A new proof of the Lagrange multiplier rule

Jan Brinkhuis and Vladimir Protassov

EI2015-39

Abstract. We present an elementary self-contained proof for the Lagrange multiplier rule. It does not refer to any substantial preparations and it is only based on the observation that a certain limit is positive. At the end of this note, the power of the Lagrange multiplier rule is analyzed.

Keywords. Lagrange multiplier rule; penalty function.

We prove the following.

Theorem (the Lagrange multiplier rule). *Assume that G is a set in $\mathbb{R}^n$ and that $f_i, 0 \leq i \leq m$ are continuously differentiable functions on G. Let $\widehat{x}$ be an interior point of G that is a minimizer of $f_0(x)$ on the set of of points in G for which $f_i(x) = 0, 1 \leq i \leq m$. Then the set of derivatives $f_i'(\widehat{x}), 0 \leq i \leq m$, the n-dimensional row vectors of the partial derivatives of the $f_i, 0 \leq i \leq m$ in $\widehat{x}$, is linearly dependent.*

The given formulation of the conclusion of the rule is equivalent to the following one, which is more familiar: if the derivatives $f_i'(\widehat{x}), 1 \leq i \leq m$ are linearly independent, then $\widehat{x}$ is a stationary point of the function $f_0(x)+\lambda_1 f_1(x)+\cdots+\lambda_m f_m(x)$ for suitable numbers (Lagrange multipliers) $\lambda_i, 1 \leq i \leq m$.

The existing proofs of the multiplier rule in the literature require effort. Either they refer to substantial preparations or they contain technical arguments. Such preparations are the implicit function theorem ([1],[2]), the inverse function theorem (of which the multiplier rule is an immediate consequence [3]), the tangent space theorem ([4], [5]), the Brouwer fixed-point theorem ([6]) or the Ekeland variational principle ([7]). Technical arguments are needed to establish inequalities and estimates ([8]). Then the resulting proofs are elementary but they tend to be relatively long and involved ([9], 10]). The new proof given in the present note requires essentially no effort: it is not only elementary, but also simple, short and insightful. Its novelty consists in the easy observations that a certain limit is positive and that this implies that boundary points are excluded as minimizers of a certain penalty function. The remainder of the proof is obvious. All elementary proofs for the multiplier rule, including the one given in this note, make use of the theorem that a continuous function assumes its minimal value over a closed ball.

Proof. By contraposition.

- *Preparations.* Assume that $\widehat{x}$ is an interior point of G that is a solution of the system of equations $f_i(x) = 0, 1 \leq i \leq m$ for which $f_i'(\widehat{x}), 0 \leq i \leq m$ are linearly independent. We will show that $\widehat{x}$ is not a minimum of the function $f_0(x)$ on the set of $x \in G$ for which $f_i(x) = 0, 1 \leq i \leq m$.

  Without loss of generality it is assumed that $f_0(\widehat{x}) = 0$, $\widehat{x} = 0$ and $n = m + 1$—the latter can be achieved by adding if necessary constraints $a_i x = 0, i \in I$, chosen in such a way that the set

$a_i, i \in I$ completes the linearly independent set $f_i'(0), 0 \le i \le m$ to a basis of the space $(\mathbb{R}^n)^T$ of n-dimensional row vectors. Define $F(x) = (f_0(x), \ldots, f_m(x))^T$ for $x \in \mathbb{R}^n$. Then $F(0) = 0$, $F'(x)$ is the $n \times n$-matrix that has as rows the derivatives $f_i'(x), 0 \le i \le m = n-1$, F' is continuous and

$$F(x) = F'(0)x + o(|x|)(x \to 0) \tag{1}$$

('Landau little o')—as the $f_i, 0 \le i \le m$ are continuously differentiable and $F(0) = 0$. The assumption that $f_i'(0), 0 \le i \le m = n-1$ are linearly independent means that the $n \times n$-matrix $F'(0)$ has rank $m+1 = n$, so $F'(0)$ is invertible. Choose for each $k \in \mathbb{N}$ a minimizer x_k of the penalty function

$$p_k(x) = |F(x) + (1/k^2, 0, \ldots, 0)^T| \tag{2}$$

on the closed ball $|x| \le 1/k$, where $|\cdot|$ is the Euclidean norm.

- *Exclusion of boundary minimizers for the penalty function.* We consider the limit

$$L = \lim_{x \to 0} \frac{|F(x) + (|x|^2, 0, \ldots, 0)^T| - |(|x|^2, 0, \ldots, 0)^T|}{|F'(0)x|}. \tag{3}$$

Here we recall that $F'(0)$ is invertible. Note that

$$\lim_{x \to 0} |x|^2/|F'(0)x| = 0.$$

Therefore,

$$L = \lim_{x \to 0} |F(x)|/|F'(0)x| = 1 > 0, \tag{4}$$

using (1) and the invertibility of $F'(0)$. It follows from (3) and (4) that for k sufficiently large and for each point $\bar{x}$ on the sphere $|x| = 1/k$, one has

$$|F(\bar{x}) + (|\bar{x}|^2, 0, \ldots, 0)^T| > |(|\bar{x}|^2, 0, \ldots, 0)^T|,$$

that is, the penalty function p_k, given by (2), takes a larger value in $\bar{x}$ than in 0—here we use also $F(0) = 0$. Therefore, boundary points of the ball $|x| \le 1/k$ cannot be minimizers of the penalty function p_k, given by (2), on this ball. That is, x_k is an interior point of this ball, $|x_k| < 1/k$.

- *Conclusion of the proof.* Moreover, for k sufficiently large, $|x_k| < 1/k$ implies that the matrix $F'(x_k)$ is invertible as $F'(0)$ is invertible and F' is continuous. From this, and as the minimizer x_k is an interior point of the closed ball $|x| \le 1/k$, it is obvious that

$$F(x_k) + (1/k^2, 0, \ldots, 0)^T = 0. \tag{5}$$

Indeed, otherwise a local decrease of the penalty function p_k, given by (2), would be possible. Here is a precise analytic verification. The stationarity vector equation

$$0 = \frac{d}{dx}|F(x) + (1/k^2, 0, \ldots, 0)^T|^2|_{x=x_k} = 2(F(x_k)^T + (1/k^2, 0, \ldots, 0))F'(x_k)$$

holds as the minimizer x_k is interior. Here we use that the minimizers of a nonnegative valued function do not change if the objective function is squared and that $|v|^2 = v^T v$ for a column vector v. This gives (5) as $F'(x_k)$ is invertible. That is, by the definition of F,

$$f_0(x_k) = -1/k^2 < 0 = f_0(0), f_i(x_k) = 0, 1 \leq i \leq m.$$

It follows that $\widehat{x} = 0$ is not a minimizer of the function $f_0(x)$ on the set of points in G for which $f_i(x) = 0, 1 \leq i \leq m$. This concludes the proof of the theorem.

Comment on the power of the rule. The power of the rule lies in the reversal of the natural order of two main tasks, elimination and differentiation. The natural order would be to eliminate m of the variables $x_1, \ldots, x_n$, using the equality constraints $f_i(x) = 0, 1 \leq i \leq m$, substitute into the objective function $f_0(x)$, and then put the partial derivatives of the resulting expression in $n - m$ variables equal to zero. The elimination task is often a nonlinear problem that is difficult or even impossible. Differentiating first gives the following statement, which is a reformulation of the conclusion of the theorem: all solutions $h \in \mathbb{R}^n$ of the system of linear equations

$$f_i'(\widehat{x})h = 0, 1 \leq i \leq m \tag{6}$$

satisfy

$$f_0'(\widehat{x})h = 0,$$

provided the vectors $f_i'(\widehat{x}), 1 \leq i \leq m$ are linearly independent. The remaining elimination task is now just a *linear* problem. It is easy to eliminate m of the n variables $h_1, \ldots, h_n$ using the system of linear equations (6). Then one can substitute in $f_0'(\widehat{x})h$. Finally one should put the coefficients of the resulting expression in $n-m$ variables equal to zero. This completes elimination without multipliers. In particular, multipliers are not essential to the power of the rule, but merely provide a convenient way to carry out the linear elimination task. See Ch 3 of [4] for more details on the power of the rule.

Acknowledgment. We would like to thank Jan van de Craats for his suggestions on the presentation.

References.

[1] Lang, S., *Calculus of Several Variables*, Reading, MA: Addison-Wesley, (1973)

[2] Luenberger, D.G., Ye Y., *Linear and Nonlinear Programming*, Springer (2008)

[3] Alexeev, V.M., Tikhomirov, V.M., Fomin, S.V., *Optimal Control*, Consultants Bureau, New York (1987)

[4] Brinkhuis, J., Tikhomirov, V.M., *Optimization: Insights and Applications*, Princeton University Press, (2005)

[5] Giaquinta, M., Modica, G., *Mathematical Analysis, An Introduction to Functions of Several Variables*, Birkhäuser, (2009)

[6] Halkin, H., *Implicit functions and optimization problems without continuous differentiability of the data*, SIAM J. Control 12, 229-236 (1974)

[7] Clarke, F.H., *Optimization and nonsmooth analysis*, Wiley (1983)

[8] Brezhneva, O., Tret'vakov, A.A., Wright, S.E., *A short elementary proof of the Lagrange multiplier theorem*, Optim. Lett. (2012) 6:1597-1601.

[9] McShane, E.J., *The Lagrange multiplier rule*, Amer. Math. Monthly 80, 922-925 (1973)

[10] Rockafellar, R.T., *Lagrange multipliers and optimality*, SIAM Rev. 35, 183-238 (1993)