BULL. AUSTRAL. MATH. SOC.
VOL. 13 (1975), 411-419.

90C25, 90C30

Sufficient Fritz John optimality conditions

B.D. Craven

The sufficient optimality conditions, of Fritz John type, given by Gulati for finite-dimensional nonlinear programming problems involving polyhedral cones, are extended to problems with arbitrary cones and spaces of arbitrary dimension, whether real or complex. Convexity restrictions on the function giving the equality constraint can be avoided by applying a modified notion of convexity to the other functions in the problem. This approach regards the problem as optimizing on a differentiable manifold, and transforms the problem to a locally equivalent one where the optimization is on a linear subspace.

1. Introduction

Gulati [9] has given sufficient optimality conditions of Fritz John type for differentiable nonlinear programming problems in finite-dimensional complex spaces. Both equality constraints, and inequality constraints involving polyhedral cones, were included. Gulati obtained sufficient conditions for optimality by adding appropriate convexity hypotheses to the Fritz John necessary conditions for optimality, given by Craven and Mond [7, 8].

Gulati's sufficient conditions are now extended to programming problems involving arbitrary cones, not necessarily polyhedral, and spaces of any dimension. These results, and Gulati's, place convexity requirements on the function specifying the equality constraint, and these may not always be fulfilled in a given problem. An alternative approach is

Received 2 September 1975.

therefore also given, in which there are no convexity restrictions on the function specifying the equality constraint, but a modified notion of convexity is applied to the other functions in the problem. This approach regards the problem as one of optimization on a differentiable manifold, and transforms the problem to a locally equivalent one in which the optimization is on a linear subspace. The quantities involved can be calculated explicitly, if the functions have second (Fréchet) derivatives.

2. Preliminaries

Let X, Y, Z, W be real or complex Banach spaces; U an open subset of X; $S \subset Y$, $T \subset Z$, and $N \subset W$ closed convex cones. Let the maps $f : U \to \mathbf{R}$ (or $f : U \to \mathbf{C}$ in the complex case), $g : U \to Y$, and $h : U \to Z$ possess linear Gâteaux derivatives (which are then continuous linear maps $f'(x) : X \to \mathbf{R}$, and so on). Define a continuous affine map $k : X \to W$ by $(\forall\, x \in X)k(x) = d + Dx$, where $d \in W$ and $D : X \to W$ is a continuous linear map. The dual cone of S is

$$S^* = \{u \in Y' : (\forall\, s \in S) \text{ re } u(s) \in \mathbf{R}_+\} ,$$

where $\mathbf{R}_+ = [0, \infty)$, and Y' is the topological dual space of Y. The transpose (or adjoint) of the continuous linear map $h'(x_0) : X \to Z$ is the continuous linear map $h'(x_0)^T : Z' \to X'$, defined by

$$z'(h'(x_0)x) = \left(h'(x_0)^T z'\right)(x)$$

for all $x \in X$, $z' \in Z'$. The topological interior of S is denoted int S; note that $(\text{int } S) + S \subset \text{int } S$, provided that int $S \neq \emptyset$.

The map $f : U \to \mathbf{R}$ is *pseudoconvex* at $a \in U$ if

$$[x \in U \text{ and } f(x) < f(a)] \Rightarrow [f'(a)(x-a) < 0] .$$

The map $g : U \to Y$ is *S-convex* (*strictly S-convex*) in U at $a \in U$ if, for each $x \in U$,

$$g(x) - g(a) - g'(a)(x-a) \in S \ (\in \text{int } S) .$$

Use will be made of the following generalization of Motzkin's alternative theorem, proved in [4] for locally convex spaces.

THEOREM 0. *Let J, K, L be real or complex Banach spaces, $H \subset K$*

https://doi.org/10.1017/S0004972700024655 Published online by Cambridge University Press

and $Q \subset L$ *closed convex cones, with* int Q *nonempty; and* $A : J \to K$ *and* $B : J \to L$ *continuous linear maps. Assume either that the cone* $A^T(H^*)$ *is weak* * *closed in* J', *or that* $A(J) = K$. *Then exactly one of the two following linear systems has a solution:*

(I) $-Ax \in H$, $-Bx \in \text{int } Q$;

(II) $(\forall\, x \in J)\ \text{re}[w(Ax)+u(Bx)] = 0$, $w \in H^*$, $0 \neq u \in Q^*$.

Define $w^*A \in K'$ by $(\forall\, x \in J)\ (w^*A)(x) = \text{re}\, w(Ax)$, and similarly for u^*B. Then (II) may be written equivalently as:

$$w^*A + u^*B = 0\ ,\quad w \in H^*\ ,\quad 0 \neq u \in Q^*\ .$$

If $X = \mathbb{C}^n$, and $\bar{z}$ denotes the complex conjugate of $z \in \mathbb{C}^n$, set $M = \{(z_1, z_2) \in X \times X : z_2 = \bar{z}_1\}$. Although M is not a complex vector space, it is shown in [4] that Theorem 0 holds with $J = M$, if the maps A and B are linear (with respect to $\mathbb{R}$). If, instead, X is any complex Banach space, then with a suitable definition of conjugate (see [4]), Theorem 0 continues to hold, assuming that the maps A and B are continuous, and linear with respect to $\mathbb{R}$.

Consider the minimization problem:

(P) $$\underset{x \in U}{\text{minimize}}\ \{\text{re}\, f(x) : -g(x) \in S,\ h(x) = 0,\ -k(x) \in N\}\ ,$$

where U, f, g, h, k, S, T, N are as described in the first paragraph. Associated to (P) is the linear system:

(F) $$r^*f'(a) + v^*g'(a) + w^*h'(a) + m^*k'(a) = 0\ ;$$

$$\text{re}\, vg(a) = 0\ ;$$

$$\text{re}\, mk(a) = 0\ ;$$

$$r \in \mathbb{R}_+\ ,\quad v \in S^*\ ,\quad w \in T^*\ ,\quad m \in N^*\ ;$$

where $a \in U$. Note that, if (P) attains a local minimum at $x = a \in U$, and if certain other conditions are fulfilled, then (F) gives the Fritz John necessary condition for the minimum [6], in the case where $T = \{0\}$.

In the complex case, if $X = M$, then (P) and (F) become (see [7], [5])

https://doi.org/10.1017/S0004972700024655 Published online by Cambridge University Press

(PM) $\underset{z\in M}{\text{minimize}}\ \{\text{re } f(z, \bar{z}) : -g(z, \bar{z}) \in S,\ h(z, \bar{z}) = 0,\ -k(z, \bar{z}) \in N\}$,

where Y, Z, W are complex Banach spaces, and k is affine with respect to R ; and

$$(\text{FM}) \qquad r\left(f_z+\overline{f_{\bar{z}}}\right) + \left(\bar{v}g_z+v\overline{g_{\bar{z}}}\right) + \left(\bar{w}h_z+w\overline{h_{\bar{z}}}\right) + \left(\bar{m}k_z+m\overline{k_{\bar{z}}}\right) = 0 ;$$

$$\text{re } vg(b, \bar{b}) = 0 ;$$

$$\text{re } mk(b, \bar{b}) = 0 ;$$

$$r \in R_+ , \quad v \in S^* , \quad w \in T^* , \quad m \in N^* ;$$

in which the partial derivatives f_z , and so on, are evaluated at $(z, \bar{z}) = (b, \bar{b}) \in M$. The notational conventions of [7] are followed here. Note that, if the dimensions are finite, and the constraint $-k(z, \bar{z}) \in N$ is omitted, then (FM) agrees with equations (1) and (2) of Gulati [9], in different notation.

3. Sufficient F. John conditions

THEOREM 1. *At $a \in U$, let* re f *be pseudoconvex, let g be strictly S-convex, let h be strictly T-convex, let* int $S \neq \emptyset$, *and let* int $T \neq \emptyset$. *Assume either that the map $C = [k'(a)\ \ k(a)]$ is surjective, or that the cone $C^T(N^*)$ is weak * closed in $X' \times C$. Then, if* (F) *has a solution r, v, w, m with r, v, w not all zero, it follows that* (P) *attains a local minimum at $x = a$.*

REMARK. This result applies both to real spaces, and to complex spaces. For complex spaces, alternative expressions for the derivatives involved will be given in Theorem 2.

Proof. If $x = a$ is *not* a minimum for (P), then there is a solution $x = x_0 \in U$ to the system

$$\text{re}[f(x)-f(a)] < 0 , \quad -g(x) \in S , \quad -h(x) = 0 , \quad -k(x) \in N .$$

Set $p = x_0 - a$. Since re f is pseudoconvex at a , $-\text{re}[f'(a)p] \in \text{int } R_+$. Since g is strictly S-convex at a ,

$$-[g'(a)p+g(a)] = \left[g(x_0)-g(a)-g'(a)p\right] + \left[-g(x_0)\right] \in (\text{int } S) + S \subset \text{int } S .$$

A similar calculation shows that $-[h'(a)p+h(a)] \in \text{int } T$, since h is

https://doi.org/10.1017/S0004972700024655 Published online by Cambridge University Press

strictly T-convex. Since k is affine,

$$-[k'(a)p+k(a)] = -k(x_0) \in N .$$

These results combine to show that there is a solution ζ to the system

$$(+) \qquad -A\zeta \in N , \quad -B\zeta \in \operatorname{int} V ,$$

where

$$A = [k'(a) \quad 0 \quad k(a)] , \quad B = \begin{bmatrix} f'(a) & 0 & 0 \\ g'(a) & g(a) & 0 \\ h'(a) & h(a) & 0 \end{bmatrix} ,$$

V is the convex cone $\mathbf{R}_+ \times S \times N$, and

$$\zeta = \begin{bmatrix} p \\ 1 \\ 1 \end{bmatrix} \in X \times \mathbf{R} \times \mathbf{R} .$$

From the properties assumed for C, either A is surjective, or the cone $A^T(N^*)$ is weak * closed. Therefore Theorem 0 applies to the system (+), showing that there is no solution $m \in N^*$, $[r \quad v \quad w] \in V^* \setminus \{0\}$ to the system

$$m^*A + [r \quad v \quad w] * B = 0 .$$

But this system is exactly (F), which by hypothesis has a solution with r, v, w not all zero.

THEOREM 2. *At $(b, \bar{b}) \in M$, let $f : M \to Y$ have pseudoconvex real part, let $g : M \to Y$ be strictly S-convex, let $h : M \to Z$ be strictly T-convex, let* int $S \neq \emptyset$, *and let* int $T \neq \emptyset$. *Assume either that the map*

$$K = \begin{bmatrix} k_z + k_{\bar{z}} & k(b, \bar{b}) \\ \overline{k_z + k_{\bar{z}}} & \overline{k(b, \bar{b})} \end{bmatrix}$$

is surjective (onto W), or that the cone $K^T(N^*)$ is weak * closed. Then, if (FM) *has a solution r, v, w, m with r, v, w not all zero, it follows that* (PM) *attains a local minimum at $(z, \bar{z}) = (b, \bar{b})$.*

REMARK. The cone $K^T(N^*)$ is to be weak * closed in $M' \times \mathbf{C}$, where M' is the space of continuous functionals on M, linear with respect to $\mathbf{R}$.

https://doi.org/10.1017/S0004972700024655 Published online by Cambridge University Press

Proof. The result follows from Theorem 1, provided that C can be identified with K. And this follows from the relations, given in [5], pages 619-622, relating derivatives with respect to z and $\bar{z}$ to derivatives with respect to real and imaginary components. Set

$$k(z, z) = \tilde{k}(x, y) \text{ , where } z = x + iy \text{ , and } \tilde{k} = \tilde{k}^r + i\tilde{k}^i .$$

If (PM) is expressed as an equivalent problem in real spaces, then m^*C becomes (if $m = m^r + im^i$)

$$\operatorname{re}\begin{bmatrix} m^r & m^i \end{bmatrix}\begin{bmatrix} \tilde{k}^r_x & \tilde{k}^r_y & \vdots & \tilde{k}^r \\ \tilde{k}^i_x & \tilde{k}^i_y & \vdots & \tilde{k}^i \end{bmatrix} = \tfrac{1}{2}\begin{bmatrix} \bar{m} & m \end{bmatrix}\begin{bmatrix} k_z + k_{\bar{z}} & \vdots & k \\ \bar{k}_z + \bar{k}_{\bar{z}} & \vdots & \bar{k} \end{bmatrix}$$

after calculation. The factor $\frac{1}{2}$ does not affect the hypothesis of the theorem.

REMARKS. The first component of m^*C equals $\operatorname{re}\left[\bar{m}k_z + m\bar{k}_{\bar{z}}\right]$, which is the real part of the corresponding term in (FM).

The cones need *not* be polyhedral, as was assumed by Gulati. However, the hypothesis that $C^T(N^*)$, or $K^T(N^*)$, is closed holds automatically in case N is polyhedral. The results apply equally to finite and to infinite dimensional spaces. In Theorems 1 and 2, the constraint $-g(x) \in S$ cannot include an affine constraint, since g must be strictly S-convex. However, affine constraints may be included in $-k(x) \in N$, where no such convex hypothesis is required.

THEOREM 3. *The conclusion of Theorem* 1 *remains valid, omitting the hypothesis that* int $T \neq \emptyset$, *and requiring that* h *is* T*-convex (rather than strictly* T*-convex), provided that* (F) *has a solution with* r *and* v *not both zero, and the hypothesis on* C *is replaced by the hypotheses that either*

$$E = \begin{bmatrix} k'(a) & k(a) \\ h'(a) & 0 \end{bmatrix}$$

is surjective, or the cone $E^T(N^* \times T^*)$ *is weak* * *closed.*

Proof. In the proof of Theorem 1, replace A and B respectively by

https://doi.org/10.1017/S0004972700024655 Published online by Cambridge University Press

$$\begin{bmatrix} k'(a) & 0 & k(a) \\ h'(a) & 0 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} f'(a) & 0 & 0 \\ g'(a) & g(a) & 0 \end{bmatrix},$$

and the corresponding cones by $N \times T$ respectively $R_+ \times S$.

REMARKS. In Theorem 3, T could be $\{0\}$; but then convexity would require h to be affine, and so includable in k . In the next section, a result is given, for real spaces, which avoids any convexity hypothesis for the equality constraint.

4. Nonconvex equality constraints

In this section, no convexity hypothesis is made for h , and the spaces considered are real. Assume now that $h : U \to Z$ is continuously Fréchet-differentiable, $h(a) = 0$, and that $C = h'(a)$ is such that $C(X) = Z$, and there is a continuous projection q of X onto the kernel $C^{-1}(0)$. It then follows [2, 3] that there is a homeomorphism φ of a neighbourhood Q of zero in the linear subspace $C^{-1}(0)$ onto $U_0 \cap \{x : h(x) = 0\}$, where $U_0 \subset U$ is a neighbourhood of $a \in X$; φ and φ^{-1} are Fréchet-differentiable at 0 and a respectively; and $\varphi'(0)(\zeta) = \zeta \in X$ for each $\zeta \in C^{-1}(0)$.

Given such a φ , the restrictions to Q of the vector space operations on $C^{-1}(0)$ may be transferred to $\varphi(Q) = U_0 \cap \{x : h(x) = 0\}$ by defining [3], for all $x, y \in \varphi(Q)$ and scalars α ,

$$x \oplus y = \varphi\left(\varphi^{-1}x + \varphi^{-1}y\right) \quad \text{and} \quad \alpha \odot x = \varphi\left(\alpha\varphi^{-1}x\right) .$$

The map $g : U \to Y$ is then called h, S-convex (strictly h, S-convex) on $\varphi(Q)$ if for all $x, y \in \varphi(Q)$ and all $\tau \in [0, 1]$,

$$(1-\tau)g(x) + \tau g(y) - g\left((1-\tau) \odot x \oplus \tau \cdot y\right) \in S \ (\in \operatorname{int} S) .$$

Thus g is h, S-convex (strictly h, S-convex) iff $g \circ \varphi$ is S-convex (strictly S-convex) at each point of Q . If g and h are twice continuously Fréchet-differentiable, then $g \circ \varphi$ is S-convex (strictly S-convex) at 0 in a neighbourhood of 0 if

$$0 \neq w \in Q \Rightarrow (g \circ \varphi)''(0)[w]^2 \in S\backslash\{0\} \ (\in \operatorname{int} S) .$$

https://doi.org/10.1017/S0004972700024655 Published online by Cambridge University Press

It is shown in [3] that, under the same hypotheses,

$$(g \circ \varphi)''(0) = g''(a) - g'(a) \circ [h'(a) \circ (1-q)]^{-1} \circ h''(a) .$$

Under these hypotheses, the problem (P), with the constraint $-k(x) \in N$ omitted, and with x restricted to a suitable neighbourhood of $x = a$, is equivalent to:

$$(EP) \qquad \underset{\zeta \in Q}{\text{minimize}} \{(f \circ \varphi)(\zeta) : -(g \circ \varphi)(\zeta) \in S\} .$$

THEOREM 4. *For problem* (P) *in real spaces, at* $a \in U$, *let* f *be pseudoconvex, let* h *satisfy the above hypotheses (so that* φ *exists), let* g *be strictly* h, S*-convex, and let* int $S \neq \emptyset$. *(The constraint* $-k(x) \in N$ *is omitted.) Then, if* (F) *has a solution* r, v, w, *not all zero, it follows that* (P) *attains a local minimum at* $x = a$.

Proof. From (F),

$$\psi \equiv r(f \circ \varphi)'(0) + v(g \circ \varphi)'(0) = [rf'(a)+vg'(a)] \circ \varphi'(0)$$
$$= -wh'(a) \circ \varphi'(0) .$$

Since $\varphi'(0)$ is the identity map of $h'(a)^{-1}(0)$ onto $h'(a)^{-1}(0) \subset X$, it follows that the restriction of ψ to domain $h'(a)^{-1}(0)$ is zero.

Since g is strictly h, S-convex, $g \circ \varphi$ is strictly S-convex. Since f is pseudoconvex, and $\varphi'(0)$ is the identity,

$$[x \in U_0,\ h(x) = 0,\ f(x) < f(a)] \Rightarrow [f'(a)(x-a) < 0 \text{ and } x-a = \zeta+o(\|\zeta\|)]$$

for some $\zeta \in Q \subset h'(a)^{-1}(0)$, where $o(\|\zeta\|)/\|\zeta\| \to 0$ as $\|x-a\| \to 0$. Hence

$$(f \circ \varphi)(\zeta) = f'(a) \circ \varphi'(0)(\zeta) < 0$$

for sufficiently small $\|\zeta\|$; so $f \circ \varphi$ is pseudoconvex, in a neighbourhood of 0.

The hypotheses of Theorem 1 are therefore satisfied for the problem (EP), for ζ in a neighbourhood of 0. Consequently (EP), and therefore (P), attains a local minimum (at 0 respectively a).

REMARKS. Sufficient conditions for g to be h, S-convex are given above, in terms of first and second derivatives of g and h. There is no need to actually calculate φ.

https://doi.org/10.1017/S0004972700024655 Published online by Cambridge University Press

The hypotheses on h, stated at the beginning of this section, are sufficient but not necessary. In order to construct φ, it is sufficient that, to each ζ satisfying $h'(a)\zeta = 0$, there is a solution $x = a + \lambda\zeta + o(\lambda)$ to $h(x) = 0$, valid for all sufficiently small positive λ, where

$$(\forall\ \varepsilon > 0)\ (\exists\ \delta > 0)\ [\|\zeta\| \leq 1 \text{ and } 0 < \lambda < \delta] \Rightarrow \|o(\lambda)\| < \varepsilon\lambda .$$

References

[1] Robert A. Abrams, "Nonlinear programming in complex space: sufficient conditions and duality", *J. Math. Anal. Appl.* 38 (1972), 619-632.

[2] B.D. Craven, "A generalization of Lagrange multipliers", *Bull. Austral. Math. Soc.* 3 (1970), 353-362.

[3] B.D. Craven, "Converse duality in Banach space", *J. Optimization Theory Appl.* 17 (1975), 229-238.

[4] B.D. Craven, "A generalized Motzkin alternative theorem" (Research Report, 13. School of Mathematical Sciences, University of Melbourne, Parkville, 1975).

[5] B.D. Craven and B. Mond, "Converse and symmetric duality in complex nonlinear programming", *J. Math. Anal. Appl.* 37 (1972), 617-626.

[6] B.D. Craven and B. Mond, "Transposition theorems for cone-convex functions", *SIAM J. Appl. Math.* 24 (1973), 603-612.

[7] B.D. Craven and B. Mond, "Real and complex Fritz John theorems", *J. Math. Anal. Appl.* 44 (1973), 773-778.

[8] B.D. Craven and B. Mond, "A Fritz John theorem in complex space", *Bull. Austral. Math. Soc.* 8 (1973), 215-220.

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

Department of Mathematics,
University of Melbourne,
Parkville, Victoria.