Available online at www.sciencedirect.com
SCIENCE DIRECT
European Journal of Operational Research xxx (2006) xxx–xxx
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/ejor

Short Communication

An elementary proof of the Fritz-John and Karush–Kuhn–Tucker conditions in nonlinear programming

Ş.İ. Birbil [a], J.B.G. Frenk [b,*], G.J. Still [c]

[a] Faculty of Engineering and Natural Sciences, Sabancı University, Orhanli-Tuzla, 34956 Istanbul, Turkey
[b] Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
[c] Department of Mathematical Sciences, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

Received 29 October 2005; accepted 12 April 2006

Abstract

In this note we give an elementary proof of the Fritz-John and Karush–Kuhn–Tucker conditions for nonlinear finite dimensional programming problems with equality and/or inequality constraints. The proof avoids the implicit function theorem usually applied when dealing with equality constraints and uses a generalization of Farkas lemma and the Bolzano-Weierstrass property for compact sets.

Keywords: Nonlinear programming; Fritz-John conditions; Karush-Kuhn-Tucker conditions

1. Introduction

Let A be an $m \times n$ matrix with rows $\mathbf{a}_k^\top$, $1 \leqslant k \leqslant m$, $\mathbf{b} \in \mathbb{R}^m$ an m-dimensional vector, and $f_i : \mathbb{R}^n \to \mathbb{R}$, $0 \leqslant i \leqslant q$ some non-affine, continuously differentiable functions. We consider the optimization problem

$$\min\{f_0(\mathbf{x}) : \mathbf{x} \in \mathscr{F}_P\},\ \mathscr{F}_P := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}_k^\top \mathbf{x} \leqslant b_k, 1 \leqslant k \leqslant m,\ f_i(\mathbf{x}) \leqslant 0, 1 \leqslant i \leqslant q\}, \tag{P}$$

and the program including equalities

$$\min\{f_0(\mathbf{x}) : \mathbf{x} \in \mathscr{F}_Q\},\ \mathscr{F}_Q := \mathscr{F}_P \cap \{\mathbf{x} \in \mathbb{R}^n : h_j(\mathbf{x}) = 0, 1 \leqslant j \leqslant r\}, \tag{Q}$$

where the functions $h_j : \mathbb{R}^n \to \mathbb{R}$, $1 \leqslant j \leqslant r$, are non-affine and continuously differentiable.

Two basic results covered in every course on nonlinear programming are the Fritz-John (FJ) and Karush–Kuhn–Tucker (KKT) necessary conditions for the local minimizers of optimization problems (P) and (Q) [7–9]. Denoting the nonnegative orthant of $\mathbb{R}^l$ by $\mathbb{R}^l_+$, the FJ necessary conditions for problem (P)

* Corresponding author. Tel.: +31 10 4081257; fax: +31 10 4527746.
E-mail addresses: sibirbil@sabanciuniv.edu (Ş.İ. Birbil), frenk@few.eur.nl (J.B.G. Frenk), g.still@math.twente.nl (G.J. Still).

0377-2217/$ - see front matter © 2006 Published by Elsevier B.V.
doi:10.1016/j.ejor.2006.04.012

are given by the following: If $\mathbf{x}_P$ is a local minimizer of problem (P), then there exist (see for example [2,5]) vectors $\mathbf{0} \neq \lambda \in \mathbb{R}_+^{q+1}$ and $v \in \mathbb{R}_+^m$ satisfying

$$\begin{aligned}&\sum_{i=0}^{q} \lambda_i \nabla f_i(\mathbf{x}_P) + \sum_{k=1}^{m} v_k \mathbf{a}_k = \mathbf{0},\\ &\lambda_i f_i(\mathbf{x}_P) = 0,\ 1 \leqslant i \leqslant q \quad \text{and } v_k(\mathbf{a}_k^\top \mathbf{x}_P - b_k) = 0,\ 1 \leqslant k \leqslant m.\end{aligned} \tag{FJP}$$

For optimization problem (Q) the resulting FJ conditions are as follows: If $\mathbf{x}_Q$ is a local minimizer of problem (Q), then there exist (see for example [2,5]) vectors $(\lambda, v) \in \mathbb{R}_+^{q+1+m}$, $\mu \in \mathbb{R}^r$ with $(\lambda, \mu) \neq \mathbf{0}$ satisfying

$$\begin{aligned}&\sum_{i=0}^{q} \lambda_i \nabla f_i(\mathbf{x}_Q) + \sum_{j=1}^{r} \mu_j \nabla h_j(\mathbf{x}_Q) + \sum_{k=1}^{m} v_k \mathbf{a}_k = \mathbf{0},\\ &\lambda_i f_i(\mathbf{x}_Q) = 0,\ 1 \leqslant i \leqslant q \quad \text{and} \quad v_k(\mathbf{a}_k^\top \mathbf{x}_Q - b_k) = 0,\ 1 \leqslant k \leqslant m.\end{aligned} \tag{FJQ}$$

If λ_0 given in conditions (FJP) and (FJQ) can be chosen positive, then the resulting necessary conditions are called the KKT conditions for problems (P) and (Q), respectively. A sufficient condition for λ_0 to be positive is given by a so-called first-order constraint qualification. In Section 2 we first give an elementary proof of the FJ and KKT conditions for problem (P). Then the same proof is given for optimization problem (Q) by using a perturbation argument but avoiding the implicit function theorem.

2. The FJ and KKT conditions for problems (P) and (Q)

For $\delta > 0$ and $\bar{\mathbf{x}} \in \mathbb{R}^n$, let $\mathcal{N}(\bar{\mathbf{x}}, \delta)$ denote a δ-neighborhood of $\bar{\mathbf{x}}$ given by

$$\mathcal{N}(\bar{\mathbf{x}}, \delta) := \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \bar{\mathbf{x}}\| \leqslant \delta\}.$$

A vector $\mathbf{x}_P$ is called a local minimizer of optimization problem (P) (respectively, for optimization problem (Q) if $\mathbf{x}_P \in \mathcal{F}_P$ (respectively, $\mathbf{x}_P \in \mathcal{F}_Q$) and there exists some $\delta > 0$ such that $f_0(\mathbf{x}_P) \leqslant f_0(\mathbf{x})$ for every $\mathbf{x} \in \mathcal{F}_P \cap \mathcal{N}(\mathbf{x}_P, \delta)$ (respectively, $\mathbf{x} \in \mathcal{F}_Q \cap \mathcal{N}(\mathbf{x}_P, \delta)$).

We introduce the active index sets $I(\mathbf{x}) := \{1 \leqslant i \leqslant q : f_i(\mathbf{x}) = 0\}$ and $K(\mathbf{x}) = \{1 \leqslant k \leqslant m : \mathbf{a}_k^\top \mathbf{x} = b_k\}$, and denote by $B(\mathbf{x})$, the matrix consisting of the corresponding active rows $\mathbf{a}_k^\top, k \in K(\mathbf{x})$.

Lemma 2.1. *If $\mathbf{x}_P$ is a local minimizer of problem (P), then* $\max\{\nabla f_i(\mathbf{x}_P)^\top \mathbf{d} : i \in I(\mathbf{x}_P) \cup \{0\}\} \geqslant 0$ *for every* $\mathbf{d}$ *such that* $B(\mathbf{x}_P)\mathbf{d} \leqslant \mathbf{0}$.

Proof. Suppose by contradiction there exists some $\mathbf{d}_0$ satisfying $B(\mathbf{x}_P)\mathbf{d}_0 \leqslant \mathbf{0}$ and

$$0 > \nabla f_i(\mathbf{x}_P)^\top \mathbf{d}_0 = \lim_{t \downarrow 0} \frac{f_i(\mathbf{x}_P + t\mathbf{d}_0) - f_i(\mathbf{x}_P)}{t}$$

for every $i \in I(\mathbf{x}_P) \cup \{0\}$. By the finiteness of the sets $\{0, \ldots, q\}$ and $\{1, \ldots, m\}$ and the continuity of f_i this implies the existence of some $t_0 > 0$ satisfying

$$f_i(\mathbf{x}_P + t\mathbf{d}_0) < 0, i \notin I(\mathbf{x}_P), f_i(\mathbf{x}_P + t\mathbf{d}_0) < f_i(\mathbf{x}_P), i \in I(\mathbf{x}_P) \cup \{0\}, A(\mathbf{x}_P + t\mathbf{d}_0) \leqslant \mathbf{b}$$

for every $0 < t \leqslant t_0$. Hence, the vector $\mathbf{x}_P + t\mathbf{d}_0$ belongs to $\mathcal{F}_P$ and satisfies $f_0(\mathbf{x}_P + t\ \mathbf{d}_0) < f_0(\mathbf{x}_P)$ for every $0 < t \leqslant t_0$. This contradicts that $\mathbf{x}_P$ is a local minimum. □

Remark 2.1. If the function f_0 is pseudo-convex and the functions f_i, $1 \leqslant i \leqslant q$ are strictly pseudo-convex, then for a feasible $\mathbf{x}_P$ the reverse implication in Lemma 2.1 also holds and in this result local minimizer is replaced by global minimizer. A proof of this will be given at the end of this section. Moreover, if $\max_{B\mathbf{d} \leqslant 0, \|d\|=1}\{\nabla f_i(\mathbf{x}_P)^\top \mathbf{d} : i \in I(\mathbf{x}_P) \cup \{0\}\} > 0$ and $\mathbf{x}_P$ feasible, then one can show that $\mathbf{x}_P$ is a local minimum of order one [5], i.e., there exists some $\delta > 0$ and $c > 0$ such that $f_0(\mathbf{x}) - f_0(\mathbf{x}_P) \geqslant c\|\mathbf{x} - \mathbf{x}_P\|$ for every $\mathbf{x} \in \mathcal{F}_P \cap \mathcal{N}(\mathbf{x}_P, \delta)$.

The proof of the *FJ* conditions for problem (P) will be based on the following generalization of Farkas lemma [6]. For completeness, a short proof, using the strong duality result for linear programming, will be given in Appendix A.

Lemma 2.2. *Let* $\Delta_s \subseteq \mathbb{R}^s_+$ *be the unit simplex. If* B *is a* $p \times n$ *matrix and* $\mathbf{c}_i \in \mathbb{R}^n, 1 \leqslant i \leqslant s$, *some given vectors, then the following conditions are equivalent*:

1. *For every* $\mathbf{d} \in \mathbb{R}^n$ *satisfying* $B\mathbf{d} \leqslant \mathbf{0}$ *it holds that* $\max_{1\leqslant i\leqslant s}\mathbf{c}_i^\top\mathbf{d} \geqslant 0$.
2. *There exists some* $\lambda \in \Delta_s$ *and* $\mu \in \mathbb{R}^p_+$ *satisfying* $\sum_{i=1}^s \lambda_i \mathbf{c}_i + B^\top\mu = \mathbf{0}$.

Proof (*FJ conditions for problem* (*P*)). By combining Lemmas 2.1 and 2.2, the FJ conditions follow. □

It is well-known that the KKT conditions follow from the FJ conditions under some constraint qualification. We say that the *Mangasarian-Fromovitz* (MF) constraint qualification for problem (P) holds at a feasible point $\mathbf{x}$ if there exists some $\mathbf{d}_0$ satisfying

$$B(\mathbf{x})\mathbf{d}_0 \leqslant 0 \quad \text{and} \quad \max_{i\in I(\mathbf{x})}\{\nabla f_i(\mathbf{x})^\top \mathbf{d}_0\} < 0.$$

We now show that at a local minimizer $\mathbf{x}_P$ of problem (P) satisfying the MF constraint qualification, the KKT conditions must hold.

Proof (*KKT conditions for problem* (*P*)). Assume that $\lambda_0 = 0$ in the FJ conditions. Applying Lemma 2.2 to the FJ conditions with $\lambda_0 = 0$ we obtain that $\max_{i\in I(\mathbf{x}_P)}\nabla f_i(\mathbf{x}_P)^\top\mathbf{d} \geqslant 0$ for every $B(\mathbf{x}_P)\mathbf{d} \leqslant \mathbf{0}$. This contradicts the MF constraint qualification. □

To prove the FJ and KKT conditions for problem (Q) without using the implicit function theorem we consider for a local minimizer $\mathbf{x}_Q$ of problem (Q) and $\delta > 0$ appropriately chosen and $\epsilon > 0$, the perturbed feasible region

$$\mathscr{F}_\delta(\epsilon) := \mathscr{F}_P \cap \mathscr{N}(\mathbf{x}_Q, \delta) \cap \{\mathbf{x} \in \mathbb{R}^n : h_j(\mathbf{x}) \leqslant \epsilon, -h_j(\mathbf{x}) \leqslant \epsilon, 1 \leqslant j \leqslant r\},$$

and the associated optimization problem

$$\min\{f_0(\mathbf{x}) + \|\mathbf{x} - \mathbf{x}_Q\|^2 : \mathbf{x} \in \mathscr{F}_\delta(\epsilon)\}. \tag{$\mathrm{Q}_\delta(\epsilon)$}$$

Since the feasible region is compact a global minimizer $\mathbf{x}_Q(\epsilon)$ exists for problem ($Q_\delta(\epsilon)$). For these global minimizers one can show the following result.

Lemma 2.3. *For any sequence* $\epsilon_l \downarrow 0$ *it follows that* $\lim_{l\uparrow\infty}\mathbf{x}_Q(\epsilon_l) = \mathbf{x}_Q$.

Proof. Let us assume to the contrary that there exists a sequence $\mathbf{x}_Q(\epsilon_l), l \in \mathbb{N}$ which does not converge to $\mathbf{x}_Q$. By $\|\mathbf{x}_Q(\epsilon_l) - \mathbf{x}_Q\| \leqslant \delta$ and the Bolzano-Weierstrass property for compact sets there exists some subsequence $\mathbf{x}_Q(\epsilon_l), l \in L \subseteq \mathbb{N}$ satisfying

$$\lim_{l\uparrow\infty,\ l\in L}\mathbf{x}_Q(\epsilon_l) = \bar{\mathbf{x}} \neq \mathbf{x}_Q. \tag{2.1}$$

By continuity $\bar{\mathbf{x}}$ must be feasible for problem (Q). Since $\mathbf{x}_Q$ is feasible for ($Q_\delta(\epsilon_l)$), $l \in L$ it follows that

$$f_0(\mathbf{x}_Q(\epsilon_l)) + \|\mathbf{x}_Q(\epsilon_l) - \mathbf{x}_Q\|^2 \leqslant f_0(\mathbf{x}_Q) \tag{2.2}$$

for every $l \in L$. Taking now the limit in relation (2.2) we find by relation (2.1) that

$$f_0(\bar{\mathbf{x}}) + \|\bar{\mathbf{x}} - \mathbf{x}_Q\|^2 \leqslant f_0(\mathbf{x}_Q)$$

and this contradicts the local optimality of $\mathbf{x}_Q$ for problem (Q). □

If $\mathbf{x}_Q$ is a strict local minimizer, i.e., $f_0(\mathbf{x}_Q) < f_0(\mathbf{x})$ for every $\mathbf{x} \in \mathscr{F}_Q \cap \mathscr{N}(\mathbf{x}_Q, \delta)$ and $\mathbf{x} \neq \mathbf{x}_Q$, we do not need in the above proof the penalty term $\|\mathbf{x} - \mathbf{x}_Q\|^2$. Using Lemma 2.3 one can now give an elementary proof of the FJ and KKT conditions for a local minimizer $\mathbf{x}_Q$ of problem (Q).

Proof (*FJ conditions for problem (Q)*). Let ϵ_l be a strictly decreasing sequence and consider the associated optimal solutions $x_Q(\epsilon_l)$ of $(Q_\delta(\epsilon_l))$. For notational convenience we denote $x_Q(\epsilon_l)$ by $\mathbf{x}^{(l)}$ and by Lemma 2.3 there exists some $l \geqslant l_0$ such that $\|\mathbf{x}^{(l)} - \mathbf{x}_Q\| < \delta$ for every $l \geqslant l_0$. Introduce now the set

$$J_l := \{1 \leqslant j \leqslant r : h_j(\mathbf{x}^{(l)}) = \epsilon_l \quad \text{or} \quad h_j(\mathbf{x}^{(l)}) = -\epsilon_l\}.$$

The set of all subsets of the finite set $\{1,\ldots,r\}$ is finite and so the sequence $J_l, l \in \mathbb{N}$ contains some subset $\overline{J} \subseteq \{1,\ldots,r\}$ such that $L := \{l \in \mathbb{N} : J_l = \overline{J}\}$ is infinite. Applying now for every $l \geqslant l_0$ and $l \in L$ the FJ conditions to problem $(Q_\delta(\epsilon_l))$ we obtain that there exist vectors $\lambda_l \in \mathbb{R}_+^{q+1}$, $\mu_l \in \mathbb{R}^{|\overline{J}|}$, $v_l \in \mathbb{R}_+^m$, $\mathbf{0} \neq (\lambda_l, \mu_l)$, satisfying

$$\begin{aligned}
&-\lambda_{0l}\mathbf{g}(\mathbf{x}^{(l)}) - \sum_{i=1}^{q} \lambda_{il} \nabla f_i(\mathbf{x}^{(l)}) - \sum_{j\in\overline{J}} \mu_{jl} \nabla h_j(\mathbf{x}^{(l)}) = \sum_{k=1}^{m} v_{kl}\mathbf{a}_k, \\
&v_{kl}(\mathbf{a}_k^\top \mathbf{x}^{(l)} - b_k) = 0,\ 1 \leqslant k \leqslant m, \quad \text{and} \quad \lambda_{il} f_i(\mathbf{x}^{(l)}) = 0,\ 1 \leqslant i \leqslant q
\end{aligned} \tag{2.3}$$

with $g(\mathbf{x}) := \nabla f_0(\mathbf{x}) + 2(\mathbf{x} - \mathbf{x}_Q)$. By relation (2.3) and Caratheodory's lemma (see Appendix A) one can find for every $l \in L$ some subset $K_l \subseteq \{1,\ldots,m\}$ and a vector $v_l^* \in \mathbb{R}_+^{|K_l|}$ satisfying

$$-\lambda_{0l}\mathbf{g}(\mathbf{x}^{(l)}) - \sum_{i=1}^{q} \lambda_{il} \nabla f_i(\mathbf{x}^{(l)}) - \sum_{j\in\overline{J}} \mu_{jl} \nabla h_j(\mathbf{x}^{(l)}) = \sum_{k\in K_l} v_{kl}^*\mathbf{a}_k \tag{2.4}$$

and the vectors $\mathbf{a}_k$, $k \in K_l$ are linearly independent. Since $\mathbf{0} \neq (\lambda_l, \mu_l)$ we may assume in relation (2.4) that the vector $(\lambda_l, \mu_l, v_l^*)$ has Euclidean norm 1. Again by selecting an infinite subsequence $L_0 \subseteq L$ if necessary we can assume $K_l = \overline{K}$ (the same) for all $l \in L_0$. By the Bolzano-Weierstrass theorem the sequence of vectors $(\lambda_l, \mu_l, v_l^*), l \in L_0$ has a converging subsequence, i.e., there exists an infinite set $L_1 \subseteq L_0$ with $\lim_{l\in L_1, l\uparrow\infty}(\lambda_l, \mu_l, v_l^*) = (\bar{\lambda}, \bar{\mu}, \bar{v})$ and $(\bar{\lambda}, \bar{\mu}, \bar{v})$ having Euclidean norm 1. Moreover, it follows by Lemma 2.3 and the continuity of h_j that $\overline{J} \subseteq \{1 \leqslant j \leqslant r : h_j(\mathbf{x}_Q) = 0\}$. Applying again Lemma 2.3 and the continuity of the gradients the desired result follows from relation (2.4) by letting $l \in L_1$ converge to infinity leading to the FJ condition:

$$\sum_{i=0}^{q} \bar{\lambda}_i \nabla f_i(\mathbf{x}_Q) + \sum_{j\in\overline{J}} \bar{\mu}_j \nabla h_j(\mathbf{x}_Q) + \sum_{k\in\overline{K}} \bar{v}_k \mathbf{a}_k = \mathbf{0}.$$

By construction the vectors $\mathbf{a}_k$, $k \in \overline{K}$, are linearly independent. Since $(\bar{\lambda}, \bar{\mu}, \bar{v})$ has Euclidean norm 1 and $\mathbf{a}_k$, $k \in \overline{K}$, are linearly independent this implies $(\bar{\lambda}, \bar{\mu}) \neq \mathbf{0}$. □

For problem (Q) we introduce the following constraint qualification: The MF constraint qualification for problem (Q) is said to hold at a feasible point $\mathbf{x}$ if

MF1. $\nabla h_j(\mathbf{x})$, $1 \leqslant j \leqslant r$ are linearly independent.
MF2. $\text{lin}\{\nabla h_j(\mathbf{x}), 1 \leqslant j \leqslant r\} \cap \text{lin}\{\mathbf{a}_k, k \in K(\mathbf{x})\} = \{\mathbf{0}\}$.
MF3. There exists some $\mathbf{d}_0$ satisfying

$$B(\mathbf{x})\mathbf{d}_0 \leqslant \mathbf{0}, \nabla h_j(\mathbf{x})^\top \mathbf{d}_0 = 0, 0 \leqslant j \leqslant r, \quad \text{and} \quad \max_{i\in I(\mathbf{x})}\{\nabla f_i(\mathbf{x})^\top \mathbf{d}_0\} < 0.$$

This is a natural condition. Without condition (MF2) a FJ point need not be a KKT point as shown by the two-dimensional optimization problem (with minimizer and FJ point $\mathbf{x}_Q = \mathbf{0}$)

$$\min\{x_1 : x_2 \leqslant 0, -x_2 \leqslant 0, x_2 - x_1^2 = 0\}.$$

Proof (*KKT conditions for problem (Q)*). To show that at a minimizer $\mathbf{x}_Q$ of problem (Q) satisfying the MF constraint qualification the KKT condition must hold we assume to the contrary that in the FJ condition for problem (Q) we have $\lambda_0 = 0$. By (MF3) it must follow that $\lambda = \mathbf{0}$ and using $(\lambda, \mu) \neq \mathbf{0}$ it follows that $\mu \neq \mathbf{0}$. Applying now (MF2) and (MF3) to the FJ conditions with $\lambda = \mathbf{0}$ and $\mu \neq \mathbf{0}$ we obtain a contradiction. □

As observed in Remark 2.1 we will now show for f_0 pseudo-convex and f_i, $1 \leqslant i \leqslant q$ strictly pseudo-convex on $\mathbb{R}^n$, that for $\mathbf{x}_P \in \mathscr{F}_P$ the condition $\max\{\nabla f_i(\mathbf{x}_P)^\top \mathbf{d} : i \in I(\mathbf{x}_P) \cup \{0\}\} \geqslant 0$ for every $\mathbf{d}$ such that $B(\mathbf{x}_P)\mathbf{d} \leqslant \mathbf{0}$ implies that $\mathbf{x}_P$ is an global minimizer of problem (P). Recall that a function $\phi : \mathbb{R}^n \mapsto \mathbb{R}$ is called pseudo-convex on $\mathbb{R}^n$ if ϕ is differentiable on $\mathbb{R}^n$ and $\nabla\phi(\mathbf{x})^\top \mathbf{d} \geqslant 0$ implies $\phi(\mathbf{x}+\mathbf{d}) \geqslant \phi(\mathbf{x})$ for every $\mathbf{x}, \mathbf{d} \in \mathbb{R}^n$. It is called strictly pseudo-convex on $\mathbb{R}^n$ if ϕ is differentiable and $\nabla\phi(\mathbf{x})^\top \mathbf{d} \geqslant 0$ implies $\phi(\mathbf{x}+\mathbf{d}) > \phi(\mathbf{x})$ for every $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{0} \neq \mathbf{d} \in \mathbb{R}^n$ [1].

Proof (*Converse of Lemma 2.1 for f_0* pseudo-convex and f_i,$1 \leqslant i \leqslant q$ strictly pseudo-convex). To prove the converse of Lemma 2.1 let us assume by contradiction that the feasible $\mathbf{x}_P$ is not an global minimizer of problem (P). Hence, there exists some $\mathbf{x}_0 \in \mathscr{F}_P$ satisfying $f_0(\mathbf{x}_0) < f_0(\mathbf{x}_P)$. By the pseudo-convexity of f_0 this implies that $\nabla f_0(\mathbf{x}_P)^\top(\mathbf{x}_0 - \mathbf{x}_P) < 0$. Also by strict pseudo-convexity of f_i,$1 \leqslant i \leqslant q$ using $f_i(\mathbf{x}_0) \leqslant 0 = f_i(\mathbf{x}_P)$,$i \in I(\mathbf{x}_P)$ and $\mathbf{x}_0 \neq \mathbf{x}_P$ we obtain that $\nabla f_i(\mathbf{x}_P)^\top(\mathbf{x}_0 - \mathbf{x}_P) < 0$ for every $i \in I(\mathbf{x}_P)$. Finally it holds that $B(\mathbf{x}_P)(\mathbf{x}_0 - \mathbf{x}_P) \leqslant \mathbf{0}$ and we arrive at a contradiction to our initial assumption. □

Combining Lemmas 2.1 and 2.2 we immediately obtain the following result [2].

Lemma 2.4. *Let f_0 be pseudo-convex and f_i,$1 \leqslant i \leqslant q$ strictly pseudo-convex. Then it follows that $\mathbf{x}_P \in \mathscr{F}_P$ is a global minimizer of (P) if and only if $\mathbf{x}_P$ satisfies the FJ conditions.*

3. Conclusion

In this note we have shown that the basic results in nonlinear programming are a natural and direct consequence of basic results in linear programming and analysis. In our proof we could avoid the implicit function theorem usually applied in the proof of the FJ conditions for problem (Q) (see for example [2,5]). The proof of the implicit function theorem [11] and its understanding is in general difficult for undergraduate/graduate students in the applied computational sciences. This concern was also the main objective for constructing an alternative elementary proof by McShane [10] for the FJ and KKT conditions for problem (Q). By not regarding separately linear and nonlinear inequalities the result in [10] is weaker than ours (also the linear independence constraint qualification for (Q) is used) and his proof uses the penalty approach of nonlinear programming (see also [3] for a similar proof). As such this technique and the technique used in this paper have their pros and cons. An advantage of the presented approach for problem (P) is the fact that it can easily identify the class of functions for which the FJ conditions for problem (P) are not only necessary but also sufficient. This seems to be difficult to show by means of the penalty approach of McShane. However, to our belief the main advantage of our proof technique is its display of a natural connection between linear and nonlinear programming.

Appendix A

In this appendix we give a short proof of Lemma 2.2 by means of the strong duality theorem for linear programming.

Proof. To verify $1 \Rightarrow 2$ we observe that

$$0 = \min_{B\mathbf{d} \leqslant \mathbf{0}} \max_{1 \leqslant i \leqslant s} \mathbf{c}_i^\top \mathbf{d} = \min_{B\mathbf{d} \leqslant \mathbf{0},\, \mathbf{c}_i^\top \mathbf{d} - z \leqslant 0,\, 1 \leqslant i \leqslant s} z. \tag{A.1}$$

This is a linear programming problem and by the strong duality theorem of linear programming (cf. [4]) we obtain

$$\min_{B\mathbf{d} \leqslant \mathbf{0},\, \mathbf{c}_i^\top \mathbf{d} - z \leqslant 0,\, 1 \leqslant i \leqslant s} z = \max\left\{ \mathbf{0}^\top \begin{pmatrix} \lambda \\ \mu \end{pmatrix} : \sum_{i=1}^{s} \lambda_i \mathbf{c}_i + B^\top \mu = \mathbf{0}, \lambda \in \Delta_s, \mu \in \mathbb{R}_+^p \right\}. \tag{A.2}$$

Applying now relations (A.1) and (A.2) we know that the feasible region of the dual problem is not empty and so there exist some $\lambda \in \Delta_s$ and $\mu \in \mathbb{R}_+^p$ satisfying $\sum_{i=1}^s \lambda_i \mathbf{c}_i^\top \mathbf{d} + B^\top \mu = \mathbf{0}$. To show the reverse implication it follows that there exists some $\lambda \in \Delta_s$ and $\mu \in \mathbb{R}_+^p$ satisfying $\sum_{i=1}^s \lambda_i \mathbf{c}_i^\top \mathbf{d} = -\mu^\top B\mathbf{d}$ for every $\mathbf{d} \in \mathbb{R}^n$. Hence for $B\mathbf{d} \leqslant \mathbf{0}$ and using $\mu \in \mathbb{R}_+^p$ we obtain $\max_{1 \leqslant i \leqslant s} \mathbf{c}_i^\top \mathbf{d} \geqslant \sum_{i=1}^s \lambda_i \mathbf{c}_i^\top \mathbf{d} \geqslant 0$. □

In our analysis we also use the following result known as Caratheodory's lemma.

Lemma A.1. *Let* $\mathbf{v} \in \mathbb{R}^m$ *be represented as cone combination* $\mathbf{v} = \sum_{k=1}^{m} v_k \mathbf{a}_k,\ v_k \geqslant 0$. *Then there is a representation* $\mathbf{v} = \sum_{k \in \overline{K}} \bar{v}_k \mathbf{a}_k,\ \bar{v}_k > 0, k \in \overline{K}$ *such that* $\mathbf{a}_k, k \in \overline{K}$ *are linearly independent.*

Proof. We can assume

$$\mathbf{v} = \sum_{k=1}^{m} v_k \mathbf{a}_k, \quad \text{with } v_k > 0, \tag{A.3}$$

and suppose that the vectors $\mathbf{a}_k, k = 1, \ldots, m$ are linearly dependent. So there is a non-trivial combination $\mathbf{0} = \sum_{k=1}^{m} \tau_k \mathbf{a}_k$. By multiplying this relation by a factor ρ and adding to (A.3) we find

$$\mathbf{v} = \sum_{k=1}^{m} (v_k + \rho\tau_k) \mathbf{a}_k$$

and see that we can choose $\rho \in \mathbb{R}$ in such a way that (at least) one of the coefficients $(v_k + \rho\tau_k)$ is zero and the others $\geqslant 0$. This can be done until the desired representation is attained. □

References

[1] M. Avriel, W.E. Diewert, S. Schaible, I. Zang, Generalized Concavity, Mathematical Concepts and Methods in Science and Engineering, vol. 36, Pleneum Press, New York, 1988.
[2] M.S. Bazaraa, H.D. Sherali, C.M. Shetty, Nonlinear Programming: Theory and Algorithms, second ed. Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley, New York, 1993.
[3] D.P. Bertsekas, Nonlinear Programming, Athena Scientific, Massachusetts, 1995.
[4] V. Chvátal, Linear Programming, W.H. Freeman and Company, New York, 1999.
[5] U. Faigle, K. Kern, G. Still, Algorithmic Principles to Mathematical Programming, Kluwer Academic publishers., Dordrecht, 2002.
[6] J. Farkas, Theorie der einfachen Ungleichungen, Journal für die reine und angewandte Mathematic 124 (1902) 1–27.
[7] F. John, Extremum problems with inequalities as side conditions, in: K.O. Friedrichs, O.E. Neugebauer, J.J. Stoker (Eds.), Studies and Essays, Courant Anniversary Volume, Wiley-Interscience, 1948.
[8] W. Karush, Minima of functions of several variables with inequalities as side conditions. Master's thesis, Department of Mathematics, University of Chicago, 1939.
[9] H.W. Kuhn, A.W. Tucker, Nonlinear programming, in: J. Neyman (Ed.), Proceedings of Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, 1951.
[10] E.J. McShane, The Lagrange multiplier rule, The American Mathematical Monthly 80 (8) (1973) 922–925.
[11] W. Rudin, Principles of Mathematical Analysis, third ed., Mc-Graw Hill, New York, 1976.