1 *Harvard University. Arrow's work was sponsored in part by the Office of Naval Research under grant number N00014-67A-0298-0019 (NR047-004).*

2 *University of North Carolina at Chapel Hill. The work of Gould and Howe was sponsored in part by the Office of Naval Research under grant number N000-14-67-A-0321-0003 (NR047-096).*

A General Saddle Point Result for Constrained Optimization

by

K.J. Arrow[1], F.J. Gould[2], S.M. Howe[2]

Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 774

September, 1971

A General Saddle Point Result for Constrained Optimization

by

K. J. Arrow[1], F. J. Gould[2], S. M. Howe[2]

I. Introduction

In the context of nonlinear programming theory, the existence of a saddle point of the Lagrangian function is known to be heavily dependent upon convexity properties of the underlying problem. In particular, for a concave program satisfying the Slater condition, with a solution x^*, there is a u^* such that x^*, u^* is a saddle point of the Lagrangian function. In 1956, motivated by game theoretical and economic implications, Arrow and Hurwicz demonstrated that the concavity assumptions could be relaxed via a modified Lagrangian approach [1]. The results in this 1956 paper were presented in terms of a specific modified Lagrangian formulation. In 1958, in a discussion of gradient methods, Arrow and Solow presented additional saddle point results in terms of a different modified Lagrangian function [2]. Another specific result along the same lines was presented by Gould and Howe in 1971 [4].

In this work, we both generalize and simplify the above presentations. Conditions will be given under which, for a nonconcave (as well as concave) program, a quite general function P will possess a saddle point corresponding

1 *Harvard University.* *Arrow's work was sponsored in part by the Office of Naval Research under grant number N00014-67A-0298-0019 (NR047-004).*

2 *University of North Carolina at Chapel Hill.* *The work of Gould and Howe was sponsored in part by the Office of Naval Research under grant number N000-14-67-A-0321-0003 (NR047-096).*

to the program solution. Specific realizations of the P function will be the modified Lagrangian expressions discussed by the above mentioned authors.

II. Reduction to an Unconstrained Problem

We formulate the nonlinear program with both equality and inequality constraints:

$$\text{(P)} \qquad \max_{x \in R^n} f(x), \quad \text{subject to} \quad \begin{array}{l} g_i(x) \le 0, \quad i = 1,\ldots,m \\ h_j(x) = 0, \quad j = 1,\ldots,p. \end{array}$$

Throughout the paper it is assumed that all functions in (P) are twice differentiable.

Corresponding to the inequality constraints (the correspondence will be clear from the P function formulation) let $\lambda(\xi,\eta,\alpha)\colon R{\times}R_+{\times}R_+ \to R$ be a multiplier function having second partial derivatives with respect to the first argument[1], where we employ the notation $\partial\lambda/\partial\xi = \lambda_1$, $\partial^2\lambda/\partial\xi^2 = \lambda_{11}$. Impose the following properties on λ

(i) for any $\alpha > 0$ $\quad \lambda_1(0,\eta,\alpha) = \eta \quad$ for every $\eta \ge 0$

(ii) for any $\alpha > 0$ $\quad \lambda_1(\xi,0,\alpha) = 0 \quad$ for every $\xi < 0$

(iii) for each fixed $\eta > 0$ $\quad \lambda_{11}(0,\eta,\alpha) \to \infty$ as $\alpha \to \infty$.

Note: (ii) implies that, for any $\alpha > 0$, $\xi < 0$, $\lambda_{11}(\xi,0,\alpha) = 0$.

Corresponding to the equality constraints let $\phi(\xi,\eta,\alpha)\colon R{\times}R{\times}R_+ \to R$ be a multiplier function having second partial derivatives with respect to the first argument, such that

(v) for any $\alpha > 0$ $\quad \phi_1(0,\eta,\alpha) = \eta \quad$ for every $\eta \in R$

(vi) for each fixed $\eta \in R$ $\quad \phi_{11}(0,\eta,\alpha) \to \infty$ as $\alpha \to \infty$.

---

[1] R_+^k *denotes the nonnegative orthant of* R^k.

Examples of the λ functions are

M1: $\quad \eta(\xi+1)^{1+\alpha}/(1+\alpha), \qquad \alpha$ an even integer

M2: $\quad (\eta/\alpha)exp(\alpha\xi)$

M3: $$\begin{cases} -\eta/4\alpha, & \xi \le -1/2\alpha \\ \eta\alpha\xi^2+\eta\xi, & \xi > -1/2\alpha \end{cases}$$

M4: $$\begin{cases} -\eta^2/4\alpha, & \xi \le -\eta/2\alpha \\ \alpha\xi^2+\eta\xi\ , & \xi > -\eta/2\alpha. \end{cases}$$

An example of the ϕ function is

M5: $\quad \alpha\xi^2 + \eta\xi.$

Now define the modified Lagrangian function $P\colon R^n \times R^m \times R^p \times R_+ \to R$ as follows:

$$P(x,\mu,\psi,\alpha) \;=\; f(x) - \sum_{i=1}^{m} \lambda(g_i(x),\mu_i,\alpha) - \sum_{j=1}^{p} \phi(h_j(x),\psi_j,\alpha).$$

If the multiplier functions are chosen as in M1 or M5 one obtains, respectively, the P functions studied by Arrow and Hurwicz [1] and by Arrow and Solow [2]. By the choice M2 one obtains the P function studied by Gould and Howe [4].

The following result, sometimes referred to as Finsler's Lemma, will be used (see Debreu [3], or the appendix to this paper):

Lemma: Let Q be a real $n\times n$ matrix and let L be a real $m\times n$ matrix. Suppose $z^TQz < 0$ for every $z \neq 0$ such that $Lz = 0$. Then, for all α sufficiently large, $z^T[Q-\alpha L^TL]z < 0$ for all $z \neq 0$. That is, if Q is negative definite on the null space of L then for α sufficiently large $Q-\alpha L^TL$ is negative definite on the whole space.

Let $L(x,\mu,\psi) = f(x)-\langle\mu,g(x)\rangle-\langle\psi,h(x)\rangle$, which is the usual Lagrangian function.

The first result of this paper is the following.

Theorem 1: Suppose (x^*,μ^*,ψ^*) satisfy the second order sufficiency conditions for x^* to be an isolated local solution to (P) and suppose strict complementarity holds. That is, we assume

(a) $\nabla L(x^*,\mu^*,\psi^*) = 0$

(b) $g_i(x^*) \le 0, \quad i = 1,\dots,m$

$h_j(x^*) = 0, \quad j = 1,\dots,p$

(c) $\mu_i^* \ge 0, \quad \mu_i^* > 0$ if and only if $g_i(x^*) = 0, \quad i = 1,\dots,m$

(d) $y^T\nabla^2 L(x^*,\mu^*,\psi^*)y < 0$ for every nonzero y such that $y^T\nabla g_i(x^*) = 0$, $i\in I$, and $y^T\nabla h_j(x^*) = 0$, $j = 1,\dots,p$, where I denotes the active inequality constraints. That is, $I = \{i: g_i(x^*)=0\}$. Then, if α is sufficiently large, the function $P(x,\mu^*,\psi^*,\alpha): R^n \to R$ has an unconstrained isolated local maximum at x^*.

Proof: For any $\alpha > 0$,

$$\begin{aligned}\nabla P(x^*,\mu^*,\psi^*,\alpha) &= \nabla f(x^*) - \sum_{i=1}^{m} \lambda_1(g_i(x^*),\mu_i^*,\alpha)\nabla g_i(x^*) \\ &\quad - \sum_{j=1}^{p} \phi_1(h_j(x^*),\psi_j,\alpha)\nabla h_j(x^*) \\ &= \nabla f(x^*) - \sum_{I} \mu_i^*\nabla g_i(x^*) - \sum_{j=1}^{p} \psi_j^*\nabla h_j(x^*) \\ &= \nabla L(x^*,\mu^*,\psi^*) = 0 \quad \text{by (a).}\end{aligned}$$

$$\begin{aligned}\nabla^2 P(x^*,\mu^*,\psi^*,\alpha) &= \nabla^2 f(x^*) - \sum_{i=1}^{m} [\lambda_1(g_i(x^*),\mu_i^*,\alpha)\nabla^2 g_i(x^*) \\ &\quad + \lambda_{11}(g_i(x^*),\mu_i^*,\alpha)\nabla g_i(x^*)\nabla^T g_i(x^*)] \\ &\quad - \sum_{j=1}^{p} [\phi_1(h_j(x^*),\psi_j^*,\alpha)\nabla^2 h_j(x^*)\end{aligned}$$

$$+ \phi_{11}(h_j(x^*),\psi_j^*,\alpha)\nabla h_j(x^*)\nabla^T h_j(x^*)]$$

$$= \nabla^2 L(x^*,\mu^*,\psi^*) - \sum_I \lambda_{11}(g_i(x^*),\mu_i^*,\alpha)\nabla g_i(x^*)\nabla^T g_i(x^*)$$

$$- \sum_{j=1}^{p} \phi_{11}(h_j(x^*),\psi_j^*,\alpha)\nabla h_j(x^*)\nabla h_j^T(x^*).$$

Defining $J^T = [\nabla g_i(x^*),..,i\epsilon I,\ \nabla h_1(x^*),..,\nabla h_p(x^*)]$, note that $y^T\nabla^2 L(x^*,\mu^*,\psi^*)y < 0$ for every nonzero y such that $Jy = 0$, by (d). By Finsler's Lemma, $\nabla^2 L(x^*,\mu^*,\psi^*)-kJ^TJ$ is negative definite for some scalar k sufficiently large. Now if α is sufficiently large we can obtain

$$\lambda_{11}(g_i(x^*),\mu_i^*,\alpha) > k \qquad \text{each } i\epsilon I$$
$$\phi_{11}(h_j(x^*),\psi_j^*,\alpha) > k \qquad \text{each } j = 1,\ldots,p$$

whence

$$\nabla^2 P(x^*,\mu^*,\psi^*,\alpha) = \nabla^2 L(x^*,\mu^*,\psi^*) - kJ^TJ$$
$$+ \sum_I [k-\lambda_{11}(g_i(x^*),\mu_i^*,\alpha)]\nabla g_i(x^*)\nabla g_i^T(x^*)$$
$$+ \sum_{j=1}^{p} [k-\phi_{11}(h_j(x^*),\psi_j^*,\alpha)]\nabla h_j(x^*)\nabla h_j^T(x^*)$$

which, for all α sufficiently large, is negative definite, since large α implies that each of the dyadic terms has a negative coefficient and is hence negative semidefinite.

□

III. A Local Saddle Point Result

To obtain the main result, it is necessary to impose the further assumptions

(iv) if $\alpha > 0$, $\eta \geq 0$ then $\lambda(\xi,\eta,\alpha)$ is monotonically nondecreasing over $\xi\epsilon(-\infty,0]$.

It should be noted that condition (iv) is satisfied by the examples M1 thru M4.

Theorem 2: If the multiplier functions satisfy properties (i)-(vi), then under the conditions of Theorem 1, if α is sufficiently large,

$$\bar{P}(x,\mu^*,\psi^*,\alpha) \le \bar{P}(x^*,\mu^*,\psi^*,\alpha) \le \bar{P}(x^*,\mu,\psi,\alpha)$$

for every x in some neighborhood N of x^* and every point $(\mu,\psi)\epsilon R_+^m \times R^p$, where

$$\bar{P}(x,\mu,\psi,\alpha) = f(x) + \sum_{i=1}^{m} [\lambda(0,\mu_i,\alpha) - \lambda(g_i(x),\mu_i,\alpha)] + \sum_{j=1}^{p} [\phi(0,\psi_j,\alpha) - \phi(h_j(x),\psi_j,\alpha)].$$

Proof: From Theorem 1, it follows immediately that $\bar{P}(x,\mu^*,\psi^*,\alpha) \le P(x^*,\mu^*,\psi^*,\alpha)$ for every x in some neighborhood N of x^*. Also, since $g_i(x^*) = 0$, $i\epsilon I$, $h_j(x^*) = 0$, $j = 1,\ldots,p$, we have

$$\bar{P}(x^*,\mu,\psi,\alpha) = f(x^*) + \sum_{i\notin I} [\lambda(0,\mu_i,\alpha) - \lambda(g_i(x^*),\mu_i,\alpha)].$$

For $i\notin I$, $\mu^*_i = 0$. Integration of (ii) from $g_i(x^*)$ to 0 shows that $\lambda(0,\mu^*_i,\alpha)-\lambda(g_i(x^*),\mu^*_i,\alpha) = 0$ for $i\notin I$. From the monotonicity property (iv), $\lambda(0,\mu_i,\alpha)-\lambda(g_i(x^*),\mu_i,\alpha) \ge 0$ for any $\mu_i \ge 0$. It follows immediately that $\bar{P}(x^*,\mu^*,\psi^*,\alpha) \le \bar{P}(x^*,\mu,\psi,\alpha)$ for all $(\mu,\psi)\epsilon R_+^m \times R^p$.

□

APPENDIX

Lemma: If $z^TQz < 0$ for every $z \neq 0$ such that $Lz = 0$ then $Q-\alpha L^TL$ is negative definite for α sufficiently large.

Proof: Let $S^\perp$ denote the nullspace of L. Hence, $S^\perp = \{z: Lz=0\}$, $S = S^{\perp\perp}$. If $y\epsilon S$, $y \neq 0$, then $y^TL^TLy = ||Ly||^2 > 0$. Let $m_1 = \min_y [y^TL^TLy: y\epsilon S, ||y||=1]$, and let $M = ||Q||$, the sup norm of the matrix Q. There are three cases to consider.

Case (i): Suppose $S^{\perp} = R^n$. Then Q is negative definite and the result is true for all $\alpha \geq 0$ since $-L^TL$ is negative semidefinite.

Case (ii): Suppose $S^{\perp} = \{0\}$. Then $S = R^n$. Then for every $v \epsilon R^n$, $v \neq 0$,

$$v^T[Q-\alpha L^TL]v = v^TQv - \alpha v^TL^TLv = v^TQv - \alpha\left[\frac{v^T}{||v||} L^TL \frac{v}{||v||}\right] ||v||^2$$

$$\leq ||v||^2M - \alpha m_1||v||^2 = ||v||^2[M-\alpha m_1] < 0 \quad \text{if} \quad \alpha > M/m_1.$$

Case (iii): Suppose $\{0\} \neq S^{\perp} \subsetneq R^n$. Then for every $z\epsilon S^{\perp}$, $z \neq 0$, $z^TQz < 0$. Hence, let $max_z[z^TQz: z\epsilon S^{\perp}, ||z||=1] = -m_2 < 0$. Suppose $v\epsilon R^n$, $v \neq 0$. Then we can write $v = y+z$ for some $y\epsilon S$, $z\epsilon S^{\perp}$. Consequently,

$$\begin{aligned}
v^T[Q-\alpha L^TL]v &= (y+z)^T[Q-\alpha L^TL](y+z) \\
&= z^TQz + y^TQz + z^TQy + y^TQy - \alpha y^TL^TLy \\
&\leq -m_2||z||^2 + 2M||y|| \, ||z|| + M||y||^2 - \alpha m_1||y||^2 \\
&= -\alpha m_1||y||^2 - m_2\left(||z|| - \frac{M}{m_2}||y||\right)^2 + \frac{M^2}{m_2}||y||^2 + M||y||^2 \\
&= ||y||^2\left(-\alpha m_1 + \frac{M^2}{m_2} + M\right) - m_2\left(||z|| - \frac{M}{m_2}||y||\right)^2 \\
&< 0 \quad \text{if} \quad \alpha > M/m_1 \ (1 + M/m_2).
\end{aligned}$$

$\square$

REFERENCES

[1] K.J. Arrow and L. Hurwicz, Reduction of constrained maxima to saddle-point problems, in: Third Berkeley Symposium on Mathematical Statistics and Probability, ed. J. Neyman (University of California Press, Berkeley, 1956).

[2] K.J. Arrow and R.M. Solow, Gradient methods for constrained maxima, with weakened assumptions, in: Studies in Linear and Non-linear Programming, ed. K. Arrow, L. Hurwicz, H. Uzawa (Stanford University Press, Stanford, 1958).

[3] G. Debreu, Definite and semidefinite quadratic forms, Econometrica XX, (1952) 295-300.

[4] F.J. Gould and S.M. Howe, "A new result on interpreting Lagrange multi-pliers as dual variables", Institute of Statistics Mimeo Series No. 738, Dept. of Statistics, University of North Carolina (Chapel Hill), January, 1971.