{for all } u_i \in U_i,$$

which completes our proof of (111).

Conclusion

In this paper we have concerned ourselves with two important aspects of vector-valued criterion optimization problems. The first was that of developing necessary conditions for the characterization of noninferior points. The necessary conditions we have obtained do not depend on the commonly made, but rather restrictive, convexity assumptions. The second was that of examining the possibility of "scalarization," i.e., of reducing a vector-valued criterion optimization problem to a family of optimization problems with scalar-valued criteria. Since it is known that scalarization by convex weighting of the components of the vector criterion function is not always possible, we have exhibited the relation between the solutions sets of certain vector-criterion problems and scalar-criteria problems derived from then by convex weighting. Finally, we have demonstrated that our results are of a very general nature by showing that they apply with equal ease to a broad class of nonlinear programming problems as well as to optimal control problems.

Since the conditions presented in this paper are considerably more general than hitherto available in the literature, it is hoped that they will open up important classes of optimization problems.

Appendix

A1. Given a subset B of a Euclidean space, defined by inequalities, i.e., $B = \{x \mid q^i(x) \leqslant 0,\ i = 1, 2,..., m\}$, where the q^i are continuously differentiable scalar-valued functions, we define the *internal cone* to B at $\bar{x} \in B$ to be the cone

$$\text{A2.} \quad IC(\bar{x}, B) = \left\{ x \,\middle|\, \frac{\partial q^i(\hat{x})}{\partial x} x < 0 \text{ whenever } q^i(\bar{x}) = 0, \quad i \in \{1, 2,..., m\} \right\}.$$

We now return to the set Ω', which was defined in (101) as

$$\text{A3.} \quad \Omega' = \{z = (x_0, x_1,..., x_k, \mathbf{v}_0, \mathbf{v}_1,..., \mathbf{v}_{k-1}) \mid x_i \in X_i', \quad i = 0, 1, 2,... k, \\ \mathbf{v}_j \in \operatorname{co} \mathbf{f}_j(x_j, U_j), \quad j = 0, 1, 2,..., k-1\}.$$

We shall prove that the set $C(\hat{z}, \Omega')$ defined in (102), as shown below, is a conical approximation for the set Ω' at $\hat{z} \in \Omega'$.

$$\text{A4.} \qquad C(\hat{z}, \Omega') = \Big\{ \delta z = (\delta x_0,..., \delta x_k, \delta \mathbf{v}_0,..., \delta \mathbf{v}_{k-1}) \mid \delta x_i \in IC(\hat{x}_i, X_i') \\ \text{for} \quad i = 0, 1,..., k,$$

and

$$\delta \mathbf{v}_i - \frac{\partial \mathbf{f}_i(\hat{x}_i, \hat{u}_i)}{\partial x} \delta x_i \in RC(\hat{\mathbf{v}}_i, \operatorname{co} \mathbf{f}_i(\hat{x}_i, U_i)) \\ \text{for} \quad i = 0, 1,..., k-1 \Big\}.$$

A5. Lemma. *The set $C(z, \Omega')$ is a conical approximation for the set Ω' at $\hat{z}$.*

Proof. First of all it is clear that $C(\hat{z}, \Omega')$ is a convex cone. Now, for $j = 1, 2,..., N$, let

$$\text{A6.} \qquad \delta z_j = (\delta x_{0j},..., \delta x_{kj}, \delta \mathbf{v}_{0j},..., \delta \mathbf{v}_{k-1j})$$

be N linearly independent vectors in $C(\hat{z}, \Omega')$, and let $S = \operatorname{co} \{\bar{\epsilon} \delta z_1, \bar{\epsilon} \delta z_2,..., \bar{\epsilon} \delta z_N\}$ where $\bar{\epsilon}$ is a positive scalar, defined below.

For any δz in S we can uniquely write

A7. $$\delta z = \bar{\epsilon} \sum_{i=1}^{N} \mu_i(\delta z)\, \delta z_i ,$$

where

$$\sum_{i=1}^{N} \mu_i(\delta z) = 1, \mu_i(\delta z) \geqslant 0, \qquad i = 1, 2,..., N.$$

Therefore,

A8. $$\delta x_i = \bar{\epsilon} \sum_{j=1}^{N} \mu_j(\delta z)\, \delta x_{ij}$$

and

A9. $$\delta \mathbf{v}_i = \bar{\epsilon} \sum_{j=1}^{N} \mu_j(\delta z)\, \delta \mathbf{v}_{ij}$$

But by definition:

A10. $$\delta \mathbf{v}_{ij} = \frac{\partial \mathbf{f}_i(\hat{x}_i , \hat{u}_i)}{\partial x} \delta x_{ij} + \mathbf{v}_{ij} ,$$

where $\mathbf{v}_{ij} \in RC(\hat{\mathbf{v}}_i , \text{co } \mathbf{f}_i(\hat{x}_i , U_i))$

From (A8), (A9), and (A10),

A11. $$\delta \mathbf{v}_i = \frac{\partial \mathbf{f}_i(\hat{x}_i , \hat{u}_i)}{\partial x} \delta x_i + \bar{\epsilon} \sum_{j=1}^{N} \mu_j(\delta z)\, \mathbf{v}_{ij} .$$

Now, let us define the positive scalar $\bar{\epsilon}$.

(a) For $j = 1, 2,..., N$ and $i = 0, 1,..., k$, δx_{ij} belongs to the convex cone $IC(\hat{x}_i , X_i')$. Hence from (A7), $\sum_{j=1}^{N} \mu_j(\delta z)\delta x_{ij}$ is also in $IC(\hat{x}_i , X_i')$ for $i = 0, 1,..., k$. Therefore there exist positive scalars $\bar{\epsilon}_i , i = 0, 1,..., k$, possible depending on $\delta z_1 , \delta z_2 ,..., \delta z_N$, such that

A12. $$\left(\hat{x}_i + \epsilon_i \sum_{j=1}^{N} \mu_j(\delta z)\, \delta x_{ij}\right) \in X_i' \qquad \text{for all} \quad 0 \leqslant \epsilon_i \leqslant \bar{\epsilon}_i$$

(b) Similarly, for $i = 0, 1,..., k - 1$,

A13. $$\sum_{j=1}^{N} \mu_j(\delta z)\, \mathbf{v}_{ij} \in RC(\hat{v}_i , \text{co } \mathbf{f}_i(\hat{x}_i , U_i)),$$

and hence there exist positive scalars ϵ_i, possible depending on $\delta z_1, \delta z_2, \ldots, \delta z_N$, such that

A14. $\hat{\mathbf{v}}_i + \epsilon_i \sum_{j=1}^{N} \mu_j(\delta_z)\, \mathbf{v}_{ij} \in \operatorname{co} \mathbf{f}_i(\hat{x}_i, U_i) \qquad \text{for all} \quad 0 \leqslant \epsilon_i \leqslant \boldsymbol{\epsilon}_i .$

We now define $\bar{\epsilon}$ to be minimum of the scalars $\bar{\epsilon}_i$, $i = 0, 1, \ldots, k$, and $\boldsymbol{\epsilon}_j$, $j = 0, 1, \ldots, k-1$.

From (A14), there exists a finite set A_i and scalars λ_α^i such that

A15. $$\bar{\epsilon} \sum_{j=1}^{N} \mu_j(\delta_z)\, \mathbf{v}_{ij} = \sum_{\alpha \in A_i} \lambda_\alpha^i \mathbf{f}_i(\hat{x}_i, u_i^\alpha) - \hat{\mathbf{v}}_i ,$$

where $u_i^\alpha \in U_i$, $\alpha \in A_i$, and $\sum_{\alpha \in A_i} \lambda_\alpha^i = 1$, $\lambda_\alpha^i \geqslant 0$.

Combining (A15) and (A11) we obtain

A16. $$\delta \mathbf{v}_i = \frac{\partial \mathbf{f}_i(\hat{x}_i, \hat{u}_i)}{\partial x} \delta x_i + \sum_{\alpha \in A_i} \lambda_\alpha^i \mathbf{f}_i(\hat{x}_i, u_i^\alpha) - \hat{\mathbf{v}}_i .$$

We can define a map ζ from S into $\Omega' - \{\hat{z}\}$ by

A17. $$\zeta(\delta z) = (y_0, y_1, \ldots, y_K, \mathbf{w}_0, \mathbf{w}_1, \ldots, \mathbf{w}_{k-1}),$$

where

A18. $$y_i(\delta z) = \delta x_i = \bar{\epsilon} \sum_{j=1}^{N} \mu_j(\delta z)\, \delta x_{ij} , \qquad i = 0, 1, \ldots, k,$$

and

A19. $$\mathbf{w}_i(\delta z) = \sum_{\alpha \in A_i} \lambda_\alpha^i \mathbf{f}_i(\hat{x}_i + \delta x_i, u_i^\alpha) - \hat{\mathbf{v}}_i , \qquad i = 0, 1, \ldots, k-1.$$

From (A12), (A15), (A18), and (A19) it is clear that ζ maps S into $\Omega' - \{\hat{z}\}$. Expanding (A19) in a Taylor series about $\hat{z}$ we find that

A20. $$\mathbf{w}_i(\delta z) = \frac{\partial \mathbf{f}_i(\hat{x}_i, \hat{u}_i)}{\partial x} \delta x_i + \sum_{\alpha \in A_i} \lambda_\alpha^i \mathbf{f}_i(\hat{x}_i, u_i^\alpha) - \hat{\mathbf{v}}_i + o_i(\delta z),$$

where

$$\frac{\| o_i(\delta z) \|}{\| \delta z \|} \to 0 \qquad \text{as} \qquad \| \delta z \| \to 0.$$

Combining (A20), (A13), (A17), and (A18), we see that

$$\zeta(\delta z) = \delta z + o(\delta z),$$

where

$$\lim_{\|\delta z\|\to 0} \frac{\| o(\delta z)\|}{\| \delta z \|} = 0$$

Since ζ is obviously continuous, $C(\hat{z}, \Omega')$ is a conical approximation for Ω' at $\hat{z}$.

References

1. V. Pareto. "Cours d'Économie Politique." Lausanne, Rouge, 1896.
2. S. Karlin. "Mathematical Methods and Theory in Games, Programming and Economics," Vol. 1. Addison-Wesley, Massachusetts, 1959.
3. G. Debreu. "Theory of Value." Wiley, New York, 1959.
4. H. W. Kuhn and A. W. Tucker. "Nonlinear Programming, Proc. of the Second Berkeley Symposium on Mathematic Statistics and Probability," pp. 481-492. Univ. of California Press, Berkeley, California, 1951.
5. L. A. Zadeh. Optimality and nonscalar-valued performance criteria. *IEEE Trans. Automat. Control* **AC-8** (1963), 59-60.
6. S. S. L. Chang. General theory of optimal processes. *J. SIAM Control* **4** (1966), 46-55.
7. A. Klinger. Vector-valued Performance Criteria. *IEEE Trans. Automat. Control* (1964), 117-118.
8. R. E. Edwards. "Functional Analysis Theory and Applications." Holt, Rinehart and Winston, New York, 1965.
9. J. Dieudonné. "Foundations of Modern Analysis." Academic Press, New York, 1960.
10. K. Fan. Fixed point and minimax theorems in locally convex topological linear spaces. *Proc. Natl. Acad. Sci., U. S.* **38** (1952), 121-126.
11. J. Farkas. Über die Theorie der einfachen Ungleichungen. *J. Reine Angew. Math.* **124** (1901), 1-27.
12. F. John. Extremum problems with inequalities as side conditions. *In* "Studies and Essays, Courant Anniversary Volume" (K. O. Friedrichs, O. E. Neugebauer, and J. J. Stoker, eds.) pp. 187-204. Wiley, New York, 1948.
13. M. Canon, C. Cullum, and E. Polak. Constrained minimization problems in finite dimensional spaces. *J. SIAM Control* **4** (1966), 528-547.
14. J. M. Holtzman. On the maximum principle for nonlinear discrete-time systems. *IEEE Trans.* AC-11, No. 2 (1966), 273-274.
15. H. Halkin. A maximum principle of the pontryagin type for systems described by nonlinear difference equations. *J. SIAM Control* **4** (1966), 90-111.
16. O. L. Mangasarian and S. Fromowitz. The Fritz John necessary conditions of optimality in the presence of equality and inequality constraints. Shell Dev. Co. Paper P1433, 1965.