Optimization: A Journal of Mathematical Programming and Operations Research
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gopt20

An elementary proof of the Lagrange multiplier theorem in normed linear spaces

Olga Brezhneva [a] & Alexey A. Tret'yakov [b c d]

[a] Department of Mathematics, Miami University, Oxford, OH 45056, USA

[b] Dorodnicyn Computing Center of the Russian Academy of Sciences, Vavilova 40, 119991 Moscow GSP-1, Russia

[c] System Research Institute, Polish Academy of Sciences, Newelska 6, 01-447 Warsaw, Poland

[d] Department of Natural Sciences, Institute of Mathematics and Physics, University of Podlasie in Siedlce, 3 Maja 54, 08-110 Siedlce, Poland

Published online: 08 Aug 2011.

Optimization
Vol. 61, No. 12, December 2012, 1511–1517

An elementary proof of the Lagrange multiplier theorem in normed linear spaces

Olga Brezhneva[a]* and Alexey A. Tret'yakov[bcd]

[a]*Department of Mathematics, Miami University, Oxford, OH 45056, USA;* [b]*Dorodnicyn Computing Center of the Russian Academy of Sciences, Vavilova 40, 119991 Moscow GSP-1, Russia;* [c]*System Research Institute, Polish Academy of Sciences, Newelska 6, 01-447 Warsaw, Poland;* [d]*Department of Natural Sciences, Institute of Mathematics and Physics, University of Podlasie in Siedlce, 3 Maja 54, 08-110 Siedlce, Poland*

(*Received 10 March 2011; final version received 11 June 2011*)

We present an elementary proof of the Lagrange multiplier theorem for optimization problems with equality constraints in normed linear spaces. Most proofs in the literature rely on advanced concepts and results, such as the implicit function theorem and the Lyusternik theorem. By contrast, the proof given in this article employs only basic results from linear algebra, the critical-point condition for unconstrained minima and the fact that a continuous function attains its minimum over a closed ball in the finite-dimensional space.

Keywords: Lagrange multiplier theorem; equality-constrained optimization

AMS Subject Classifications: 90C46; 49K27

1. Introduction

Consider the following problem:

$$\begin{aligned} &\min_{x\in X} f(x) \\ &\text{subject to} \quad g(x) = (g_1(x), \dots, g_m(x)) = 0, \end{aligned} \tag{1}$$

where $f: X \to \mathbb{R}$, $g_i: X \to \mathbb{R}$, $i = 1, \dots, m$ and X is a normed linear space.

In this article we give a new elementary proof of the Lagrange multiplier theorem for problem (1). The proof for the finite-dimensional case is given in [6]. The article complements our recent article [5], where we gave an elementary proof of the Karush–Kuhn–Tucker theorem for optimization problems with inequality constraints in normed linear spaces.

We denote the dual of space X by X^*.

*Corresponding author. Email: brezhnoa@muohio.edu

ISSN 0233–1934 print/ISSN 1029–4945 online

http://dx.doi.org/10.1080/02331934.2011.603323
http://www.tandfonline.com

Theorem 1 (Lagrange multiplier theorem: necessary conditions for optimality) *Suppose that x^* is a local minimizer for problem* (1). *Assume that f and g_i, $i = 1, \dots, m$, are twice Frechet differentiable at x^*, and that the set $\{g_i'(x^*) \mid i = 1, \dots, m\}$ is linearly independent.*

Then there exist unique multipliers $\lambda_i^* \in \mathbb{R}$, $i = 1, \dots, m$, such that

$$f'(x^*) = \sum_{i=1}^{m} \lambda_i^* g_i'(x^*). \tag{2}$$

Traditional proofs (e.g. [1]) of Theorem 1 are based on the implicit function theorem. To our knowledge the following is the simplest approach to prove Theorem 1 using the implicit function theorem.

Assume that (2) does not hold. Then $f'(x^*)$, $g_1'(x^*), \dots, g_m'(x^*)$ are linearly independent. Consider mapping

$$R(\varepsilon, x) = \begin{pmatrix} f(x) - f(x^*) + \varepsilon \\ g(x) \end{pmatrix}, \quad \varepsilon > 0.$$

Notice that $R_x'(0, x^*)$ is onto and so R satisfies the assumptions of the implicit function theorem. Then there exists $x(\varepsilon)$ such that $R(\varepsilon, x(\varepsilon)) = 0$. Hence, $f(x(\varepsilon)) = f(x^*) - \varepsilon$ and $g(x(\varepsilon)) = 0$, which contradicts the assumption that x^* is a local minimizer.

However, the proof of the implicit function theorem is not generally considered elementary. Other proofs of Theorem 1 are based on the penalty function approach [3], elimination of variables [2], separation theorem and some other advanced results. Bhakta and Roychaudhuri [4] give a proof that relies on Farkas lemma, which is based on the 'strict separation axiom'. Ioffe and Tikhomirov [7] give a proof that is based on the Lyusternik theorem, which requires advanced analysis concepts. In contrast to the proofs mentioned above, our short proof of Theorem 1 uses only Fermat's rule, which states that an unconstrained optimum occurs at a critical point, the Weierstrass theorem and some basic facts from linear algebra.

2. Proof of the Lagrange multiplier theorem

We need the following lemma proved in [5].

Lemma 1 *Let X be a normed linear space. Let ξ_i, $i = 1, \dots, n$, be linearly independent elements in X^* for some $n \geq 1$. Then there exists a set of n linearly independent elements of X: $\eta_1, \dots, \eta_n$, such that matrix A_n, defined by*

$$A_n = \begin{bmatrix} \langle \xi_1, \eta_1 \rangle & \dots & \langle \xi_1, \eta_n \rangle \\ \vdots & \ddots & \vdots \\ \langle \xi_n, \eta_1 \rangle & \dots & \langle \xi_n, \eta_n \rangle \end{bmatrix}, \tag{3}$$

is invertible.

Proof (Proof of Theorem 1) Assume on the contrary that (2) does not hold. Then the elements $\xi_1 = g_1'(x^*), \dots, \xi_m = g_m'(x^*)$, $\xi_{m+1} = f'(x^*)$ are linearly independent. Hence, by Lemma 1, there exists a set of $m + 1$ linearly independent elements of X,

$\eta_1, \ldots, \eta_{m+1}$, such that matrix $A_{m+1}(x^*)$ defined in (3) with $n = m+1$ is invertible. Consider the linear system,

$$A_{m+1}(x^*)\begin{bmatrix}\alpha_1\\ \vdots\\ \alpha_m\\ \alpha_{m+1}\end{bmatrix} = \begin{bmatrix}\langle g_1'(x^*), \eta_1\rangle & \ldots & \langle g_1'(x^*), \eta_{m+1}\rangle\\ \vdots & \ddots & \vdots\\ \langle g_m'(x^*), \eta_1\rangle & \ldots & \langle g_m'(x^*), \eta_{m+1}\rangle\\ \langle f'(x^*), \eta_1\rangle & \ldots & \langle f'(x^*), \eta_{m+1}\rangle\end{bmatrix}\begin{bmatrix}\alpha_1\\ \vdots\\ \alpha_m\\ \alpha_{m+1}\end{bmatrix} = \begin{bmatrix}0\\ \vdots\\ 0\\ -1\end{bmatrix}. \quad (4)$$

Since matrix $A_{m+1}(x^*)$ is invertible, system (4) has a solution $(\bar{\alpha}_1, \ldots, \bar{\alpha}_{m+1})$. Define h and $\eta(\alpha)$ in X by

$$h = \bar{\alpha}_1\eta_1 + \cdots + \bar{\alpha}_{m+1}\eta_{m+1}, \quad \eta(\alpha) = \alpha_1\eta_1 + \cdots + \alpha_m\eta_m, \quad (5)$$

where $\alpha = (\alpha_1, \ldots, \alpha_m)$. Note that there is a constant C_1 such that

$$\|\eta(\alpha)\| \le C_1\|\alpha\| \quad (6)$$

for all $\alpha \in \mathbb{R}^m$.

In addition, for any $x, y \in X$, we define $g'(x)$ and $g'(x)y$ by

$$g'(x) = \begin{bmatrix}g_1'(x)\\ \vdots\\ g_m'(x)\end{bmatrix} \quad \text{and} \quad g'(x)\,y = \begin{bmatrix}\langle g_1'(x), y\rangle\\ \vdots\\ \langle g_m'(x), y\rangle\end{bmatrix}.$$

Without loss of generality, we assume $\|g'(x^*)\| = 1$ and $\|h\| = 1$. Then, by (4),

$$g'(x^*)h = 0_m, \quad \langle f'(x^*), h\rangle = -1. \quad (7)$$

Now, define matrix $A_m(x)$ by

$$A_m(x) = \begin{bmatrix}\langle g_1'(x), \eta_1\rangle & \ldots & \langle g_1'(x), \eta_m\rangle\\ \vdots & \ddots & \vdots\\ \langle g_m'(x), \eta_1\rangle & \ldots & \langle g_m'(x), \eta_m\rangle\end{bmatrix}.$$

As follows from the proof of Lemma 1 in [5], all leading principal submatrices of $A_{m+1}(x^*)$, defined in (4), are invertible, and, therefore, matrix $A_m(x^*)$ is invertible.

By the assumption that the set $\{g_i'(x^*) \mid i = 1, \ldots, m\}$ is linearly independent, there exists $\gamma > 0$ such that for every $\alpha \in \mathbb{R}^m$,

$$\|\, g'(x^*)\eta(\alpha)\| \ge \gamma\|\alpha\|. \quad (8)$$

Choose $\epsilon > 0$ so that by Taylor's theorem, there exist a constant $C > 0$ and function $\omega(s)$ such that

$$g(x^* + s) = g(x^*) + g'(x^*)s + \omega(s) \quad (9)$$

and

$$\|\omega(s)\| \le C\|s\|^2 \quad (10)$$

for all s, $\|s\| \le \epsilon$.

Define d and δ as

$$d = \frac{6C}{\gamma}, \quad \delta = \min\left\{\frac{\gamma}{8CC_1}, \frac{\sqrt{\gamma}}{2\sqrt{Cd}C_1}, 1, \frac{\epsilon}{1 + C_1 d}\right\}. \tag{11}$$

Let $t \in (0, \delta)$. Consider $\psi(t,\alpha) = \|g(x^* + th + \eta(\alpha))\|^2$. We will show that the absolute minimizer of $\psi(t,\alpha)$ for a fixed t is an interior point of the ball $B(0, dt^2) = \{\alpha \mid \|\alpha\| \le dt^2\}$.

Using the assumptions $g(x^*) = 0$ and $g'(x^*)h = 0$, we get

$$\psi(t,\alpha) = \| g(x^*) + g'(x^*)(th + \eta(\alpha)) + \omega(th + \eta(\alpha))\|^2 = \| g'(x^*)\eta(\alpha) + \omega(th + \eta(\alpha))\|^2.$$

Inequality (6) with $\|\alpha\| = dt^2$ yields

$$\|\eta(\alpha)\| \le C_1 dt^2. \tag{12}$$

Then using (8)–(12) and $\|h\| = 1$ for each α such that $\|\alpha\| = dt^2$, we get

$$\begin{aligned}
\sqrt{\psi(t,\alpha)} &= \| g'(x^*)\eta(\alpha) + \omega(th + \eta(\alpha))\| \\
&\ge \| g'(x^*)\eta(\alpha)\| - \|\omega(th + \eta(\alpha))\| \\
&\ge \gamma\|\alpha\| - C\|th + \eta(\alpha)\|^2\| \\
&\ge \gamma dt^2 - Ct^2\|h\|^2 - 2Ct\,\|h\|\,\|\eta(\alpha)\| - C\|\eta(\alpha)\|^2 \\
&\ge \gamma\, \mathrm{d}t^2 - Ct^2 - 2Ct\,(C_1 dt^2) - CC_1^2 d^2 t^4 \\
&\ge \gamma dt^2 - \frac{1}{4}\gamma dt^2 - \frac{1}{4}\gamma dt^2 - \frac{1}{4}\gamma dt^2 \\
&= \frac{1}{4}\gamma dt^2.
\end{aligned}$$

At the same time, using (10), (11) and $\|h\| = 1$, we have

$$\sqrt{\psi(t,0)} = \|\omega(th)\| \le Ct^2\|h\|^2 \le Ct^2 < \frac{1}{4}\gamma dt^2 \le \sqrt{\psi(t,\alpha)}.$$

Thus, an absolute minimizer $\alpha(t)$ of $\psi(t,\cdot)$ is an interior point of the ball $B(0, dt^2)$. Therefore, $\alpha(t)$ is a critical point of $\psi(t,\eta)$ for a fixed t with $\|\alpha(t)\| < dt^2$. Hence, using chain rule

$$0 = \psi'_\alpha(t, \alpha(t)) = 2A_m(x^* + th + \eta(\alpha(t)))\, g(x^* + th + \eta(\alpha(t))).$$

Since matrix $A_m(x^*)$ is invertible, we may assume that $A_m(x^* + th + \eta(\alpha(t)))$ is invertible. Then

$$g(x^* + th + \eta(\alpha(t))) = 0. \tag{13}$$

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

$$f(x^* + th + \eta(\alpha(t))) = f(x^*) + t\langle f'(x^*), h\rangle + \langle f'(x^*), \eta(\alpha(t))\rangle + u(t)$$

and $\lim_{t\to 0}|u(t)|/t = 0$. Since $\|\alpha(t)\| < dt^2$, inequality (6) implies $\|\eta(\alpha(t))\| < C_1 dt^2$. Then using $\langle f'(x^*), h\rangle = -1$ we get

$$f(x^* + th + \eta(\alpha(t))) < f(x^*)$$

and $g(x^* + th + \eta(\alpha(t))) = 0$ for all $t \in (0, \delta)$. This contradicts the minimality of x^*, thereby proving the theorem. ■

Remark In the proof of the theorem we used Taylor's theorem in the form (9) twice with $s = th + \eta(\alpha)$ and $s = th$. Then the requirement $\|s\| \le \epsilon$ is equivalent to $\|th + \eta(\alpha)\| \le \epsilon$ and $\|th\| \le \epsilon$. Both inequalities follow from (11), (12) and $\|h\| = 1$.

3. Conclusion

Notice that our consideration subsumes an elementary proof of a special version of an implicit function theorem. Namely, as was shown above, $g(x) = 0$ has a solution $x(t) \in N(x^*)$, $t \in (0, \delta)$, which can be written in the form

$$x(t) = x^* + th + \eta(\alpha(t)), \quad x(t) \in M(x^*) = \{x \in N(x^*) \mid g(x) = 0\}, \tag{14}$$

where $N(x^*)$ is some neighbourhood of x^* in X. The same approach can also be used to prove the classical implicit function theorem.

As we mentioned in the introduction, Lyusternik theorem [7] implies that an element h from the kernel of $g'(x^*)$ is tangent to the set $M(x^*)$. Recall that a vector $h \in X$ is said to be tangent to the set $M(x^*)$ if there exists $\delta > 0$ and a mapping $t \to r(t)$ of the interval $(0, \delta)$ into X such that

$$x(t) = x^* + th + r(t) \in M(x^*) \quad \forall t \in (0, \delta), \quad \lim_{t\to 0} \frac{\|r(t)\|}{t} = 0.$$

The proof of Theorem 1 yields the same property; namely by (7), (12) and (14), an element h from the kernel of $g'(x^*)$ is tangent to the set $M(x^*)$.

The approach presented in the article can also be used in analysis of optimization problems, where feasible points $x_p(t)$ have the form

$$x_p(t) = x^* + th_1 + t^2 h_2 + \cdots + t^p h_p + \omega_p(t), \quad x_p(t) \in M(x^*)$$

and $\|\omega_p(t)\| \le Ct^{p+1}$. We illustrate this by a specific case of $p = 2$ considered in the following lemma.

Lemma 2 *Suppose that $g_i : X \to \mathbb{R}$, $i = 1, \ldots, m$, are three times Frechet differentiable at x^*. Assume that $g(x^*) = (g_1(x^*), \ldots, g_m(x^*)) = 0$ and that the set $\{g_i'(x^*) \mid i = 1, \ldots, m\}$ is linearly independent. Let $h_1 = h$ be defined in* (5) *and*

$$h_2 = -\frac{1}{2}(g'(x^*))^T\big[g'(x^*)(g'(x^*))^T\big]^{-1} g''(x^*)[h_1]^2, \tag{15}$$

where $g''(x^*)[h_1]^2 = g''(x^*)(h_1, h_1)$. Then there exist $\sigma > 0$ and $C_3 > 0$ such that

$$x_2(t) = x^* + th_1 + t^2 h_2 + \omega_2(t), \quad x_2(t) \in M(x^*), \quad \|\omega_2(t)\| \le C_3 t^3, \tag{16}$$

for any $t \in (0, \sigma)$.

Proof Notice that by (7), $g'(x^*)h_1 = 0$. Let $\eta(\alpha)$ be defined by (5) and γ be defined in (8). Choose $\epsilon_2 > 0$ so that by Taylor's theorem, there exist a constant $C_2 > 0$ and function $\bar{\omega}(t, \alpha)$ such that

$$g\big(x^* + th_1 + t^2h_2 + t^3\eta(\alpha)\big) = g(x^*) + tg'(x^*)(h_1 + th_2 + t^2\eta(\alpha)) + \frac{1}{2}t^2g''(x^*)[h_1]^2 + \bar{\omega}(t,\alpha)$$

and $\|\bar{\omega}(t,\alpha)\| \le C_2t^3$ for all $t \in (0, \epsilon_2)$ and all $\|\alpha\| \le \rho$, where

$$\rho = \frac{4C_2}{\gamma}. \tag{17}$$

Consider $t \in (0, \sigma)$, where $\sigma \le \epsilon_2$ is sufficiently small. Introduce

$$\psi_2(t,\alpha) = \| g\big(x^* + th_1 + t^2h_2 + t^3\eta(\alpha)\big)\|^2, \quad \alpha \in \mathbb{R}^m.$$

Similarly to the proof of Theorem 1, for any α with $\|\alpha\| = \rho$, using (8), (15) and (17), we get,

$$\begin{aligned}
\sqrt{\psi_2(t,\alpha)} &= \left\| t^2g'(x^*)h_2 + t^3g'(x^*)\eta(\alpha) + \frac{1}{2}t^2g''(x^*)[h_1]^2 + \bar{\omega}(t,\alpha)\right\| \\
&= \left\| -\frac{1}{2}t^2g''(x^*)[h_1]^2 + t^3g'(x^*)\eta(\alpha) + \frac{1}{2}t^2g''(x^*)[h_1]^2 + \bar{\omega}(t,\alpha)\right\| \\
&\ge \|t^3g'(x^*)\eta(\alpha)\| - \|\bar{\omega}(t,\alpha)\| \\
&\ge t^3\gamma\rho - C_2t^3 \\
&= \frac{3}{4}t^3\gamma\rho.
\end{aligned}$$

At the same time, (17) implies $C_2 = \frac{1}{4}\gamma\rho$ and, hence,

$$\sqrt{\psi_2(t,0)} = \| g(x^* + th_1 + t^2h_2)\| \le \|\bar{\omega}(t,0)\| \le C_2t^3 < \frac{3}{4}t^3\gamma\rho.$$

Thus for every fixed $t \in (0,\sigma)$, there exists an absolute minimizer $\bar{\alpha}(t)$ of $\psi(t,\cdot)$ with $\|\bar{\alpha}(t)\| < \rho$. Notice that by (6), $\|\eta(\bar{\alpha}(t))\| < C_1\rho$. Then, similar to (13), $g(x^* + th_1 + t^2h_2 + t^3\eta(\bar{\alpha})) = 0$ and, therefore, (16) holds with $\omega_2(t) = t^3\eta(\bar{\alpha}(t))$ and $C_3 = C_1\rho$. ■

Acknowledgements

The authors thank the anonymous reviewer for the valuable comments and suggestions which helped us improve the content and presentation of this article. The work of the second author is partially supported by the Russian Foundation for Basic Research, grant no 08-01-00619, and by the Leading Scientific Schools, grant no 4096.2010.1.

References

[1] V.M. Alekseev, V.M. Tikhomirov, and S.V. Fomin, *Optimal Control*, Consultants Bureau, New York, London, 1987.

[2] S.A. Ashmanov and A.V. Timohov, *Optimization Theory in Problems and Exercises*, Nauka, Moscow, Russia, 1991 (in Russian).
[3] D.P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1999.
[4] P.C. Bhakta and S. Roychaudhuri, *Optimization in normed linear spaces*, J. Math Anal. Appl. 134 (1988), pp. 460–470.
[5] O. Brezhneva and A.A. Tret'yakov, *An elementary proof of the Karush–Kuhn–Tucker theorem in normed linear spaces for problems with a finite number of inequality constraints*, Optimization 60 (2011), pp. 613–618.
[6] O. Brezhneva, A.A. Tret'yakov, and S.E. Wright, *A short elementary proof of the Lagrange multiplier theorem*, Optim. Lett. (2011), DOI: 10.1007/s11590-011-0349-4. Available at http://www.springerlink.com/content/v2504h68v5246481/
[7] A.D. Ioffe and V.M. Tikhomirov, *Theory of Extremal Problems*, North-Holland, Amsterdam, The Netherlands, 1979.