Optim Lett (2009) 3:7–10
DOI 10.1007/s11590-008-0096-3

ORIGINAL PAPER

A simple and elementary proof of the Karush–Kuhn–Tucker theorem for inequality-constrained optimization

Olga A. Brezhneva · Alexey A. Tret'yakov · Stephen E. Wright

Received: 14 June 2008 / Accepted: 18 July 2008 / Published online: 6 August 2008

Abstract We present an elementary proof of the Karush–Kuhn–Tucker Theorem for the problem with nonlinear inequality constraints and linear equality constraints. Most proofs in the literature rely on advanced optimization concepts such as linear programming duality, the convex separation theorem, or a theorem of the alternative for systems of linear inequalities. By contrast, the proof given here uses only basic facts from linear algebra and the definition of differentiability.

Keywords Nonlinear programming · Karush–Kuhn–Tucker theorem

In this paper, we prove the classical necessary conditions for optimality in the following nonlinear programming problem:

$$\min_x \{f(x) \mid g(x) \le 0, \ Ax = b\}, \tag{1}$$

O. A. Brezhneva (✉) · S. E. Wright
Department of Mathematics and Statistics, Miami University, Oxford, OH 45056, USA
e-mail: brezhnoa@muohio.edu

S. E. Wright
e-mail: wrightse@muohio.edu

A. A. Tret'yakov
Dorodnicyn Computing Center of the Russian Academy of Sciences, Vavilova 40, 119991 Moscow GSP-1, Russia

A. A. Tret'yakov
System Research Institute, Polish Academy of Sciences, Newelska 6, 01-447 Warsaw, Poland

A. A. Tret'yakov
University of Podlasie in Siedlce, 3 Maja 54, 08-110 Siedlce, Poland
e-mail: tret@ap.siedlce.pl

where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, $A \in \mathbb{R}^{k\times n}$, and $b \in \mathbb{R}^k$. The vector inequality $g(x) \leq 0$ is understood to hold entry-wise. The Karush–Kuhn–Tucker Theorems are the analogues of the Lagrange multiplier theorem for problems involving inequality constraints.

We assume that f and g are differentiable at a given point x^*. We use g' to denote the matrix whose rows are the gradients of the corresponding entries in the vector function g, and g'_I to denote the submatrix of g' given by rows with indices in $I = \{i \mid g_i(x^*) = 0\}$.

Theorem 1 (Karush, Kuhn, Tucker). *Suppose that x^* is a local minimizer for problem* (1) *and that the matrix*

$$\begin{bmatrix} g'_I(x^*) \\ A \end{bmatrix}$$

has full row rank. Then there exists $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^k$ satisfying the Karush–Kuhn–Tucker conditions:

(a) $\lambda^* \geq 0$;
(b) $(\lambda^*)^T g(x^*) = 0$;
(c) $f'(x^*) + (\lambda^*)^T g'(x^*) + (\mu^*)^T A = 0$.

This theorem was originally derived by Karush [5] and Kuhn and Tucker [6]. Prékopa [10] discusses the historical precedents for the theorem, whereas a modern exposition of such multiplier rules is given by Rockafellar [11].

Karush, Kuhn and Tucker imposed a weaker hypothesis than the full-rank condition used here. Even in the form stated in Theorem 1, however, most proofs rely on advanced concepts such as linear programming duality, the convex separation theorem, or a theorem of the alternative for systems of linear inequalities. For example, the proofs in Bazaraa et al. [1] and Birbil et al. [4] each employ a theorem of the alternative, whereas Pourciau [9] presents a proof using convex separation. Although these proofs allow for a weaker hypothesis of the form $\{d \mid Ad = 0,\ g'(x^*)d < 0\} \neq \emptyset$, verifying that hypothesis may still require an understanding of the theory of linear inequalities. Prior knowledge of such optimization concepts can be avoided if f and g are assumed to be *continuously* differentiable, in which case the known proofs instead use real analysis concepts (and allow for consideration of nonlinear equality constraints). A standard proof along these lines uses the implicit-function theorem (e.g., [7]), which is not generally considered elementary. A more elementary proof is given by McShane [8] (cf., [2,3,11]), which relies instead on a penalty reformulation and several applications of the Bolzano-Weierstrass theorem.

By contrast, the proof given below uses only basic facts from linear algebra and the definition of differentiability. This makes our proof particularly suitable for use in courses not aimed specifically at optimization theory, such as mathematical modeling, mathematical methods for science or engineering, problems seminars, or any other circumstance where a quick treatment is required. Our proof and hypotheses are sufficient for many such uses.

In the remainder of this note, we shall assume that the hypotheses of Theorem 1 hold. Without loss of generality, we may assume that $I = \{1, \ldots, m\}$, because the continuity of g_i for $i \notin I$ prevents g_i from taking the value 0 on some neighborhood of x^*. For such inactive constraints, condition (b) in the theorem amounts to choosing $\lambda_i^* = 0$; note that condition (b) is trivially satisfied for $i \in I$.

We need the following standard lemma (cf. [1, Exercise 4.43] or [4, Lemma 2.1]).

Lemma 1 *For every scalar* $\zeta < 0$ *and vector* $z < 0$*, there are no solutions* d *to the linear system*

$$\begin{bmatrix} f'(x^*) \\ g'(x^*) \\ A \end{bmatrix} d = \begin{bmatrix} \zeta \\ z \\ 0 \end{bmatrix}. \tag{2}$$

Proof Suppose (2) does have a solution d for some negative (ζ, z). By the assumed differentiability, there exist functions $u : \mathbb{R} \to \mathbb{R}$ and $w : \mathbb{R} \to \mathbb{R}^m$ satisfying $\lim_{t\to 0} |u(t)|/t = 0$, $\lim_{t\to 0} \|w(t)\|/t = 0$, and

$$\begin{aligned} \begin{bmatrix} f(x^* + td) \\ g(x^* + td) \\ A(x^* + td) \end{bmatrix} &= \begin{bmatrix} f(x^*) + tf'(x^*)d + u(t) \\ g(x^*) + tg'(x^*)d + w(t) \\ Ax^* + tAd \end{bmatrix} \\ &= \begin{bmatrix} f(x^*) \\ 0 \\ b \end{bmatrix} + t\left(\begin{bmatrix} \zeta \\ z \\ 0 \end{bmatrix} + \begin{bmatrix} u(t)/t \\ w(t)/t \\ 0 \end{bmatrix} \right) \end{aligned}$$

for all $t \neq 0$. Thus, for all sufficiently small $t > 0$, we have $f(x^* + td) < f(x^*)$, $g(x^* + td) < 0$, and $A(x^* + td) = b$. This contradicts the minimality of x^*, thereby proving the lemma. □

Traditional arguments proceed from Lemma 1 by invoking a theorem of the alternative for the linear system (2), whereas our proof consists of two linear algebraic steps.

Proof of Theorem 1 First, we observe that some $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^k$ must satisfy condition (c) of Theorem 1. Otherwise, the matrix

$$\begin{bmatrix} f'(x^*) \\ g'(x^*) \\ A \end{bmatrix}$$

would have full row rank, implying that (2) admits a solution d for *any* choice of (ζ, z). Second, we show that these multipliers must actually satisfy $\lambda^* \geq 0$. Suppose not, so that some entry in λ^* is negative. We can then choose a vector $q \in \mathbb{R}^m$ satisfying $q < 0$ and $q^T\lambda^* > 0$. For example, we could take $q = \lambda_i^*\mathbf{e} - (1 + |\mathbf{e}^T\lambda^*|)\mathbf{e}^{(i)}$ for any index i having $\lambda_i^* < 0$, where $\mathbf{e}^{(i)}$ and $\mathbf{e}$ denote (respectively) the ith coordinate vector and the vector of all ones in $\mathbb{R}^m$. By the assumption of full row rank, we know that the system

$$\begin{bmatrix} g'(x^*) \\ A \end{bmatrix} d = \begin{bmatrix} q \\ 0 \end{bmatrix}$$

has a solution d. Combining this with condition (c) of Theorem 1 yields

$$\begin{bmatrix} f'(x^*) \\ g'(x^*) \\ A \end{bmatrix} d = \begin{bmatrix} -(\lambda^*)^T g'(x^*)d - (\mu^*)^T Ad \\ g'(x^*)d \\ Ad \end{bmatrix} = \begin{bmatrix} -(\lambda^*)^T q \\ q \\ 0 \end{bmatrix}.$$

Therefore, the linear system (2) has a solution d for the specific choice of $\zeta = -(\lambda^*)^T q < 0$ and $z = q < 0$. This contradiction shows that we must have $\lambda^* \geq 0$ after all, as desired. $\square$

References

1. Bazaraa, M.S., Sherali, H.D., Shetty, C.M.: Nonlinear Programming: Theory and Algorithms, 2nd edn. Wiley, New York (1993)
2. Beltrami, E.J.: A constructive proof of the Kuhn–Tucker multiplier rule. J. Math. Anal. Appl. **26**, 297–306 (1969)
3. Bertsekas, D.P.: Nonlinear Programming. Athena Scientific, Boston (1995)
4. Birbil, S.I., Frenk, J.B.G., Still, G.J.: An elementary proof of the Fritz–John and Karush–Kuhn–Tucker conditions in nonlinear programming. Eur. J. Oper. Res. **180**, 479–484 (2007)
5. Karush, W.: Minima of Functions of Several Variables with Inequalities as Side Conditions. Master's thesis. Department of Mathematics, University of Chicago (1939)
6. Kuhn, H.W., Tucker, A.W.: Nonlinear programming. In: Neyman, J. (ed.) Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, pp. 481–492. University of California Press, Berkeley (1950)
7. Luenberger, D.G.: Introduction to Linear and Nonlinear Programming. Addison-Wesley, Reading (1973)
8. McShane, E.J.: The Lagrange multiplier rule. Am. Math. Monthly **80**, 922–925 (1973)
9. Pourciau, B.H.: Modern multiplier rules. Am. Math. Monthly **87**, 433–452 (1980)
10. Prékopa, A.: On the development of optimization theory. Am. Math. Monthly **87**, 527–542 (1980)
11. Rockafellar, R.T.: Lagrange multipliers and optimality. SIAM Rev. **35**, 183–238 (1993)