Sharpening the Karush-John optimality conditions

Arnold Neumaier and Hermann Schichl

Institut für Mathematik, Universität Wien
Strudlhofgasse 4, A-1090 Wien, Austria
email: Arnold.Neumaier@univie.ac.at, Hermann.Schichl@esi.ac.at
WWW: http://www.mat.univie.ac.at/~neum/

July 23, 2003

Abstract. We present a refined version of the Karush-John first order optimality conditions which reduces the number of constraints for which a constraint qualification is needed. This version is a generalization both of the Karush-John conditions and of the first order optimality conditions for concave constraints.

Keywords: optimality conditions, Kuhn-Tucker conditions, Karush-John conditions

2000 MSC Classification: 90C30

1 Introduction

The Karush-John optimality conditions (often also called Fritz John conditions, after Fritz John, who rediscovered Karush's results; see Karush [5], John [4] and Schrijver [12, p.220] for a history) play a central role in constrained optimization,, characterizing the solutions of smooth nonlinear programming problems by first order necessary conditions. However, the Karush-John conditions in their most general form pose difficulties in applications, because the factor in front of the gradient term may be zero or very small. Therefore, most of the local solvers require a constraint qualification, like that of Mangasarian & Fromovitz [9], to be able to reduce the Karush-John conditions to the more convenient Kuhn-Tucker conditions [7]. Thorough discussions of such constraint qualifications can be found in Bazaraa et al. [2] and Mangasarian [8].

Deterministic global optimization algorithms cannot take this course, since it is not known beforehand whether the global optimum satisfies an assumed constraint qualification. Therefore, they have to use the Karush-John conditions in their general form (cf., e.g., Kearfott [6]). Unfortunately, the additional constraints needed involve all multipliers and are very inconvenient for the solution process.

In this article we derive a stronger form of the Karush-John optimality conditions for general smooth nonlinear programming problems. These conditions imply the (known) result that for concavely (or linearly) constrained problems no constraint qualification is needed, and the derived Kuhn-Tucker conditions requires linear independence constraint qualifications for fewer constraints than the conditions found in the literature. The new conditions are incorporated in the COCONUT environment [3] for deterministic global optimization.

2 Known results

The situation is simplest when the constraints are concave, i.e., of the form $F(x) \geq 0$ with convex F. Due to the concave structure of the feasible set descent from nonoptimal points can be achieved using linear paths.

2.1 Theorem. (First order optimality conditions for concave constraints)
Let $\widehat{x} \in \mathbb{R}^n$ be a solution of the nonlinear program

$$\begin{aligned} &min\ \ f(x) \\ &s.t.\ \ \ F(x) \geq 0 \end{aligned} \tag{1}$$

where $f : C_0 \to \mathbb{R}$ and $F : C_0 \to \mathbb{R}^r$ are continuously differentiable functions defined on their domain of definition.

If F is convex then there is a vector $\widehat{z} \in \mathbb{R}^r$ such that

$$g(\widehat{x}) = F'(\widehat{x})^T\widehat{z}, \quad \inf(\widehat{z}, F(\widehat{x})) = 0. \tag{2}$$

Here $\inf(u, v)$ denotes the infimum (componentwise minimum) of the vectors u and v; a condition of the form $\inf(u, v) = 0$ is generally referred to as a **complementarity condition**.

The proof can be found, e.g., in [8, Section 7.3]; the assumptions amount to the reverse convex constraint qualification by Arrow et al. [1].

In general, curved paths may be needed to get descent, and without some condition on the geometry one does not arrive at the Kuhn-Tucker conditions (2). The conditions for the general case, stated, e.g., in [8, Section 11.3], follow from the following, equivalent statement when slack variables are introduced to eliminate general inequality constraints. Our formulation is adapted to the use of interval analysis [6, 10, 11] in branch and bound methods for global optimization. A **box** is a set of the form

$$\boldsymbol{x} = [\underline{x}, \overline{x}] = \{x \in \mathbb{R}^n \mid \underline{x} \leq x \leq \overline{x}\},$$

with componentwise inequalities; the bounding vectors are allowed to contain infinite entries indicating missing bounds.

2.2 Theorem. (Karush-John first order optimality conditions)
Let $\widehat{x} \in \mathbb{R}^n$ be a solution of the nonlinear program

$$\begin{aligned} &min\ \ f(x) \\ &s.t.\ \ \ F(x) = 0 \\ &\qquad\ \ x \in \boldsymbol{x}, \end{aligned} \tag{3}$$

where $f : C_0 \to \mathbb{R}$ and $F : C_0 \to \mathbb{R}^r$ are continuously differentiable functions defined on their domain of definition.

Then there are a constant $\kappa \geq 0$ and a vector $\widehat{z} \in \mathbb{R}^r$, not both zero, such that

$$\widehat{y} := \kappa g(\widehat{x}) - F'(\widehat{x})^T\widehat{z} \tag{4}$$

satisfies the two-sided complementarity condition

$$\begin{array}{ll} \widehat{y}_k \geq 0 & \text{if } \widehat{x}_k = \underline{x}_k, \\ \widehat{y}_k \leq 0 & \text{if } \widehat{x}_k = \overline{x}_k, \\ \widehat{y}_k = 0 & \text{otherwise.} \end{array} \tag{5}$$

In the following we prove a common generalization of both Theorem 2.1 and Theorem 2.2. Apart from the inverse function theorem, our mail tool is the following transposition theorem, which is a simple application of the Lemma of Farkas.

2.3 Theorem. *Let $B \in \mathbb{R}^{m\times n}$, and let (I, J, K) be a partition of $\{1, \ldots, m\}$. Then exactly one of the following holds:*

(i) $(Bp)_I = 0$, $(Bp)_J \geq 0$, $(Bp)_K > 0$ *for some* $p \in \mathbb{R}^n$,

(ii) $B^T q = 0$, $q_{J\cup K} \geq 0$, $q_K \neq 0$ *for some* $q \in \mathbb{R}^m$.

3 A refinement of the Karush–John conditions

We consider concave and nonconcave constraints separately, and introduce slack variables to transform all nonconcave constraints into equations. Thus we may write the nonlinear optimization problems without loss of generality in the form

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & C(x) \geq 0 \\ & F(x) = 0. \end{array} \tag{6}$$

The form (6), which separates the concave constraints (including bound constraints and general linear constraints) and the remaining nonlinear constraints, is most useful to prove our strong form of the Karush-John conditions. However, in computer implementations, a transformation to this form is not ideal, and the slack variables should be eliminated again from the optimality conditions.

3.1 Theorem. (General first order optimality conditions)
Let $\widehat{x} \in \mathbb{R}^n$ be a solution of the nonlinear program (6), where $f : U \to \mathbb{R}$, $C : U \to \mathbb{R}^m$, and $F : U \to \mathbb{R}^r$ are functions continuously differentiable on a neighborhood U of $\widehat{x}$. In addition, C shall be convex on U. Then there is a constant $\kappa \geq 0$ and there are vectors $\widehat{y} \in \mathbb{R}^m$, $\widehat{z} \in \mathbb{R}^r$ such that

$$\kappa g(\widehat{x}) = C'(\widehat{x})^T \widehat{y} + F'(\widehat{x})^T \widehat{z}, \tag{7}$$

$$\inf(\widehat{y}, C(\widehat{x})) = 0, \tag{8}$$

$$F(\widehat{x}) = 0, \tag{9}$$

and

$$\kappa, \widehat{z} \text{ are not both zero.} \tag{10}$$

In contrast, the standard Karush-John condition asserts in this case only that $\kappa, \widehat{y}, \widehat{z}$ are not all three zero. Thus the present version gives more information in case that $\kappa = 0$; therefore, weaker constraint qualifications are needed to ensure that $\kappa \neq 0$ (in which case one can scale the multipliers so that $\kappa = 1$).

Proof. We begin by noting that a feasible point $\widehat{x}$ of (6) is also a feasible point for the optimization problem

$$\begin{aligned} \min \;\; & f(x) \\ \text{s.t.} \;\; & Ax \geq b \\ & F(x) = 0, \end{aligned} \tag{11}$$

where J is the set of all components j for which $C(\widehat{x})_j = 0$ and

$$A = C'(\widehat{x})_{J:}, \quad b = C'(\widehat{x})_{J:}\widehat{x}.$$

For the indices k corresponding to the set N of inactive constraints, we choose $y_N = 0$ to satisfy condition (8). Since C is convex, we have $C(x) \geq C(\widehat{x}) + C'(\widehat{x})(x - \widehat{x})$. Restricted to the rows J we get $C(x)_J \geq C'(\widehat{x})_{J:}(x - \widehat{x})$. This fact implies that problem (6) is a relaxation of problem (11) on a neighborhood U of $\widehat{x}$. Note that since C is continuous we know that $C(x)_j > 0$ for $k \in N$ in a neighborhood of $\widehat{x}$ for all constraints with $C(\widehat{x})_j > 0$. Since, by assumption, $\widehat{x}$ is a local optimum of a relaxation of (11) and a feasible point of (11), it is a local optimum of (11) as well. Together with the choice $y_N = 0$ the Karush-John conditions of problem (11) are again conditions (7)–(9). So we have successfully reduced the problem to the case where C is an affine function and all constraints are active at $\widehat{x}$.

Thus, in the following, we consider a solution $\widehat{x}$ of the optimization problem (11) satisfying

$$A\widehat{x} = b. \tag{12}$$

If $\operatorname{rk} F'(\widehat{x}) < r$ then $z^T F'(\widehat{x}) = 0$ has a solution $z \neq 0$, and we can solve (7)–(10) with $y = 0$, $\kappa = 0$. Hence we may assume that $\operatorname{rk} F'(\widehat{x}) = r$. This allows us to select a set R of r column indices such that $F'(\widehat{x})_{:R}$ is nonsingular. Let B be the $(0,1)$-matrix such that Bs is the vector obtained from $s \in \mathbb{R}^n$ by discarding the entries indexed by R. Then the function $\Phi : C \to \mathbb{R}^n$ defined by

$$\Phi(x) := \begin{pmatrix} F(x) \\ Bx - B\widehat{x} \end{pmatrix}$$

has at $x = \widehat{x}$ a nonsingular derivative

$$\Phi'(\widehat{x}) = \begin{pmatrix} F'(\widehat{x}) \\ B \end{pmatrix}.$$

Hence, by the inverse function theorem, Φ defines in a neighborhood of $0 = \Phi(\widehat{x})$ a unique continuously differentiable inverse function Φ^{-1} with $\Phi^{-1}(0) = \widehat{x}$. Using Φ we can define a curved search path with tangent vector $p \in \mathbb{R}^n$ tangent to the nonlinear constraints satisfying $F'(\widehat{x})p = 0$. Indeed, the function defined by

$$s(\alpha) := \Phi^{-1}\begin{pmatrix} 0 \\ \alpha Bp \end{pmatrix} - \widehat{x}$$

for sufficiently small $\alpha \geq 0$, is continuously differentiable, with

$$s(0) = \Phi^{-1}(0) - \widehat{x} = 0, \quad \begin{pmatrix} F(\widehat{x} + s(\alpha)) \\ Bs(\alpha) \end{pmatrix} = \Phi\left(\Phi^{-1}\begin{pmatrix} 0 \\ \alpha Bp \end{pmatrix}\right) = \begin{pmatrix} 0 \\ \alpha Bp \end{pmatrix},$$

hence

$$s(0) = 0, \quad F(\widehat{x} + s(\alpha)) = 0, \quad Bs(\alpha) = \alpha Bp. \tag{13}$$

Differentiation of (13) at $\alpha = 0$ yields

$$\begin{pmatrix} F'(\widehat{x}) \\ B \end{pmatrix} \dot{s}(0) = \begin{pmatrix} F'(\widehat{x})\dot{s}(0) \\ B\dot{s}(0) \end{pmatrix} = \begin{pmatrix} 0 \\ Bp \end{pmatrix} = \begin{pmatrix} F'(\widehat{x}) \\ B \end{pmatrix} p,$$

hence $\dot{s}(0) = p$, i.e., p is indeed a tangent vector to $\widehat{x} + s(\alpha)$ at $\alpha = 0$.

Now we consider a direction $p \in \mathbb{R}^n$ such that

$$g^T p < 0, \quad g = g(\widehat{x}), \tag{14}$$

$$Ap > 0, \tag{15}$$

$$F'(\widehat{x})p = 0. \tag{16}$$

(In contrast to the purely concave case, we need the strict inequality in (15) to take care of curvature terms.) Since $A\widehat{x} \geq b$ and (15) imply $A(\widehat{x} + s(\alpha)) = A(\widehat{x} + \alpha\dot{s}(0) + o(\alpha)) = A\widehat{x} + \alpha(Ap + o(1)) \geq b$ for sufficiently small $\alpha \geq 0$, (13) implies feasibility of the points $\widehat{x} + s(\alpha)$ for small $\alpha \geq 0$. Since

$$\frac{d}{d\alpha} f(\widehat{x} + s(\alpha)) \Big|_{\alpha=0} = g^T \dot{s}(0) = g^T p < 0,$$

f decreases strictly along $\widehat{x} + s(\alpha)$, α small, contradicting the assumption that $\widehat{x}$ is a solution of (6). This contradiction shows that the condition (14)–(16) are inconsistent. Thus, the transposition theorem 2.3 applies with

$$\begin{pmatrix} -g^T \\ A \\ F'(\widehat{x}) \end{pmatrix}, \quad \begin{pmatrix} \kappa \\ y \\ z \end{pmatrix} \quad \text{in place of } B, q,$$

and shows the solvability of

$$-g\kappa + A^T y + F'(\widehat{x})^T z = 0, \quad \kappa \geq 0, \quad y \geq 0, \quad \begin{pmatrix} \kappa \\ y \end{pmatrix} \neq 0.$$

If we put $\widehat{z} = z$, let $\widehat{y}$ be the vector with $\widehat{y}_J = y$ and zero entries elsewhere, and note that $\widehat{x}$ is feasible, we find (7)–(9).

Because of (10), it remains to discuss the case where $\kappa = 0$ and $z = 0$, and therefore

$$A^T y = 0, \quad y \neq 0. \tag{17}$$

In this case, $b^T y = (A\widehat{x})^T y = \widehat{x}^T A^T y = 0$. Therefore any point $x \in U$ satisfies $(Ax - b)^T y = x^T A^T y - b^T y = 0$, and since $y \geq 0$, $Ax - b \geq 0$, we see that the set

$$K := \{i \mid (Ax)_i = b_i \text{ for all } x \in U\}$$

contains all indices i with $y_i \neq 0$ and hence is nonempty.

Since U is nonempty, the system $A_{K:}x = b_K$ is consistent, and hence equivalent to $A_{L:}x = b_L$, where L is a maximal subset of K such that the rows of A indexed by L are linearly independent. If M denotes the set of indices complementary to K, we can describe the feasible set equivalently by the constraints

$$A_{M:}x \geq b_M, \quad \begin{pmatrix} A_{L:}x - b_L \\ F(x) \end{pmatrix} = 0.$$

In this modified description of the feasible set has no equality constraints implicit in the inequality $A_{M:}x \geq b_M$. For a solution $\widehat{x}$ of the equivalent optimization problem with these constraints, we find as before a number $\kappa \geq 0$ and vectors y_M and $\begin{pmatrix} y_L \\ z \end{pmatrix}$ such that

$$\kappa g(\widehat{x}) = A_{M:}^T y_M + \begin{pmatrix} A_{L:} \\ F'(\widehat{x}) \end{pmatrix}^T \begin{pmatrix} y_L \\ z \end{pmatrix}, \tag{18}$$

$$\inf(y_M, A_{M:}\widehat{x} - b_M) = 0, \tag{19}$$

$$F(x) = 0, \quad A_{K:}\widehat{x} - b_K = 0, \tag{20}$$

$$\kappa, \begin{pmatrix} y_L \\ z \end{pmatrix} \text{ are not both zero.} \tag{21}$$

Clearly, this yields vectors $\widehat{y} = y$ and $\widehat{z} = z$ satisfying (7) and (8), but now $y_{K \setminus L} = 0$. The exceptional situation $\kappa = 0$, $z = 0$ can now no longer occur. Indeed, as before, all indices i with $y_i \neq 0$ lie in K; hence $y_M = 0$ and (18) gives $A_{L:}^T y_L = 0$. Since, by construction, the rows of $A_{L:}$ are linearly independent, this implies $y_L = 0$, contradicting (21). This completes the proof. □

Theorem 2.1 is the special case where F is zero-dimensional, and Theorem 2.2 follows directly from the special case where $C(x) = \begin{pmatrix} x - \underline{x} \\ \overline{x} - x \end{pmatrix}$.

3.2 Corollary. *Under the assumptions of Theorem* 3.1, *if the* **constraint qualification**

$$C'(\widehat{x})_{J:}^T y_J + F'(\widehat{x})^T z = 0, \ y_J \geq 0 \quad \Rightarrow \quad z = 0 \tag{22}$$

holds then the conclusion of Theorem 3.1 *holds with* $\kappa = 1$.

A slightly more restrictive constraint qualification, which, however, is easier to interpret geometrically, is obtained if we remove the nonnegativity condition from the assumption in (22). The resulting condition

$$C'(\widehat{x})_{J:}^T y_J + F'(\widehat{x})^T z = 0 \quad \Rightarrow \quad z = 0 \tag{23}$$

implies that $\mathrm{rk} F'(\widehat{x}) = r$ (to see this put $y = 0$). Writing the left hand side of (23) as $y_J^T C'(\widehat{x})_{J:} = -z^T F'(\widehat{x})$, we see that (23) forbids precisely common nonzero vectors in the **row spaces** (spanned by the rows) of $C'(\widehat{x})_{J:}$ and $F'(x)$, respectively. Thus we get the following useful form of the optimality conditions:

3.3 Corollary. *Under the assumption of Theorem* 3.1, *if* $\mathrm{rk}\, F'(\widehat{x}) = r$ *and if the row spaces of* $F'(\widehat{x})$ *and* $C'(\widehat{x})_{J:}$, *where* $J = \{i \mid C(\widehat{x})_i = 0\}$, *have trivial intersection only, then there are vectors* $\widehat{y} \in \mathbb{R}^m$, $\widehat{z} \in \mathbb{R}^r$ *such that*

$$g(\widehat{x}) = C'(\widehat{x})^T \widehat{y} + F'(\widehat{x})^T \widehat{z}, \tag{24}$$

$$\inf(\widehat{y}, C(\widehat{x})) = 0, \tag{25}$$

$$F(\widehat{x}) = 0. \tag{26}$$

(24)–(26) are the **Kuhn-Tucker conditions** for the nonlinear program (6), cf. [7]. The traditional linear independence constraint qualification requires in place of the assumptions in Corollary 3.3 the stronger condition that the rows of $\begin{pmatrix} F'(\widehat{x}) \\ C'(\widehat{x}) \end{pmatrix}$ are independent. In contrast, our condition allows arbitrary dependences among the rows of $C'(\widehat{x})$.

Note that in view of (8), the condition (10) can be written (after rescaling) in the equivalent form

$$\kappa \geq 0, \quad \kappa + u^T\widehat{y} + \widehat{z}^T D\widehat{z} = 1, \tag{27}$$

where u is an arbitrary nonnegative vector with $u_J > 0$, $u_N = 0$ and D is an arbitrary diagonal matrix with positive diagonal entries. This form is numerically stable in that all multipliers are bounded and near degeneracies – which would produce huge multipliers in the Kuhn-Tucker conditions – are revealed by small values of κ. The lack of a constraint qualification (which generally cannot be established in finite precision arithmetic anyway) therefore simply appears as the limit $\kappa = 0$.

The formulation (27) is particularly useful for the rigorous verification of the existence of a solution of our refined Karush-John conditions in the vicinity of an approximate solution; cf. KEARFOTT [6, Section 5.2.5] for the corresponding use of the standard Karush-John conditions. The advantage of our stronger formulation is that in case there are only few nonconcave constraints, condition (27) involves only a few variables and hence is a much stronger constraint if constraint propagation techniques [6, 13] are applied to the optimality conditions.

References

[1] K.J. Arrow, L. Hurwicz and H. Uzawa, Constraint qualifications in maximization problems, Naval Research Logistics Quarterly 8 (1961), 175–191.

[2] M.S. Bazaraa, H.D. Sherali and C.M. Shetty, Nonlinear Programming. Theory and Algorithms, 2nd edition, Wiley, New York 1993.

[3] COCONUT, COntinuous CONstraints - Updating the Technology. `http://www.mat.univie.ac.at/~neum/glopt/coconut/`

[4] F. John. Extremum problems with inequalities as subsidiary conditions. pp. 543–560 in: *Fritz John, Collected Papers, Vol.2* (J. Moser, ed.), Birkhäuser, Boston (1985).

[5] W. Karush. *Minima of Functions of Several Variables with Inequalities as Side Constraints.* M.Sc. dissertation, Dept. of Mathematics, Univ. of Chicago, Chicago, Ill. (1939).

[6] R. B. Kearfott. *Rigorous Global Search: Continuous Problems.* Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

[7] H. W. Kuhn and A. W. Tucker. Nonlinear programming. In J. Neyman, editor, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 481–492. Univ. of California Press, Berkeley, California, 1951.

[8] O.L. Mangasarian, Nonlinear Programming, McGraw-Hill New York 1969. Reprinted as: Classics in Applied Mathematics, SIAM, Philadelphia 1994.

[9] O. L. Mangasarian and S. Fromovitz, *The Fritz John necessary optimality conditions in the presence of equality and inequality constraints*, J. Math. Anal. Appl. 17 (1967), 37–47.

[10] A. Neumaier, *Interval Methods for Systems of Equations*, Cambridge University Press, Cambridge, 1990.

[11] A. Neumaier. *Introduction to Numerical Analysis.* Cambridge Univ. Press, Cambridge, 2001.

[12] A. Schrijver. *Theory of Linear and Integer Programming.* Wiley, Chichester, 1986.

[13] P. Van Hentenryck, L. Michel, and Y. Deville. *Numerica - A Modeling Language for Global Optimization.* MIT Press, Cambridge, MA, 1997.