STATE UNIVERSITY OF NEW YORK AT STONY BROOK

COLLEGE OF ENGINEERING

REPORT NO. 110

A CONSTRUCTIVE PROOF OF THE KUHN-TUCKER MULTIPLIER RULE

by

E. J. Beltrami

Dept. Applied Analysis

SUNY, Stony Brook

MAY 1968

A CONSTRUCTIVE PROOF OF THE KUHN-TUCKER MULTIPLIER RULE

by

E. J. Beltrami
Department of Applied Analysis
University of the State of New York
at Stony Brook
Stony Brook, New York

The report was presented as a paper at the SIAM 1968 National Meeting in Toronto, Canada, June 11-14.

May 1968

Introduction

One way to approach constrained minimization problems on E^n and, in particular, nonlinear programs is to append increasingly large cost or penalty terms to the objective function in such a way that the minima of the augmented but unconstrained functions converge to the constrained minimum in the limit. In this paper we discuss the validity of this penalty concept, due to R. Courant [1], and then apply it to obtain constructive proofs of the Kuhn-Tucker and Lagrange multiplier rules.

In section 2 we establish the penalty argument and show by several examples that this approach allows one to obtain the minima of constrained problems even when the multiplier rules are invalid. This is followed, in section 3, by a proof of the appealing fact that rank is lower semi-continuous on the space of m by n matrices. We use this result in section 4 to establish that if the rank of the Jacobian matrix G of constraints is invariant in some ball about the minimum then the Lagrange multiplier rule is applicable. We then show that this statement includes the classical cases where G is of maximal rank or the constraints are linear. The proof depends on the penalty argument and avoids the use of the implicit function theorem. In section 5 the Kuhn-Tucker rule is established by passing to the limit with the necessary conditions for unconstrained problems. The limiting argument is shown to hold under a suitable regularity assumption without appealing to the usual procedure involving Farkas lemma. One virtue of this constructive proof is that the multipliers are explicitly obtained as limits of certain quantities in a natural way. If we let G denote the Jacobian matrix of active constraints at the minimum then the regularity assumption is satisfied if G is of maximal rank, or if the constraints are linear, or if the feasible set is non-empty and the constraints convex. The proof of this is given in section 6.

This paper extends the arguments used in an earlier note [2] where we simply gave some numerical examples of how the multiplier rule can be verified computationally or vindicated in practice by using the penalty approach. This constructive approach suggests a computationally feasible algorithm for solving nonlinear programs and, in fact, an iterative procedure for doing this was devised by Kelley, et. al. [3]. Their algorithm has been implemented numerically to obtain explicit solutions to non-convex programs.

I wish to thank R. Duffin and W. Anderson who, during their stay in Stony Brook, added to the content of this paper by constructive remarks and examples. In particular the proof of lemma 1 is due to W. Anderson.

2. A Penalty Argument

The problem considered in this paper is to minimize f subject to m constraints $g_j = 0$, $0 \leq j \leq s$, and $g_j \leq 0$, $s < j \leq m$, where m is unrestricted and $s < n$. The functions f, g_j are real valued and defined on E^n.

All constraints will be written as equality constraints. In order to do this we define u_j by $u_j(x) = 1$ whenever $g_j(x)>0$, for $j > s$, and by $u_j(x) = 1$ otherwise; here $x \epsilon E^n$. We then observe that the original m constraints are satisfied if and only if the m equality constraints $g_j^2 u_j = g_j^2 u_j^2 = 0$ hold. Moreover, if the m vector g with components g_j belongs to C^1 then so does $(gu)^2$ where gu is the m vector with components $g_j u_j$.

Now let K be an m by m diagonal matrix with positive diagonal elements k_j. Then $K_n \to \infty$ means that all entries $k_{n,j}$ in K_n increase without bound as $n \to \infty$. The quadratic form $\sum_{j \leq m} k_{n,j}(g_j u_j)^2$ will be denoted by $(gu, K_n gu)$ and we define an augmented objective function f_n by $f + \frac{1}{2}(gu, K_n\, gu)$ for each n.

The sense of the theorem below is that the minima of the unconstrained f_n tend to the minimum of the constrained f as $n \to \infty$ since $(gu, K_n\, gu)$ necessarily tends to zero as the cost of violating the constraints increases without bound. For this reason K is called a matrix of penalty constants.

Our main result in this section is based on a theorem given in the NYU notes [4] and in a paper by Butler-Martin [5]. An extension valid for convex functionals on a Hilbert space is given in [6]. Another variant of the penalty argument on E^n is to be found in several papers by Fiacco-McCormick (e.q., [7].)

<u>Theorem 1</u> Let f, g_j^2 be lower semi-continuous (l.s.c.) on a closed set Ωo in E^n and suppose that either Ωo is bounded or that $f(x) \to +\infty$ as $\|x\| \to \infty$. If Ωo has a non-empty intersection with the feasible set Ω_1 defined by $\{x | (gu)^2 = 0\}$ then, for every sequence $K_n \to \infty$, there exists a corresponding sequence x_n which minimizes f_n on Ωo and such that $x_{n_\nu} \to x^o$ for some subsequence. The point x^o is a minimum of f on $\Omega o \cap \Omega_1$ and $f(x_n) \to f(x^o)$.

The penalty argument has the defect that it may yield fictitious solutions when the problem is ill-posed. To illustrate consider the problem of minimizing distance from the origin in the plane subject to the linear constraints $x + y - 1 = 0$ and $x + y - 2 = 0$. The problem clearly does not possess a solution but, acting in ignorance, let as form the augmented function $f_n = x^2 + y^2 + n[(x + y - 1)^2 + (x + y - 2)^2]$. Here the penalty matrix consists of diagonal entries n. The unconstrained minimum of f_n is found by setting ∇f_n to zero from which we obtain $x_n = y_n = 3n / 2(1 + 2n) \to 3/4$ as $n \to \infty$. Thus $f_n \to \infty$ but $f(x_n, y_n) \to f(3/4, 3/4)$.

We mentioned above that Fiacco and McCormick have devised a penalty approach to optimization. Their method generates a sequence of points which lie within the feasible set for inequality constraints. By contrast with their "interior" method the Courant method is an "outside" technique for it may be shown that either the approximations x_n terminate after a finite number of steps or at least one inequality constraint is violated at each iterate.

In the following sections we give examples of how the penalty argument obtains a solution to properly formulated problems even if the multiplier rules fail.

3. A Lemma Concerning Rank

We begin with the following result whose proof is due to W. Anderson (1967, unpublished).

Lemma 1 All n by m matrices of rank $\leq$ r form a closed set in the norm topology.

Proof Let A be an n x m matrix of rank $\geq$ r. The generalized inverse A^+ exists and if B lies in the ball $\| B - A \| < 1 / \| A^+ \|$ then from $B = A + (B - A)$ we obtain $BA^+ = AA^+ + (B - A)A^+ = P_{\mathcal{R}(A)} + (B - A) A^+$ (for definition and properties of generalized inverse see, for example, [10]). Thus $BA^+ = \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix} + \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$

and $\| \delta \| \leq \| (B - A) A^+ \| < 1$ so that $I + \delta$ is invertible. Hence

$$BA^+ \begin{pmatrix} 0 & 0 \\ 0 & (I+\delta)^{-1} \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & (I+\delta) \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & (I+\delta)^{-1} \end{pmatrix} = \begin{pmatrix} 0 & \beta(I+\delta)^{-1} \\ 0 & I \end{pmatrix}$$

and since the right side has rank $\geq$ r, B must also. It follows that matrices of rank $\geq$ r form an open set since a similar open ball can be formed for each such A. Matrices of rank $\leq r$ are then a closed set.

Another proof of this lemma, which I believe is due to R. Duffin is that if $A_n \to A$ in norm where rank $A_n \leq r$ then all minors of A_n of order $> r$ have zero determinant. Each such minor converges to something whose determinant must then also be zero. Hence rank $A \leq r$ and we have a closed set.

If we are given a function f on a subset of E^n then f is lower semi-continuous if and only if $\{x \mid f(x) \leq \alpha\}$ is closed for all α (see, for example, [8], pg. 40). Hence Lemma 1 is equivalent to asserting that rank is l.s.c. on the class of n by m matrices. We will use this result in the following form.

Lemma 2 If $A_n \to A$ then rank $A \leq$ rank A_n for all large n.

I recently became aware that P. R.Halmos has also independently established the lower semi-continuity of rank in order to prove that reducible operators on E^n form a closed set. His proof will appear in an appendix to a forthcoming paper on irreducible operators.

4. Lagrange Multipliers

In this section we restrict ourselves to equality constraints (the case $s = m < n$ and $u \equiv 1$). Suppose that x^o is a local minimum of f on E^n subject to the m constraints $g = 0$ in an open region where f, g are C^1 functions. There is a closed ball Ωo about x^o contained within this region such that x^o is a global minimum on Ωo. Since the constraints are assumed to be satisfied it follows that $\Omega = \Omega o \cap \Omega_1$ is non-empty where, as before, Ω_1 is the feasible set $\{x \mid g^s = 0\}$.

One may safely regard x^0 as the unique global minimum on Ω for, as we show in the next section, there is no loss in generality in our assuming this.

Let G denote the n by m Jacobian matrix associated with the vector valued mapping g, and denote by G_f the n by (m+1) matrix which augments G by adding the column vector ∇f. Then G_f consists of m + 1 columns $\nabla f, \nabla g_1, \ldots, \nabla g_m$. By suitable rearrangement of columns we can always assume that if G is of rank r then the first r columns of G are linearly independent. We now prove a slight extension of the usual Lagrange multiplier argument.

Theorem 2 Let f, g_j be C^1 functions in an open set containing a local constrained minimum x^0 of f. If there is an open ball about x^0 in which rank G is invariant then there exists multipliers λ_j for which

$$\nabla f(x^0) = \sum_{j\leq r} \lambda_j \nabla g_j (x^0) \tag{1}$$

where r is the rank of G.

Proof Let f_n denote the augmented but unconstrained objective functions described in section 2. By theorem 1 there exists a sequence x_n which tends to x^0 (since we can assume that x^0 is the unique global minimum) and such that f_n is minimized on Ωo by x_n. For large enough n the sequence of minimizing points is interior to Ωo and so

$$\nabla f_n = \nabla f + \sum_{j\leq m} k_{n,j} g_j \nabla g_j = 0 \tag{2}$$

at x_n. Since f, g are C^1, G and G_f at x_n tend to G, G_f at x^0. By hypothesis rank G is r for large n and we can rewrite (2) as

$$\nabla f(x_n) = \sum_{j\leq r} \mu_{n,j} \nabla g_j (x_n) \tag{3}$$

for suitable scalars $\mu_{n,j}$. Hence rank G_f is also r for such n from which it follows, using Lemma 2, that $r = \text{rank } G(x^o) \leq \text{rank } G_f(x^o) \leq \text{rank } G_f(x_n) = r$, or ∇f is dependent on $\nabla g_1, \ldots, \nabla g_r$ at x^o. This proves the theorem.

The multiplier rule includes the usual well known cases, as we prove in the next result, which is a corollary of the above theorem.

<u>Theorem 3</u> If G is of maximal rank m at x^o or if the constraints are linear then the rank of G is invariant in some ball about x^o and the result of theorem 2 continues to hold.

<u>Proof</u> If G is of maximal rank the theorem follows from the fact that matrices of rank $\leq m-1$ are closed, by Lemma 1, so that those of rank m are open. If the constraints are linear then G is constant and so invariance of rank is immediate in this case.

Note that if rank G is maximal then the multipliers λ_j are uniquely determined since $\nabla g_1, \ldots, \nabla g_m$ forms a basis in E^m, but if rank $G < m$ then the multipliers are not uniquely given.

The next example illustrates how the penalty argument can find a minimum even when the multiplier rule breaks down. Consider the problem of minimizing distance from the origin in E^2 subject to the constraint $g(x,y) = y^2 - (x-1)^3 = 0$. An inspection of the graph of g shows that the global minimum is attained at (1,0). However $\nabla g = \begin{pmatrix} -3(x-1)^2 \\ 2y \end{pmatrix}$ is zero at (1,0) and non-zero in a neighborhood of this point. Hence rank G does not satisfy the conditions of theorem 2 and the multiplier rule is not expected to hold. In fact any attempt to find the constrained minimum using the multiplier rule quickly leads to an unresolvable difficulty, as the reader may verify. However we proceed by means of the penalty argument and form the augmented functions $f_n(x,y) = x^2 + y^2 + n(y^2 - (x-1)^3)^2$ where the penalty matrix is simply the scalar n in this case. Set ∇f_n to zero and solve to obtain points (x_n, y_n) with $y_n = 0$ and with x_n tending to 1 as $n \to \infty$. The constrained minimum is thus obtained in the limit as constraint violations are increasingly penalized. In the next section another example of the same kind is discussed for problems with inequality constraints.

The Kuhn-Tucker Rule

The proof given in the previous section resulted in a multiplier rule because of a dependence argument (rank G = rank G_f at x^0). Now, however, we wish to use a more constructive argument to derive a general result which will include theorem 2 as a special case. The reason for a different proof is that in case of inequality constraints we are not content with merely showing that multipliers exist but wish to establish them as limiting values of certain non-negative quantities.

We need some additional notation. Let J denote the set of indices j corresponding to active or binding inequality constraints. We let Ω_j be the open set $\{x \mid g_j(x) < 0, j \in J\}$. The set Ω_j may be empty but when it is non-empty it will be possible to conclude a useful result (see theorem 5).

For notational simplicity we restrict ourselves in the next theorem to the case of inequality constraints (s=0) and indicate later how to treat the cases $s > 0$.

We define G, G_f in a slightly different way then in section 3. The matrix G is the n by ℓ Jacobian of constraints for which $j \in J$ and G_f is G augmented by the column vector ∇f. The next theorem was first given by Kuhn-Tucker [9], using a different regularity assumption.

Theorem 4 Let F, g_j be C^1 functions on an open set in E^n containing a local constrained minimum x^o of f. If the constraints satisfy the regularity assumption $(\nabla g_j, h) < 0$ for some h in E^n, for all $j \in J$, then there exists multipliers $\lambda_j \geq 0$ for which

$$\nabla f(x^o) = - \sum_{j \in J} \lambda_j \nabla g_j(x^o) \tag{4}$$

If rank G is maximal at x^o the multipliers are uniquely determined.

Proof Let Ωo be a closed ball about the constrained minimum x^o contained within the region in which f, g_j are C^1 and for which x^o is a global minimum on Ωo. If Ω_1 is the feasible set $\{x \mid g^2 u = 0\}$ then $\Omega = \Omega o \cap \Omega_1$ is non-empty. Replace f by the objective function $f(x) + \| x - x^o \|^2 = f^o(x)$ so that f^o has a unique global minimum on Ωo. If f_n are the augmented but unconstrained functions corresponding to f^o then, by theorem 1, there exists a sequence x_n which minimize f_n on Ωo and such that x_n tend to x^o. For large n the x_n are interior to Ωo and so $\nabla f_n(x_n) = 0$ or

$$\nabla f^o(x_n) = - \sum_{j \leq m} \lambda_{n,j} \nabla g_j(x_n) \tag{5}$$

where $\lambda_{n,j} = k_{n,j} \, g_j(x_n) \, u_j(x_n) \geq 0$. When n is large enough $\lambda_{n,j} = 0$ for $j \notin J$ since x^o belongs to the open set $\{x \mid g_j(x) < 0, j \notin J\}$ and $x_n \to x^o$. Hence (5) can be written as a sum over $j \in J$ for $n \geq N$. By the regularity assumption and because the ∇g_j are continuous it follows that $(\nabla g_j(x_n), h) \leq - \delta < 0$ for some $\delta > 0$ when n is large, since $x_n \to x^o$. Moreover $(\nabla f, h)$ is bounded by some constant M on the closed ball Ωo since ∇f^o is also continuous. From (5) we obtain $M \geq (\nabla f, h) \geq \delta \sum_{j \in J} \lambda_{n,j} \geq \delta \lambda_{n,j}$. Hence the $\lambda_{n,j}$ are bounded and so there exists a subsequence, also denoted by $\lambda_{n,j}$, such that $\lambda_{n,j} \to \lambda_j \geq 0$ as $n \to \infty$. Since f^o, g are C^1, a passage to the limit as $x_n \to x^o$ shows that $\nabla f^o(x^o) = \nabla f(x^o) = -\sum_{j \in J} \lambda_j \nabla g_j (x^o)$. If rank G is maximal at x^o the ∇g_j form a basis in E^{ℓ} and so the multipliers are unique.

In order to obtain a version of theorem 4 when $s > 0$ it is merely necessary to insist that the regularity assumption hold not only for $j \in J$ (active inequality constraints) but for equality constraints as well. However in the latter case it is no longer true in general that the multiplier approximations $\lambda_{n,j}$ be non-negative and so the boundedness of $\lambda_{n,j}$ must be established in a slightly different manner. What we do, in fact, is to write all s equality constraints as inequalities by introducing 2s constraints $\hat{g}_j \leq 0$ defined by $\hat{g}_j = g_j$, $j \leq s$, and $\hat{g}_{j+s} = -g_j$, $j \leq s$. Since $\hat{g}_j u_j \geq 0$ we can now proceed as in theorem 4. Details are left to the reader.

In the general situation in which $s > 0$ the matrix G becomes an n by $\ell + s \leq m$ Jacobian of constraints for which $j \leq s$ and $j \in J$ (note that when $j \in J$ then $j > s$).

In the next section we investigate to what extent the regularity assumption can be expected to hold. For the moment we want to illustrate the extent of the multiplier rule vis à vis the penalty argument. Consider the problem of minimizing $f(x,y) = x$, subject to the constraints $g_1(x,y) = (y-2) - (3-x)^3 \leq 0$ and $g_2(x,y) = -(y-2) - (3-x)^3 \leq 0$. It is not hard to determine that the minimum exists at (3,2) and that if $h = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$ is an arbitrary vector in E^2 then $(\nabla g_1, h) = -h_2$ and $(\nabla g_2, h) = h_2$ at (3,2) which implies that $h_2 = 0$. Thus the regularity assumption is violated and, in fact, the multiplier rule is not valid since $\nabla f = \binom{1}{0}$ and so $(\nabla f, h) = h_1 < 0$ for suitable h. But then, should (4) hold we arrive at a contradiction since $0 > (\nabla f, h) = -\Sigma \lambda_j (\nabla g_j, h) \geq 0$ at x^o when $\lambda_j \geq 0$. The penalty argument yields a solution, however, as in the example of section 3. In fact let f_n be the augmented objective function $x + n\{[(y-2)]\}-(3-x)^3] u_1(x) + [(y-2) + (3-x)^3] u_2(x)\}$ where u_1, u_2 are zero when the constraints g_1, g_2 are satisfied. Since $\nabla f_n = 0$ we obtain (x_n, y_n) given by $\left(3 - \frac{1}{\sqrt[5]{2n}}, 2\right)$ which tends to (3,2), as we had hoped it would.

The Regularity Assumption

It is a worthwhile task to isolate the several important cases in which the regularity hypothesis or constraint qualification $(\nabla g_j, h) < 0$ is satisfied. The next two theorems do this for us.

Theorem 5 If the set Ω_j is non-emtpy and the constraints g_j convex for $j \in J$ then there exists some h in E^n for which $(\nabla g_j, h) < 0$ at x^o. Hence the regularity assumption is satisfied for $j \in J$ and so theorem 4 holds for $s = 0$. On the other hand, if there exists some open ball about x^o in which rank G is invariant then the regularity assumption holds for all those $j \leq s$ and $j \in J$ for which $\{\nabla g_j\}$ form a maximal linearly independent set at x^o. In the latter case the multiplier rule of theorem 4 remains true for all $s > 0$.

Proof With Ω_j non-empty and g_j convex let $h = x - x^o$ for any $x \in \Omega_j$. Then $(\nabla g_j (x^o), h) \leq g_j(x) - g_j(x^o) < 0$ for $j \in J$. In the case where rank G is invariant in some ball we are assured that rank $G(x_n) = r$ for some fixed $r \leq \ell + s$ and for all n large enough. Let $G_r(x^o)$ designate an r by n submatrix of $G(x^o)$ having rank r and consisting of r linearly independent columns $\nabla g_j(x^o)$. Then we know that for all large n, rank G_r = rank G = r continues to hold since, by lemma 1, matrices of maximal rank form an open set. But equation (5) then tells us that rank $G_f(x_n)$ is also r for all such large n and hence the sum in (5) is extended over the ∇g_j belonging to G_r. Without loss of generality we may denote the columns of G_r by $\nabla g_1, \ldots, \nabla g_r$ and since they form a basis in E^r at x^o there exists a dual basis $\nabla\tilde{g}_j(x^o)$ for which $(\nabla g_j, \nabla\tilde{g}_i) = \delta_{ij}$. Now let $h = -\sum_{j\leq r} \nabla\tilde{g}_j$; then $(\nabla g_j(x^o), h) = -1 < 0$ so that the regularity assumption holds for those $j \in J$ and $j \leq s$ for which ∇g_j belongs to G_r. Since the sum in (5) is extended over precisely this subset of j values for all large n the proof of theorem 4 can be completed as before.

As in section 3 we can state the following special case of theorem 5, proof of which is identical to that of theorem 3.

Theorem 6 If rank $G(x^o)$ is maximal or if the constraints are linear then the rank of G is invariant in some ball about x^o and the result of theorem 4 continues to hold.

Because of theorem 5 the Lagrange multiplier of section 3 rule is obtained from the Kuhn-Tucker theorem by letting s = m. We note, incidentally, that for linear constraints the proof of the multiplier rule follows more directly from the fact that the linear manifold generated by the gradients of active constraints is closed and does not vary with n. Hence since $\nabla f(x_n)$ belongs to this manifold so does the limit $\nabla f(x^o)$. This suffices to prove theorem 2. For inequality constraints we can use a similar argument to prove the Kuhn-Tucker rule with $\lambda_j \geq 0$ since the cone generated by the ∇g_j is also closed.

In summary, a constructive proof of the multiplier rule has been given provided that the constraint qualification $(\nabla g, h) < 0$ holds at x^o. Moreover the constraint qualification is satisfied whenever rank G is invariant in some ball about x^o. This last fact will hopefully sharpen the applicability of the Kuhn-Tucker theorem.

References

1. Courant, R. "Variational Methods for the Solution of Problems of Equilibrium and Vibration." Bull. AMS 49, pp. 1-23, 1943.

2. Beltrami, E. J. "A Computational Approach to Necessary Conditions in Mathematical Programming." ICC Bull. 6, pp. 267-275, 1967.

3. Kelley, H. J.; W. Denham; I. Johnson; P. Wheatley, "An Accelerated Gradient Method for Parameter Optimization with Non-linear Constraints." J.A.S. 13, pp. 166-169, 1966.

4. Courant, R. "Calculus of Variations and Supplementary Notes." NYU lecture notes, 1956-7.

5. Butler, T.; A. Martin, "A Method of Courant for Minimizing Functionals." J. Math. Phy. 41, pp. 291-299, 1962.

6. Beltrami, E. J. "On Infinite Dimensional Convex Programs." J. Computer & System Sci., to appear in 1968.

7. Fiacco, A. V.; G. McCormick, "Computational Algorithm for the Sequential Unconstrained Minimization Technique for Non-linear Programming." Management Science 10, pp. 601-617, 1964.

8. Goldstein, A. A., "Constructive Real Analysis." Harper & Row (N.Y.) 1967.

9. Kuhn, H. W.; Tucker, A. W., "Nonlinear Programming." in Proceedings of 2nd Berkeley Symposium in Mathematical Statistics, Univ. of Calif., Press, 1961, pp. 481-492.

10. Ben Israel, A.; A. Charnes, "Contribution to the Theory of Generalized Inverses." SIAM J. 11, pp. 667-697, 1963.