AD/A-002 959

ON A CHARACTERIZATION OF OPTIMALITY IN CONVEX PROGRAMMING

A. Ben-Israel, et al

Texas University at Austin

Prepared for:

Office of Naval Research
Technion - Israel Institute of Technology

October 1974

Unclassified
Security Classification

i

AD/A-002959

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Center for Cybernetic Studies University of Texas at Austin	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE
On a Characterization of Optimality in Convex Programming

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

5. AUTHOR(S) (First name, middle initial, last name)
A. Ben-Israel
A. Ben-Tal

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
October 1974	18	5

8a. CONTRACT OR GRANT NO. N00014-67-A-0126-0008;0009 b. PROJECT NO. NR 047-021 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) Research Report CCS 196 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT
This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Office of Naval Reseaarch (Code 434) Washington, D. C.

13. ABSTRACT

Necessary and sufficient conditions for optimality are given, for convex programming problems, without constraint qualification, in terms of a single mathematical program, which can be chosen to be bilinear.

DD FORM 1 NOV 65 1473 (PAGE 1)
S/N 0101-807-6811

Unclassified
Security Classification

A-31408

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Convex programming						
Bilinear programming						
Linear programming						

Research Report
CCS 196

ON A CHARACTERIZATION OF OPTIMALITY IN CONVEX PROGRAMMING

by

A. Ben-Israel*
A. Ben-Tal**

October 1974

*Technion, Israel Institute of Technology, Haifa, Israel
**Center for Cybernetic Studies, The University of Texas at Austin

This research was partly supported by Project No. NR 047-[illegible]. ONR Contracts N00014-67-A-0126-0008 and N00014-67-A-0126-[illegible] with the Center for Cybernetic Studies, The University of Texas. Reproduction in whole or in part is permitted for any purpose of the United States Government.

CENTER FOR CYBERNETIC STUDIES

A. Charnes, Director
Business-Economics Building, 512
The University of Texas
Austin, Texas 78712
(512)471-1821
(512)471-4894

ABSTRACT

Necessary and sufficient conditions for optimality are given, for convex programming problems, without constraint qualification, in terms of a single mathematical program, which can be chosen to be bilinear.

1. Introduction

This paper is a sequel to [1], where optimality conditions for convex programming, not requiring constraint qualification, were given in terms of a family of linear programs, expressing the "logical" conditions (5) and (6) below.

Here the same is achieved by a single problem, which depends on some positive-definite functions to be chosen. For the case where the constraint functions are strictly convex in their "actual" variables, this characterization of optimality is given in § 2. In particular, it is possible to characterize optimality by the single bilinear program (AL), given following Example 1 below.

For the convex case, we give a sample result in § 3, characterizing optimality in the case where the constraint functions are faithfully convex, [4].

2. The strictly convex case.

For a given function $f^k: R^n \to R$, we define its restriction $f^{\lceil k \rceil}$ as follows. Let $\lceil k \rceil \subset \{1, 2, \ldots, n\}$ denote the index set of the variables $\{x_j\}$ on which f^k actually depends

$$\lceil k \rceil \stackrel{\Delta}{=} \{j: \text{ There exist } x_i = \xi_i,\ i \neq j, \text{ such that the function } f^k(\xi_1, \ldots, \xi_{j-1}, \cdot, \xi_{j+1}, \ldots, \xi_n) \text{ is not a constant}\}.$$

For any $x \in R^n$ the subvector $x_{\lceil k \rceil}$ is obtained by deleting the components $\{x_j: j \notin \lceil k \rceil\}$. The restriction $f^{\lceil k \rceil}$ is the function

$f^{[k]}: R^{card[k]} \to R$ obtained by restricting f^k to $x_{[k]}$.

Consider the programming problem

$$\text{(P)} \qquad \min f^0(x)$$
$$\text{s.t. } f^k(x) \le 0 \qquad k \in P \stackrel{\Delta}{=} \{1, 2, \ldots, p\}.$$

For a feasible solution x^*, i.e.,

$$(1) \qquad f^k(x^*) \le 0, \qquad k \in P,$$

we denote the set of binding constraints by

$$(2) \qquad P^* \stackrel{\Delta}{=} \{k: k \in P, f^k(x^*) = 0\}.$$

A characterization of optimality is given in the following

Theorem 1. Let

(i) the problem (P) have convex functions $\{f^k: k \in \{0\} \cup P\}$ assumed differentiable,

(ii) x^* be a feasible solution of (P) at which the restrictions of the binding constraints $f^{[k]}$, $k \in P^*$, are strictly convex[1],

(iii) $\phi^k: R^{card[k]} \to R$ be any positive definite function, i.e.

$$\phi^k(z) > 0, \quad 0 \ne z \in R^{card[k]}$$
$$\phi^k(0) = 0 \qquad\qquad k \in P^*.$$

Then x^* is optimal if and only if $\sigma = 0$ is the optimal value of the program

---

[1]This assumption is weaker than strict convexity of the functions $\{f^k: k \in P^*\}$, unless $[k] = \{1, \ldots, n\}$ for all $k \in P^*$.

max α

s.t.

(3) $$d^t \nabla f^o(x^*) + \alpha \leq 0$$

(4) $$d^t \nabla f^k(x^*) + \alpha\phi^k(d_{[k]}) \leq 0, \quad k \in P^*.$$

Proof.

Let d stand for directions such that, for $0 < \varepsilon$ sufficiently small, $x^* + \varepsilon d$ is feasible and $f^o(x^* + \varepsilon d) < f^o(x^*)$. Then the optimality of x^* is equivalent to the nonexistence of such d. Using the convexity properties of f^o and $\{f^{[k]}: k \in P^*\}$ it follows that the optimality of x^* is equivalent[2] to the nonexistence of d satisfying

(5) $$d^t \nabla f^o(x^*) < 0$$

(6) $$d^t \nabla f^k(x^*) \leq 0$$

with equality only if $d_{[k]} = 0$, $k \in P^*$.

If.

Let x^* be non optimal, i.e., let there exist a $\bar{d}$ satisfying (5) and (6). Let

$$\bar{\alpha} \triangleq -\min\{\bar{d}^t \nabla f^o(x^*), \ \max\{\frac{\bar{d}^t \nabla f^k(x^*)}{\phi^k(\bar{d}_{[k]})} : \bar{d}^t \nabla f^k(x^*) < 0\}\}.$$

Then $\bar{\alpha}$ is positive and

---

[2] The details are as in the proof of [1, Theorem 1].

(7) $$\bar{d}^t \nabla f^o(x^*) + \bar{\sigma} \leq 0$$

(8) $$\bar{d}^t \nabla f^k(x^*) + \bar{\sigma}\, \varphi^k(\bar{d}_{[k]}) \leq 0, \quad k \in P^*,$$

showing that the program (A) has a positive optimal value.

Only if.

Let the program (A) have a positive optimal value, i.e., let there exist a vector $\bar{d}$ and a scalar $\bar{\sigma}$ satisfying (7) and (8). Then

$$\bar{d}^t \nabla f^o(x^*) \leq -\bar{\sigma} < 0$$

and

$$\bar{d}^t \nabla f^k(x^*) \leq -\bar{\sigma}\, \varphi^k(\bar{d}_{[k]}) \leq 0, \; k \in P^*,$$

so that

$$\bar{d}^t \nabla f^k(x^*) = 0 \Rightarrow \varphi^k(\bar{d}_{[k]}) = 0$$

$$\Rightarrow \bar{d}_{[k]} = 0, \text{ since } \varphi^k \text{ is positive definite.}$$

Therefore $\bar{d}$ satisfies (5) and (6) showing that x^* is not optimal.

Remarks

1. The convexity assumptions in Theorem 1, and in related results below, can be weakened in the manner of [3].

2. Similarly, differentiability is not essential here since the results can be stated in terms of directional derivatives.

3. Since $d = 0$, $\sigma = 0$ is a feasible solution of (A), the optimal value of (A) is clearly nonnegative. If nonzero, this optimal value is unbounded. It could be bounded (if so desired) by normalizing d, say

$$(9) \qquad -1 \le d_i \le 1, \qquad i = 1, \ldots, n.$$

It should be noted that our results hold in cases where classical optimality conditions, [2] (which do require some constraint qualification) fail. This is illustrated in the following

Example 1. The problem is

$$\begin{aligned}
\min \quad & f^0(x) = e^{x_1} + e^{-x_2} + x_3 \\
\text{s.t.} \quad & f^1(x) = e^{x_1} - 1 \le 0 \\
& f^2(x) = e^{-x_2} - 1 \le 0 \\
& f^3(x) = (x_1-1)^2 + x_2^2 - 1 \le 0 \\
& f^4(x) = x_1^2 + x_2^2 + e^{-x_3} - 1 \le 0.
\end{aligned}$$

Here the sets $[k]$, $k \in P$, are $[1] = \{1\}$, $[2] = \{2\}$, $[3] = \{1, 2\}$ and $[4] = \{1, 2, 3\}$. The restrictions $f^{[k]}$, $k \in P$, are strictly convex.[3] The feasible solutions are

$$\begin{bmatrix} 0 \\ 0 \\ x_3 \end{bmatrix} : \quad x_3 > 0$$

and the optimal solution is $x^* = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, at which point the Kuhn-Tucker condition

$$\nabla f^0(x^*) + \sum \lambda_i \nabla f^i(x^*) = 0, \quad \lambda_i \ge 0,$$

---

[3] Note that the original functions f^1, f^2, and f^3 are not strictly convex.

does not hold since $\nabla f^0(x^*) = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$, $\nabla f^1(x^*) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\nabla f^2(x^*) = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$,

$\nabla f^3(x^*) = \begin{bmatrix} -2 \\ 0 \\ 0 \end{bmatrix}$, $\nabla f^4(x^*) = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}$.

Choosing the positive definite functions φ^k of Theorem 1 as the ℓ_1 - norm

$$\varphi^k(z) = \sum_i |z_i|, \qquad k \in P^*, \tag{10}$$

problem (A) becomes

$$\begin{array}{ll} \max \sigma & \\ \text{s.t.} & \\ & d_1 - d_2 + d_3 + \sigma \leq 0 \\ & d_1 + \sigma|d_1| \leq 0 \\ & -d_2 + \sigma|d_2| \leq 0 \\ & -2d_1 + \sigma(|d_1| + |d_2|) \leq 0 \\ & -d_3 + \sigma(|d_1| + |d_2| + |d_3|) \leq 0 \end{array}$$

whose optimal solution can be found, by inspection, to be $\sigma = 0$.

In applying Theorem 1, the positive definite functions $\{\varphi^k : k \in P^*\}$ should be chosen so as to simplify the problem (A) as much as possible. Such a choice is the ℓ_1 - norm (10) for which the problem (A) reduces to the following bilinear program

(AL) $\max \sigma$

s.t.

$$d^t \nabla f^o(x^*) + \sigma \leq 0 \tag{11}$$

$$d^t \nabla f^k(x^*) + \sigma \sum_{i \in \Gamma_k} |d_i| \leq 0, \quad k \in P^*, \tag{12}$$

whose constraints, or fixed σ, are in fact linear.

But the case where problem (A) of Theorem 1 assumes the simplest form, i.e., a linear program, is where

$$\{j: \frac{\partial f^{o}(x^*)}{\partial x_j} \neq 0\} \subset [k], \qquad \forall\, k \in P^*. \tag{13}$$

This incidence condition, of the type studied in [1, 14], implies that any d satisfying (5) cannot satisfy (6) with an equality. Thus only strict inequalities need be checked in (6), and the optimality of x^* is therefore equivalent to the nonconsistency of the system

$$d^t \, \nabla f^{o}(x^*) < 0 \tag{5}$$

$$d^t \, \nabla f^{k}(x^*) < 0, \qquad k \in P^*, \tag{14}$$

which by the theorem of the alternative is equivalent to the consistency of

$$\sum_{i \in \{0\} \cup P^*} \lambda_i \nabla f^{i}(x^*) = 0$$

(15)

$$\lambda_i \geq 0, \text{ at least one } \lambda_i \neq 0, \; i \in \{0\} \cup P^*.$$

known as the Fritz John condition.

The incidence condition (13) is a special case of the regularization conditions studied in [1], under which the consistency of (15) characterizes the optimality of x^*. Other regularization conditions are the well known constraint qualifications which guarantee the necessity of the Kuhn-Tucker condition.

The following theorem gives an alternative characterization of optimality. Its proof will be omitted, since it much resembles the proof of Theorem 1.

<u>Theorem 2.</u> Under the assumptions of Theorem 1, the feasible solution x^* is optimal if, and only if, for every positive τ, the optimal value of the following problem is zero.

(B.τ) $$\min \; d^t \nabla f^0(x^*)$$

s.t.

(8) $$d^t \nabla f^k(x^*) + \tau \phi^k(d_{\sigma_k}) \le 0, \quad k \in P^*,$$

(9) $$-1 \le d_i \le 1, \quad i = 1, \ldots, n.$$

<u>Remark.</u>

1. A possible advantage of problem (B.τ) over the previously given problem (A), is that the direction found here is of steeper descent.

2. For any $\tau > 0$, let $d(\tau)$ denote an optimal solution of (B.τ). Clearly

$$\tau_1 < \tau_2 \Rightarrow d(\tau_1)^t \nabla f^0(x^*) \le d(\tau_2)^t \nabla f^0(x^*).$$

Thus the optimality of x^* is equivalent to

(10) $$\liminf_{\tau \to 0^+} d(\tau)^t \nabla f^0(x^*) = 0.$$

3. A special case where only one value of σ, say $\sigma = 1$, needs checking in Theorem 2 is where the functions $\{\varphi^k : k \in P^*\}$ have the property

$$\text{(16)} \qquad \lim_{\epsilon \to 0^+} \frac{\varphi^k(\epsilon z)}{\epsilon} = 0, \quad \forall z.$$

Such a choice, $\varphi^k(z) = \sum z_i^2$, was discussed in [1, Corollary 1.1].

4. The simplest form that problem (B.σ) admits, is a linear program. This is obtained by choosing the positive definite functions $\{\varphi^k : k \in P^*\}$ in (8) as the ℓ_1- norm (10). Then (B.σ) becomes

(Bl.σ) $$\min \; d^t \nabla f^o(x^*)$$

s.t.

$$\text{([illegible])} \qquad d^t \nabla f^k(x^*) + \sigma \sum_{i \in [1,n]} |d_i| \le 0, \quad k \in P^*.$$

$$\text{(9)} \qquad -1 \le d_i \le 1, \quad i = 1, \ldots, n.$$

5. For $\sigma = 0$, problem (B.σ) becomes

(B.0) $$\min \; d^t \nabla f^o(x^*)$$

s.t.

$$d^t \nabla f^k(x^*) \le 0, \quad k \in P^*$$

$$-1 \le d_i \le 1, \quad i = 1, \ldots, n.$$

The fact that here the optimal value is zero

$$\text{(18)} \qquad d(0)^t \; \nabla f^o(x^*) = 0$$

is equivalent to the Kuhn-Tucker condition

$$\text{(19)} \qquad \nabla f^o(x^*) + \sum_{k \in P^*} \lambda_k \nabla f^k(x^*) = 0$$

$$\lambda_k \ge 0, \quad k \in P^*$$

which is sufficient for the optimality of x^*.

6. A heuristic procedure for checking the optimality of a given feasible solution x^* is:

a) Solve the linear program (B.0).

b) If its optimal value is zero then, by the previous remark, x^* is optimal.

c) If (18) does not hold, solve the linear program $(BL.\tau)$ for some small $\tau > 0$.

d) If its optimal value $d(\tau_o)^t \nabla f^o(x^*)$ is negative then x^* is nonoptimal, and $d(\tau_o)$ is a direction of descent. Otherwise, solve $(B.\tau_1)$ for $\tau_1 = \frac{\tau_o}{2}$, etc. Use a reasonable stopping rule.

7. Note that it is possible for (17) to hold, even though $d(o)^t \nabla f^o(x^*) < 0$. To illustrate this we can use any example where the Kuhn Tucker conditions do not hold at an optimal point x^*. Thus for example 1, problem (B.0) becomes

$$
\begin{aligned}
\min\ & d_1 - d_2 - d_3 \\
\text{s.t.}\ & d_1 \le 0 \\
& -d_2 \le 0 \\
& -2d_1 \le 0 \\
& -d_3 \le 0 \\
& -1 \le d_i \le 1 \quad i = 1, 2, 3.
\end{aligned}
$$

The optimal solution here $d(o)^t = (0, 1, 0)$ with optimal value $= -1$.

[illegible]. A result for the convex case.

Consider again the problem

$$\text{(P)} \qquad \min \; f^o(x)$$
$$\text{s.t.} \quad f^k(x) \le 0, \quad k \in P,$$

where the function $\{f^k : k \in \{0\} \cup P\}$ are convex, but without further assumptions on the restrictions $f^{[k]}$. In [1, §5] it was shown that at an optimal solution x^*, the logical condition (6) here becomes

$$(20) \qquad d^t \nabla f^k(x^*) \le 0,$$

with equality only if $d \in D_k^*$, $k \in P^*$,

where D_k^* is the cone of directions of constancy of f^k at x^*, defined by

$$(21) \qquad D_k^* \triangleq \{d : \exists \bar{\alpha} > 0 \ni f^k(x^* + \alpha d) = f^k(x^*), \; \forall \alpha \in [0, \bar{\alpha}]\}.$$

Generally this cone is neither polyhedral nor convex, see, e.g. the examples in [1, §5]. However, this cone is quite manageable for the following important family of convex functions

$$(22) \qquad f^k(x) = g^k(A_k x + b_k) + a_k^t x + \alpha_k$$

where

$g^k : R^m \to R$ is a strictly convex function

$A_k : R^n \to R^m$ is a linear transformation

$b_k \in R^m$

$a_k \in R^n$

$\alpha_k \in R$.

These functions are the faithfully convex functions introduced and studied by Rockafellar in [4], [5]. For the function f^k given by (22) and subject to the above assumptions the cone D_f^* is simply

(23) $$D_k^* = N\begin{pmatrix} A_k \\ a_k \end{pmatrix}$$

the null space of the $(m + 1) \times n$ matrix $\begin{pmatrix} A_k \\ a_k \end{pmatrix}$, independently of x^*.

Thus the analogous result to Theorem 1 is

Theorem 3. Let

(i) the problem (P) have a convex objective function f^o and convex constraint functions $\{f^k: k \in P\}$ of the type (22), all assumed differentiable.

(ii) $\varphi^k: R^{m+1} \to R$ be any positive definite functions, $k \in P$.

Then a feasible solution x^* is optimal if, and only if $\sigma = 0$ is the optimal value of the problem

$$\max \ \sigma$$

$$\text{s.t.} \quad d^t \ \nabla f^o(x^*) + \sigma \le 0$$

$$d^t \ \nabla f^k(x^*) + \sigma \varphi^k\left(\begin{pmatrix} A_k \\ a_k \end{pmatrix} d\right) \le 0$$

Proof.

Follows from (20), (23) as in the proof of Theorem 1.

$\square$

The remaining results of §2 can similarly be adapted to the convex case.

REFERENCES

1. A. Ben-Tal, A. Ben-Israel and S. Zlobec. "Characterization of optimality in convex programming without constraint qualification," Technion's reprint series No. AMT-25, July 1974. Technion, Israel Institute of Technology, Haifa, Israel.

2. H. W. Kuhn and A. W. Tucker. "Nonlinear programming," Proc. Second Berkeley Symp. on Mathematical Statistics and Probability. J. Neyman, editor. University of California Press, Berkeley, 1951.

3. J. Ponstein. "Seven kinds of convexity," SIAM Review, Vol. 9 No. 1, January 1967, 115-119.

4. R. T. Rockafellar. "Ordinary Convex Programs without a duality gap," J. Optimiz. Theory Appl., Vol. 7, No. 3, 1971, 143-148.

5. R. T. Rockafellar. "Some convex programs whose duals are linearly constrained," in Nonlinear Programming, Proc. of Symp. held at MRC, University of Wisconsin, Madison, Wisconsin. J. B. Rosen, O. L. Mangasarian, K. Ritter, editors, Academic Press, New York, 1970.