Optim Lett (2012) 6:1597–1601
DOI 10.1007/s11590-011-0349-4

ORIGINAL PAPER

A short elementary proof of the Lagrange multiplier theorem

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

Received: 26 August 2010 / Accepted: 23 May 2011 / Published online: 14 June 2011
© Springer-Verlag 2011

Abstract We present a short elementary proof of the Lagrange multiplier theorem for equality-constrained optimization. Most proofs in the literature rely on advanced analysis concepts such as the implicit function theorem, whereas elementary proofs tend to be long and involved. By contrast, our proof uses only basic facts from linear algebra, the definition of differentiability, the critical-point condition for unconstrained minima, and the fact that a continuous function attains its minimum over a closed ball.

Keywords Nonlinear programming · Lagrange multiplier theorem · First-order necessary conditions

In a recent paper [4], we offered a brief, elementary proof of the Karush–Kuhn–Tucker (KKT) conditions for an optimization problem constrained by nonlinear inequalities and linear equalities. In the present article, we complement that work with a short

O. Brezhneva (✉)
Department of Mathematics, Miami University, Oxford, OH 45056, USA
e-mail: brezhnoa@muohio.edu

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

S. E. Wright
Department of Statistics, Miami University, Oxford, OH 45056, USA
e-mail: wrightse@muohio.edu

elementary proof of the classical first-order necessary conditions for local optimality in a nonlinear equality-constrained optimization problem. Specifically, we prove the following.

Lagrange multiplier theorem. Suppose $\bar{x}$ is a local minimizer for

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

where $f : \mathbf{R}^n \to \mathbf{R}$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ are continuously differentiable at $\bar{x}$. If the $m \times n$ Jacobian matrix $g'(\bar{x})$ has full row rank, then there exists a (unique) $\bar{\lambda} \in \mathbf{R}^m$ satisfying

$$f'(\bar{x}) = \bar{\lambda}^T g'(\bar{x}). \tag{2}$$

The traditional proof (e.g., [8]) of this important theorem uses the implicit function theorem, which is not generally considered elementary. Other short proofs make use of deep or advanced results such as the Brouwer fixed-point theorem [7] or the Ekeland variational principle [5], which also allow one to relax the continuous differentiability hypothesis. In an article focusing on inequality constraints, Pourciau [10] proves the multiplier rule for equality-constrained problems by invoking the convex interior mapping theorem, which is again based on the Brouwer fixed-point theorem. A truly elementary proof is given by McShane [9], who relies instead on a penalty reformulation, repeated application of the Bolzano–Weierstrass theorem, and the (Weierstrass) theorem that a continuous function on a closed ball in Euclidean space attains its minimum. Such proofs have since become common (e.g., [2,3,11]), and a penalty-based proof by Beltrami [1] replaces the full-rank assumption by a constant-rank assumption. Moreover, the proofs presented in most of the works cited above also allow for the consideration of inequality constraints, as needed for the more general KKT theorem. Our approach can likewise be embedded in a simple proof of the KKT theorem, as we indicate below.

In contrast to the proofs cited above, our short proof of the Lagrange theorem uses only: (a) the definition of (Fréchet) differentiability; (b) the Fermat theorem that an unconstrained optimum occurs at a critical point; (c) a single application of the Weierstrass theorem; and (d) some basic facts from linear algebra. The proof also requires a straightforward verification of some inequalities, which can be left as a routine exercise. This makes our proof particularly suitable for use in a first undergraduate course in optimization or analysis, and also for 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.

Before giving the proof, we explain the underlying ideas. The proof constructs elements $x \approx \bar{x}$ such that $g(x) = 0$ and $f(x) < f(\bar{x})$, thereby contradicting the minimality of $\bar{x}$. While the idea of utilizing a path through $\bar{x}$ is essentially classical (e.g., see the discussion of *linear solvability* in [6]), our proof of the existence of such a path is new and emphasizes important optimization concepts. Specifically, we choose $x = \bar{x} + th + G\eta(t)$ for all sufficiently small t, where h is in the kernel of $g'(\bar{x}) =: G^T$

and $\eta(t)$ is an absolute minimizer of $\psi(t,\eta) := \|g(\bar{x}+th+G\eta)\|^2$ for each fixed t. The key idea is that the function $\psi(t,\eta)$ grows quadratically with respect to $\eta \approx 0$, so that $\psi(t,\eta) > \psi(t,0)$. Hence, $\eta(t)$ must be an interior point of some ball centered at 0 and is therefore a critical point. Such use of quadratic growth is a common feature of penalty and regularization methods, and also plays an important role in the study of optimal control problems. The novelty of its use in the current proof is that it provides for direct treatment of the equality constraint: the critical-point condition $\psi'_\eta(t,\eta(t)) = 0$ implies $g(\bar{x}+th+G\eta(t)) = 0$ under the assumption that the Jacobian matrix $g'(\bar{x})$ has full row rank.

Proof of the Lagrange multiplier theorem. Assume that $g'(\bar{x})$ has full row rank and that no $\bar{\lambda}$ satisfies (2). Because (2) fails for all $\bar{\lambda}$, the matrix

$$\begin{bmatrix} f'(\bar{x}) \\ g'(\bar{x}) \end{bmatrix}$$

has linearly independent rows; it therefore defines a surjective mapping of $\mathbf{R}^n$ onto $\mathbf{R}^{1+m}$. In particular, there exists a vector $h \in \mathbf{R}^n$ for which $f'(\bar{x})h = -1$ and $g'(\bar{x})$ $h = 0$.

Without loss of generality, we assume $\|h\| = 1 = \|g'(\bar{x})\|$. Let $G = [g'(\bar{x})]^T$ and let $B(y,\beta)$ denote the closed ball of radius β about y. By the full-rank assumption and $\|G\| = 1$, there exists $\gamma \in (0,1]$ so that

$$\gamma\|\eta\| \le \left\|G^T G\eta\right\| \le \|\eta\|, \ \forall \eta \in R^m. \tag{3}$$

Next, for $z \in R^n$ let $\omega(z) = g(\bar{x}+z) - [g(\bar{x}) + g'(\bar{x})z] = g(\bar{x}+z) - G^T z$ denote the remainder in the definition of differentiability of g at $\bar{x}$. In other words, there exist $\epsilon \in (0,1)$ and a strictly increasing function $\rho : [0,\epsilon] \to [0,\infty)$ with $\lim_{\tau\to 0^+} \rho(\tau)/\tau = 0 = \rho(0)$, so that

$$\|\omega(z)\| \le \rho(\|z\|), \ \forall z \in B(0,\epsilon). \tag{4}$$

In particular, we have $\rho(\tau) = o(\tau)$. Moreover, since g is continuously differentiable at $\bar{x}$, it is differentiable on $B(\bar{x},\epsilon)$, with $g'(x)G$ invertible for all $x \in B(\bar{x},\epsilon)$. Now choose $\delta \in (0,\epsilon/2)$ so that

$$\gamma \ge 2\sqrt{\rho(2\tau)/\tau}, \ \forall \tau \in (0,\delta). \tag{5}$$

Next we define $\psi : (-\delta,\delta) \times R^m \to R$ by

$$\psi(t,\eta) = \|g(\bar{x}+th+G\eta)\|^2 = \left\|G^T G\eta + \omega(th+G\eta)\right\|^2, \tag{6}$$

where we use the assumptions that $g(\bar{x}) = g'(\bar{x})h = 0$.

Using (3)–(6) one readily verifies that each η with $\|\eta\| = \sqrt{|t|\rho(2|t|)} > 0$ satisfies

$$\sqrt{\psi(t,\eta)} \geq \gamma\|\eta\| - \rho\left(\|th + G\eta\|\right) \geq \gamma\|\eta\| - \rho\left(|t| + \|\eta\|\right)$$
$$\geq \sqrt{|t|\rho(2|t|)}\left(\gamma - \sqrt{\rho(2|t|)/|t|}\right) \geq \rho(2|t|) > \rho(|t|) \geq \sqrt{\psi(t,0)}.$$

Consequently, an absolute minimizer $\eta(t)$ of $\psi(t,\eta)$ over all $\eta \in B\left(0, \sqrt{|t|\rho(2|t|)}\right)$ is necessarily an interior point of that ball with $\|\eta(t)\| = o(t)$. Therefore, $\eta(t)$ must be a critical point of $\psi(t,\eta)$ for a fixed nonzero t, and hence

$$0 = \psi'_\eta(t,\eta(t)) = 2g(\bar{x} + th + G\eta(t))^T g'(\bar{x} + th + G\eta(t))G.$$

By the invertibility of $g'(x)G$ in a neighborhood of $\bar{x}$, this implies that $g(\bar{x} + th + G\eta(t)) = 0$.

By the assumed differentiability of f at $\bar{x}$, there exists a function $u : R \to R$ such that

$$f(\bar{x} + th + G\eta(t)) = f(\bar{x}) + tf'(\bar{x})h + f'(\bar{x})G\eta(t) + u(t)$$

and $\lim_{t\to 0}|u(t)|/t = 0$. Since $\|\eta(t)\| = o(t)$ and $f'(\bar{x})h = -1$, we have

$$f(\bar{x} + th + G\eta(t)) < f(\bar{x})$$

and $g(\bar{x}+th+G\eta(t)) = 0$ for all nonzero $t \in (-\delta,\delta)$. This contradicts the minimality of $\bar{x}$, thereby proving the theorem. □

One can also handle inequality constraints by extending the above argument in the manner presented by Brezhneva et al. [4]. In this way, one obtains a very simple proof of the KKT theorem under the linear independence constraint qualification. Moreover, if one is willing to invoke linear programming duality or a theorem of the alternative, then a standard argument ([8]) allows one to extend such a proof to cover the more general Mangasarian–Fromovitz constraint qualification. Such an argument is quite a bit simpler and, we believe, more intuitive than the elementary proof of the KKT theorem recently given by Birbil et al. [3].

Acknowledgments The authors thank the anonymous reviewers for the valuable comments and suggestions which helped us improve the presentation of this paper. The work of Alexey A. Tret'yakov is partially supported by the Russian Foundation for Basic Research, Grants # 08-01-00619 and # 11-01-00786-a, and by the Leading Scientific Schools, Grant # 4096.2010.1.

References

1. Beltrami, E.J.: A constructive proof of the Kuhn-Tucker multiplier rule. J. Math. Anal. Appl. **26**, 297–306 (1969)
2. Bertsekas, D.P.: Nonlinear programming. Athena Scientific, Boston (1995)
3. 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. Euro. J. Oper. Res. **180**, 479–484 (2007)

4. Brezhneva, O.A., Tret'yakov, A.A., Wright, S.E.: A simple and elementary proof of the Karush–Kuhn–Tucker theorem for inequality-constrained optimization. Optim. Lett. **3**, 7–10 (2009)
5. Clarke, F.H.: Optimization and nonsmooth analysis. Wiley, New York (1983)
6. Craven, B.D.: Mathematical programming and control theory. Chapman and Hall, London (1978)
7. Halkin, H.: Implicit functions and optimization problems without continuous differentiability of the data. SIAM J. Control **12**, 229–236 (1974)
8. Luenberger, D.G.: Introduction to linear and nonlinear programming. Addison-Wesley, Reading (1973)
9. McShane, E.J.: The Lagrange multiplier rule. Amer. Math. Monthly **80**, 922–925 (1973)
10. Pourciau, B.H.: Modern multiplier rules. Amer. Math. Monthly **87**, 433–452 (1980)
11. Rockafellar, R.T.: Lagrange multipliers and optimality. SIAM Rev. **35**, 183–238 (1993)