J. Austral. Math. Soc. 19 (Series B), (1976), 462–468.

SUFFICIENT FRITZ JOHN OPTIMALITY CONDITIONS FOR NONDIFFERENTIABLE CONVEX PROGRAMMING

B. D. CRAVEN and B. MOND

(Received 19 February 1976)

Abstract

The Fritz John necessary conditions for optimality of a differentiable nonlinear programming problem have been shown, given additional convexity hypotheses, to be also sufficient (by Gulati, Craven, and others). This sufficiency theorem is now extended to minimization (suitably defined) of a function taking values in a partially ordered space, and to (convex) objective and constraint functions which are not always differentiable. The results are expressed in terms of subgradients.

1. Introduction

For a constrained minimization problem with differentiable objective and constraint functions, Lagrangean conditions of Fritz John type (see, e.g., [7], [6]) are well known to be necessary for a minimum to be attained, without assuming any constraint qualification, as would be required for Kuhn–Tucker conditions. Under some additional convexity hypotheses, the Fritz John necessary conditions have been shown [1, 2, 10, 11, 5] also to be sufficient for a minimum.

These results are now extended in two directions. First, the objective function now maps into a partially ordered space, instead of the real line **R**, and *minimum* is defined, in terms of a convex cone, as in Craven [4]. This definition of *minimum* has also been discussed, more recently, by Borwein [3], who called it *weak minimum*. Second, the hypothesis that the functions are (linearly Gâteaux) differentiable is given up. Given the convexity hypotheses, which are required in any case for a sufficient Fritz John theorem, Fritz John necessary conditions have been obtained [9], for functions *not* always differentiable, in terms of directional derivatives, and subdifferential (the latter as in [12]). These generalized Fritz John conditions, with suitable convexity, are now shown, in Section 3, to be also *sufficient* for a minimum.

In Section 4, this sufficiency theorem is applied to a minimization problem which involves square roots of quadratic forms (or equivalently seminorms), in both objective function and constraints. Earlier papers (e.g. [8]) have included the square root function (which is not differentiable everywhere) in the objective function only. The present results are given for real spaces only; extensions to complex spaces would be readily possible.

The subdifferential of the square root of a quadratic form is expressed (Section 4, Lemma) in terms of an auxiliary variable w, and additional constraints. This accords with the analogous expressions obtained by other methods in [8], for complex spaces, and suggests how the latter could be extended.

If a minimization algorithm, seeking to compute a "Kuhn–Tucker point" where the Kuhn–Tucker conditions are fulfilled, arrives instead at a point where only the Fritz John conditions are fulfilled, this can still characterize the desired minimum, provided that some appropriate convexity hypotheses are also satisfied.

2. Preliminaries

Let X, W, and Y be real normed spaces (whether finite or infinite dimensional); let X_0 be an open subset of X; let $P \subset W$ and $S \subset Y$ be closed convex cones, such that P has nonempty interior $\operatorname{int} P$. The minimization problem:

$$\text{(MP):} \qquad \underset{x \in X_0}{\text{Minimize}} \ f(x) \text{ such that } -g(x) \in S,$$

where $f\colon X_0 \to W$ and $g\colon X_0 \to Y$, attains a (local) *minimum* at $x = a \in X_0$ if $f(x) - f(a) \notin -\operatorname{int} P$ for all x in a neighbourhood of a satisfying $-g(x) \in S$.

The functions $g\colon X_0 \to Y$ is *S-convex* if

$$\lambda g(x) + (1-\lambda) g(x') - g(\lambda x + (1-\lambda)x') \in S$$

whenever $x, x' \in X_0$ and $0 < \lambda < 1$; g is *strictly S-convex* iff g is $(\operatorname{int} S)$-convex; the latter requires that $\operatorname{int} S \neq \emptyset$. A function $h\colon X_0 \to \mathbf{R}$ is then *convex* if it is $\mathbf{R}_+$-convex, where $\mathbf{R}_+ = [0, \infty)$. The convex function h has a *subgradient* u at $a \in X_0$ if

$$h(x) - h(a) \geqq u(x-a)$$

for all $x \in X_0$; the subdifferential $\partial h(a)$ denotes the set of all subgradients u of h at a. The *directional derivative* of h at a in the direction $x \in X$ is $h'(a, x)$, defined as

$$\lim_{\lambda \downarrow 0} \lambda^{-1}[h(a + \lambda x) - h(a)].$$

Rockafellar [12], Section 23, shows that, *if h is convex*, then $\lambda^{-1}[h(a+\lambda x)-h(a)]$ is a nondecreasing function of $\lambda>0$, which therefore converges, as $\lambda\downarrow 0$, to a (possibly infinite) limit $h'(a,x)$, which satisfies $(\forall\lambda>0)$ $h'(a,\lambda x)=\lambda h'(a,x)$. For $X=\mathbf{R}^n$, Rockafellar shows that $h'(a,\cdot)$ is a proper closed convex function, defined for all $x\in X$; $\partial h(a)$ is a (weak *) closed convex subset of the dual space X', given by the intersection

$$\bigcap_{x\in X}\{v\in X':\ h'(a,x)\geqq vx\};$$

$\partial h(a)$ is a norm-bounded subset of X'; and

$$h'(a,x)=\sup_{v\in\partial h(a)} vx \tag{1}$$

When X_0 is an open subset of an infinite-dimensional normed space X, assume also that $h'(a,\cdot)$ is bounded above on some neighbourhood; then $h'(a,\cdot)$ is continuous, by Rockafellar [13], Theorem 8; then (1) follows from [13], Theorem 11.

If $h: X_0\to\mathbf{R}$ is strictly $\mathbf{R}_+$-convex, and $a\in X_0$, $x\in X_0$, $a\neq x$, then

$$(\forall k\in\partial h(a))\qquad h(x)-h(a)>k(x-a).$$

For the definition of strict $\mathbf{R}_+$-convexity gives, for $\lambda>0$, that

$$\begin{aligned} h(x)-h(a) &> \lambda^{-1}[h(a+\lambda(x-a))-h(a)]\\ &\geqq h'(a,x-a) \end{aligned}$$

since the function of λ is nondecreasing

$$\geqq k(x-a)\ \text{ for each } k\in\partial h(a), \text{ by (1).}$$

Given a convex cone $S\subset Y$, its *dual cone* is

$$S^*=\{y'\in Y':\ y'(S)\subset\mathbf{R}_+\}.$$

Note that if $\operatorname{int} S\neq\emptyset$, then $0\neq s^*\in S^*$ and $s\in\operatorname{int} S$ imply that $s^*s>0$.

Under some additional hypotheses, not assumed here, necessary conditions (of Fritz John type) for the minimization problem (MP) to attain a minimum at $x=a\in X_0$ are

$$(\mathbf{F}):\ \begin{cases} \sum_j k_j+\sum_i q_i=0; & \sum_i (s_i^*)\circ g(a)=0;\\ s_i^*\in S^*\quad (i=1,2,\cdots); & q_i\in\partial(s_i^*\circ g)(a);\\ p_j^*\in P^*\quad (j=1,2,\cdots); & k_j\in\partial(p_j^*\circ f)(a); \end{cases}$$

where Σ denotes finite sums, and at least one of the p_j^*, or at least one of the s_i^*, does not vanish.

The necessary condition **(F)** is proved in [9] under hypotheses weaker than convexity, and with a corresponding generalization of subdifferential. If, in particular, the functions f and g have linear Gâteaux derivatives $f'(a)$ and $g'(a)$, then $\partial f(a) = \{f'(a)\}$ and $\partial g(a) = \{g'(a)\}$, and **(F)** reduces to the Fritz John necessary condition

$$p^* \circ f'(a) + s^* \circ g'(a) = 0; \qquad s^* \circ g(a) = 0;$$
$$s^* \in S^*; \quad p^* \in P^*; \quad (s^*, p^*) \neq (0,0).$$

Under suitable hypotheses, **(F)** will now be shown to be also a sufficient condition for a minimum of (MP).

3. Sufficient Fritz John theorem

THEOREM. *Let X, W, and Y be normed spaces; let $P \subset W$, $S \subset Y$, $Q \subset Y$ be closed convex cones, with* $\operatorname{int} P$ *and* $\operatorname{int} Q$ *nonempty; let X_0 be an open subset of X; let the functions $f\colon X_0 \to W$ and $g\colon X_0 \to Y$ be resp. P-convex and strictly Q-convex. At $a \in X_0$, let $g'(a, \cdot)$ be continuous, and let the generalized Fritz John conditions* **(F)** *hold, and satisfy also $s_i^* \in Q^*$ for each i. Then the problem (MP) attains a minimum at $x = a$.*

REMARKS. If $\operatorname{int} S \neq \varnothing$, then it suffices to take $Q = S$. But $\operatorname{int} S$ may well be empty, for example if (MP) includes equality constraints. In this case, a convex cone Q is required, satisfying $Q \supset S$ and $\operatorname{int} Q \neq \varnothing$. Since then $Q^* \subset S^*$, the requirement $s_i^* \in Q^*$ is more stringent than $s_i^* \in S^*$. If it happens that $S \subset \operatorname{int} Q$, then g will be strictly Q-convex if it is S-convex.

PROOF. If $x = a$ is *not* a minimum for (MP), then there exists $x_0 \in X_0$ with $x_0 \neq a$, $-g(x_0) \in S$, and $f(x_0) - f(a) \in -\operatorname{int} P$. Set $z = x_0 - a$. Now $p_j^* \circ f$ is a convex function, since f is P-convex and $p_j^* \in P^*$; if some $p_j^* \neq 0$, then $k_j \in \partial(p_j^* \circ f)(a)$ with $0 \neq p_j^* \in P^*$; hence

$$k_j z \leqq p_j^* \circ f(x_0) - p_j^* \circ f(a) < 0$$

since $0 \neq p_j^* \in P^*$ and $f(x_0) - f(a) \in -\operatorname{int} P$. If some $s_i^* \neq 0$, then $q_i \in \partial(s_i^* \circ g)(a)$ with $0 \neq s_i^* \in Q^*$ by hypothesis; since g is strictly Q-convex,

$$-q_i z - s_i^* \circ g(a) = \{s_i^* \circ [g(x_0) - g(a)] - q_i z\} + s_i^* \circ (-g(x_0))$$
$$\in (\operatorname{int} \mathbf{R}_+) + \mathbf{R}_+ = \operatorname{int} \mathbf{R}_+,$$

noting that the function $s_i^* \circ g(\cdot)$ is strictly $\mathbf{R}_+$-convex. This, with the equation $(\Sigma s_i^*) \circ g(a) = 0$ of **(F)**, shows that $\Sigma_i q_i z < 0$, provided that some $s_i^* \neq 0$.

By the hypothesis of (**F**), there is some s_i^* or p_j^* nonzero; hence $\Sigma_i q_i z + \Sigma_j k_j z < 0$, which contradicts the equation $\Sigma_i q_i + \Sigma_j k_j = 0$ in (**F**). Hence $x = a$ is a minimum for (MP).

4. Applications

REMARK. If P and S are polyhedral cones, then so also are P^* and S^*. The variables p_j^* and s_i^* can then be restricted to the generators of their respective cones.

LEMMA. *Define the function* $g: \mathbf{R}^n \to \mathbf{R}$ *by* $g(x) = (x^TAx)^{1/2}$, *where A is a (symmetric) positive semidefinite $n \times n$ matrix. Then g is convex, and*

$$z \in \partial g(a) \quad \textit{iff} \quad \{z = w^TA,\ w^TAw \leqq 1,\ g(a) = w^TAa\}. \tag{2}$$

PROOF. If $g(a) \neq 0$, then g has derivative $g'(a) = a^TA/(a^TAa)^{1/2}$ at a. From the Schwarz inequality:

$$w^TAx \leqq |w^TAx| \leqq (x^TAx)^{1/2}(w^TAw)^{1/2}, \tag{3}$$

it follows that $g(x) - g(a) \geqq g'(a)(x - a)$, and therefore that g is convex on $\mathbf{R}^n \backslash N$, where $N = \{a: g(a) = 0\} = A^{-1}(0)$. Since g is continuous on $\mathbf{R}^n$, it then follows that g is convex on $\mathbf{R}^n$. Now

$$z \in \partial g(a) \quad \text{iff} \quad (\forall x)\ g(x) - g(a) \geqq z(x - a).$$

If $g(a) = 0$ and if A is positive definite, then (3) shows that $z \in \partial g(a)$ iff $z = w^TA$ where $w^TAw \leqq 1$; this extends to A positive semidefinite by diagonalizing A, then expressing it as the direct sum of a positive definite matrix and a zero matrix. Also $g(a) = 0 \Leftrightarrow Aa = 0 \Rightarrow g(a) = w^TAa$.

Suppose now that $g(a) \neq 0$. Then $\partial g(a) = \{g'(a)\}$, where $g'(a) = w^TA$, and $w = a/g(a)$ satisfies $w^TAw = 1$ and $w^TAa = g(a)$.

Conversely, suppose that $z = w^TA$, $w^TAw \leqq 1$, and $g(a) = w^TAa$. Then

$$z(x - a) = w^TAx - g(a) \leqq (w^TAw)^{1/2}(x^TAx)^{1/2} - g(a) \leqq g(x) - g(a),$$

and hence $z \in \partial g(a)$.

Consider now the minimization problem:

(MP'): $\quad$ Minimize $F(x) + (x^TB_{(\cdot)}x)^{1/2}$ subject to

$$-G(x) - (x^TD_{(\cdot)}x)^{1/2} \in \mathbf{R}_+^m,$$

where $F: \mathbf{R}^n \to \mathbf{R}^r$ and $G: \mathbf{R}^n \to \mathbf{R}^m$ are (linear Gâteaux) differentiable functions, B_j ($j = 1, 2, \cdots, r$) and D_i ($i = 1, 2, \cdots, m$) are positive semidefinite matrices, $x^TB_{(\cdot)}x$ denotes the vector in $\mathbf{R}^r$ whose jth component is x^TB_jx, and $x^TD_{(\cdot)}x$ is similarly defined. Now (MP') is of the form (MP), with $S = \mathbf{R}_+^m$ and taking $P = \mathbf{R}^r_+$. For this P, minimization at a, namely $f(x) - f(a) \notin -\operatorname{int} P$, is equivalent to $(\exists j)\ f_j(x) - f_j(a) \geqq 0$, where f_j is the jth component of $f(x) = F(x) + (x^T B_{()} x)^{1/2}$.

For (MP′), the Fritz John conditions (**F**) take the form

$$(\mathbf{F}'):\ \begin{cases} \sum_{j=1}^{r} \tau_j [F_j'(a) + v_j^T B_j] + \sum_{i=1}^{m} \lambda_i [G_i'(a) + w_i^T D_i] = 0; \\ \lambda_i \geqq 0, \quad \lambda_i [G_i(a) + (a^T D_i a)^{1/2}] = 0, \\ w_i^T D_i w_i \leqq 1, \quad w_i^T D_i a = (a^T D_i a)^{1/2}, \quad (i = 1, 2, \cdots, m); \\ \tau_j \geqq 0, \quad v_j^T B_j v_j \leqq 1, \quad v_j^T B_j a = (a^T B_j a)^{1/2}, \quad (j = 1, 2, \cdots, r). \end{cases}$$

In (**F**′), at least one of the λ_i or τ_j is nonzero. Here the characterization, in the above lemma, of the subgradient of a square root function has been used.

Assume now that F is $\mathbf{R}^r_+$-convex (thus each component F_j is convex); that G is $\mathbf{R}^m_+$-convex; and that, for each $i = 1, 2, \cdots, m$, G_i is strictly convex or D_i is positive definite. Then $g(x) = G(x) + (x^T D_{()} x)^{1/2}$ is a strictly $\mathbf{R}^m_+$-convex function of x. The sufficient Fritz John theorem then shows that these hypotheses, with (**F**′), imply that (MP′) attains a minimum at $x = a$.

The cone $\mathbf{R}^m_+$ can be replaced by a polyhedral cone S, and an analogous result obtained. Here the s_i^* become the generators of S^*. In order that the function $x \mapsto s_i^*(x^T D_{()} x)^{1/2}$ is convex, it is necessary, and sufficient, that $S^* \subset \mathbf{R}^m_+$, hence that $S \supset \mathbf{R}^m_+$.

Similar conclusions hold with $x^T B_j x$ replaced by $(x^T - b_j^T) B_j (x - b_j)$, for constant vectors b_j.

Acknowledgement

The authors are indebted to a referee for a shorter proof of the converse part of the Lemma in section 4, and for several minor corrections.

References

[1] C. R. Bector, Suresh Chandra, and T. R Gulati, Complex nonlinear programming with equality constraints, *Proc. Fourth Manitoba Conference on Numerical Mathematics* (1974), 205–216.

[2] C. R. Bector and T. R. Gulati, Sufficient optimality conditions in presence of quasiconvex equality and inequality constraints, *Zeitschrift für Mathematische Operationsforschung und Statistik*, to appear.

[3] J. M. Borwein, Optimization with respect to partial orderings, D. Phil. thesis, Oxford, 1974.

[4] B. D. Craven, Nonlinear programming in locally convex spaces, *J. Optim. Theor. Appl.* 10 (1972), 197–210.

[5] B. D. Craven, Sufficient Fritz John optimality conditions, *Bull. Austral. Math. Soc.* 13 (1975), 411–419.

[6] B. D. Craven, Lagrangean conditions for optimization in Banach spaces, University of Melbourne, School of Mathematical Sciences, Research Report No. 6 (1976)

[7] B. D. Craven and B. Mond, Transposition theorems for cone-convex functions, *S.I.A.M. J. Appl. Math.* 24 (1973), 603–612.

[8] B. D. Craven and B. Mond, A class of nondifferentiable complex programming problems, *Mathematische Operationsforschung und Statistik* 6 (1975) 581–591.

[9] B. D. Craven and B. Mond, Lagrangean conditions for quasidifferentiable optimization, University of Melbourne, School of Mathematical Sciences, Research Report No. 7 (1976).

[10] T. R. Gulati, A Fritz John type sufficient optimality theorem in complex space, *Bull. Austral. Math. Soc.* 11 (1974), 219–224.

[11] T. R. Gulati, Optimality criteria and duality in complex mathematical programming, Ph.D. thesis, Indian Institute of Technology, Hauz-Khas, New Delhi (1975).

[12] R. T. Rockafellar, *Convex analysis*, Princeton University Press, Princeton, New Jersey, 1970.

[13] R. T. Rockafellar, Conjugate duality and optimization, Soc. Ind. Appl. Math., Philadelphia, 1974. (Vol. 16 of Regional Conference Series in Applied Mathematics.)

Department of Mathematics,
University of Melbourne,
Parkville 3052,
Australia.

Department of Mathematics,
La Trobe University,
Bundoora 3083,
Australia.