WIS-CS-74-201
Computer Sciences Department
The University of Wisconsin
1210 West Dayton Street
Madison, Wisconsin 53706

Received December 20, 1973

UNCONSTRAINED LAGRANGIANS IN NONLINEAR PROGRAMMING

by

O. L. Mangasarian

Computer Sciences Technical Report No. 201

January 1974

UNCONSTRAINED LAGRANGIANS IN NONLINEAR PROGRAMMING[1)3)]

by

O. L. Mangasarian[2)]

Abstract

The main purpose of this work is to associate a wide class of Lagrangian functions with a nonconvex, inequality and equality constrained optimization problem in such a way that unconstrained stationary points and local saddlepoints of each Lagrangian are related to Kuhn-Tucker points or local or global solutions of the optimization problem. As a consequence of this we are able to obtain duality results and two computational algorithms for solving the optimization problem. One algorithm is a Newton algorithm which has a local superlinear or quadratic rate of convergence. The other method is a locally linearly convergent method for finding stationary points of the Lagrangian and is an extension of the method of multipliers of Hestenes and Powell to inequalities.

1) Supported by NSF Grant GJ 35292.

2) Computer Sciences Department, University of Wisconsin
1210 West Dayton Street, Madison, Wisconsin 53706

3) This is a revision of Technical Report #174 dated March 1973. The major difference consists in section 4 where the computational algorithms are considerably simpler than before and employ a different Lagrangian. For simplicity a smaller class of Lagrangians have been used here.

1. INTRODUCTION

In 1970 Rockafellar [21] introduced a Lagrangian for inequality constrained convex programming problems for which an unconstrained saddlepoint corresponded to a solution of the convex programming problem. Moreover this Lagrangian was once differentiable everywhere if the objective and constraint functions of the convex programming problem were also differentiable everywhere. In 1971 Arrow, Gould and Howe [1] considered a general class of Lagrangians (including Rockafellar's) for nonconvex programming problems and established local saddle point properties for this class of Lagrangians. For their class of Lagrangians however, the saddlepoint was in general nonnegatively constrained just as it is in the classical Kuhn-Tucker [11] Lagrangian for nonlinear programming. The local saddlepoint property was obtained by the presence of a convexifying parameter in their Lagrangian which made the Hessian of the Lagrangian positive definite for large enough, but finite, values of the parameter. This elegant idea of local convexification was first introduced by Arrow and Solow in 1958 [2] in connection with equality constrained problems and was later independently reconsidered in a different algorithmic context by Hestenes [8,9] and Powell [19] in 1969 and by Haarhoff and Buys [7] in 1970. Miele, Moseley and Cragg [14,15] have conducted numerical experiments on these ideas for equality constrained problems. More recently Rockafellar [22] gave an illuminating derivation of his Lagrangian for inequality constrained problems from the Arrow-Solow Lagrangian for equality constrained problems by the use of slack variables.

A primary purpose of this work is to relate Kuhn-Tucker points of nonconvex, inequality and equality constrained nonlinear programming problems to unconstrained stationary points of a wide class of Lagrangian functions. Such a relation is important because it can bring to bear all the algorithms and results of nonlinear equations theory [17, 18] on nonlinear programming. As a consequence of this relationship we present in this work local and global duality results (section 3), a new superlinearly or quadratically convergent algorithm (algorithm 4.7 and theorem 4.8), and a linearly convergent extension to inequality constraints and to more general Lagrangians of the method of multipliers (algorithm 4.9 and theorem 4.10).

The difference between our approach and that of Rockafellar [21,22,23] is that Rockafellar's results are valid only for convex problems, whereas in our approach convexity plays only a minor role in some of the peripheral results. In [24] Rockafellar extends the results for his specific Lagrangian to nonconvex optimization problems and relates global solutions of the optimization problem to global saddlepoints of his Lagrangian. Our results are principally aimed at related local stationary points of the two problems and are established for a general class of Lagrangians. Also Rockafellar's Lagrangian is differentiable only one globally, whereas ours are twice differentiable globally. This is an important distinction in the application of Newton type algorithms which require twice differentiability. In obtaining this twice differentiability property we lose the general concavity of Rockafellar's Lagrangian in the dual variably y. However our Lagrangians are concave in y for primal feasible points (see remark 2.13 below). The difference between our approach and that of Arrow, Gould and Howe [1] is that for their general result the Lagrangian saddlepoint is constrained by nonnegativity constraints whereas the stationary points of our Lagrangians are completely unconstrained. Also, the conditions imposed on our Lagrangians are different from their conditions. In addition we give a new general formulation for unconstrained Lagrangians together with new concrete realizations.

We shall be concerned throughout this paper with the following problem

1.1 minimize $f(x)$

subject to $g_i(x) \leqq 0 \quad i=1,\ldots,m$

$$g_i(x) = 0 \quad i=m+1,\ldots,k$$

where f, and $g_i, i=1,\ldots,k$, are functions from R^n into R. We shall associate with this problem a real valued Lagrangian function L in such a way that Kuhn-Tucker points of 1.1 are related to unconstrained stationary and saddlepoints of L. This is done in section 2 of the paper where in addition, we give sufficient conditions different from those of [1], for the Hessian of L with respect to x to be positive definite. This latter result is important in establishing the local duality results of section 3 and the convergence of the algorithms of section 4. In section 3 we establish duality results between problem 1.1 and an equality constrained dual problem, problem 3.2. We establish a weak duality theorem 3.3 in the presence of convexity, a duality theorem 3.4 and a converse duality theorem 3.10 in which convexity plays a secondary role. In particular we relate, among other things, points satisfying Kuhn-Tucker conditions to points satisfying second order optimality conditions, without any convexity assumptions. In section 4 we present two computational algorithms for the solution of 1.1 based upon finding stationary points of a Lagrangian M obtained by augmenting L. Algorithm 4.7 is a Newton method for finding a zero of the gradient of M and for which we establish under suitable conditions a superlinear or quadratic rate of convergence. Algorithm 4.9 is an extension of the method of multipliers [8,9,19,7] to inequalities and for which we establish a local linear rate of convergence.

We shall make use of the following notation. For the point $\bar{x}$ satisfying the constraints of problem 1.1 we shall define the following sets: $I = \{i \mid g_i(\bar{x})=0, i=1,\ldots,m\}$, $J = \{i \mid g_i(\bar{x})<0, i=1,\ldots,m\}$, $E = \{i \mid i=m+1,\ldots,k\}$ and $K = I \cup E$. For the Lagrangian $L:R^n \times R^k \times (0,\infty) \to R$, $\nabla L(x,y,\alpha)$ will denote its $(n+k)$-dimensional gradient with respect to (x,y), $\nabla_1 L(x,y,\alpha)$ its n-dimensional gradient with respect to x, $\nabla_2 L(x,y,\alpha)$ its k-dimensional gradient with respect to y, $\nabla^2 L(x,y,\alpha)$ its $(n+k) \times (n+k)$ Hessian matrix with respect to (x,y). The submatrices of $\nabla^2 L(x,y,\alpha)$ with be denoted by $\nabla_{11} L(x,y,\alpha)$, $\nabla_{12} L(x,y,\alpha)$, $\nabla_{21} L(x,y,\alpha)$ and $\nabla_{22} L(x,y,\alpha)$. All vectors are either row or column vectors depending on the context. A superscript T will denote the transpose and will be used only in denoting the transpose of a matrix or the tensor product of two vectors.

2. EQUIVALENCE OF KUHN-TUCKER POINTS AND UNCONSTRAINED STATIONARY AND SADDLEPOINTS

A primary objective of this work is to relate points that satisfy the Kuhn-Tucker conditions for problem 1.1 to unconstrained stationary points and saddlepoints of an appropriately defined Lagrangian L. For that purpose we begin by defining such a Lagrangian as follows

2.1 $$L(x,y,\alpha) = f(x) + \sum_{i=1}^{m} (\psi(\alpha g_i(x)+y_i)_+ - \psi(y_i)) + \sum_{i=m+1}^{k} (\psi(\alpha g_i(x)+y_i) - \psi(y_i))$$

where $\alpha > 0$ and

$$\psi: R \to R \ , \ \psi(\zeta)_+ = \begin{cases} \psi(\zeta) & \text{if } \zeta \geqq 0 \\ 0 & \text{if } \zeta < 0 \end{cases}$$, and ψ satisfies the following

conditions

2.2 (a) ψ is twice differentiable on R and $\psi''(\zeta) > 0$ for $\zeta \neq 0$

(b) ψ' maps R onto R and $\psi'(0) = 0$

(c) $\psi(0) = 0$

It immediately follows from the above conditions that

2.2 (d) ψ' is a strictly increasing function on R

(e) ψ is a nonnegative convex function on R

The motivation behind the above Lagrangian is the following. For the case of equality constraints <u>only</u>, it is easy to see that for

$$L(x,y,\alpha) = f(x) + \sum_{i=m+1}^{k} (\psi(\alpha g_i(x)+y_i) - \psi(y_i)),$$

the condition $\nabla L(x,y,\alpha) = 0$ is equivalent to $\nabla f(x) + \sum_{i=m+1}^{k} \alpha\psi'(\alpha g_i(x)+y_i)\nabla g_i(x) = 0$

and $\psi'(\alpha g_i(x)+y_i) - \psi'(y_i) = 0$, $i = m+1, \ldots, k$. Since ψ' is strictly increasing, the last equality gives that $\alpha g_i(x)+y_i = y_i$ or $g_i(x) = 0$, $i = m+1, \ldots, k$, and the gradient with respect to x becomes $\nabla f(x) + \sum_{i=m+1}^{k} \alpha\psi'(y_i)\nabla g_i(x) = 0$. These are precisely the Kuhn-Tucker conditions for the equality constrained problem with classical Lagrange multipliers $u_i = \alpha\psi'(y_i)$, $i = m+1, \ldots, k$. The case of inequality constraints $g_i(x) \leqq 0, i=1, \ldots, m$, is handled by introducing the slack variable variables z_i and writing $g_i(x)+z_i^2 = 0, i=1, \ldots, m$. Using the Lagrangian just introduced for equality constraints we have

$$L(x,z,y,\alpha) = f(x) + \sum_{i=1}^{m} (\psi(\alpha g_i(x)+\alpha z_i^2+y_i) - \psi(y_i)).$$

The variable z can be eliminated now by setting the gradient of L with respect to z equal to zero, a condition which must be satisfied by the Lagrangian for equality constraints. This gives the condition that z_i must satisfy $2\alpha z_i\psi'(\alpha g_i(x)+\alpha z_i^2+y_i) = 0, i=1, \ldots, m$. This condition is satisfied if we set $z_i = \alpha^{-\frac{1}{2}}(-\alpha g_i(x)-y_i)^{\frac{1}{2}}$ when $\alpha g_i(x) + y_i < 0$ and $z_i = 0$ when $\alpha g_i(x) + y_i \geqq 0$. The Lagrangian $L(x,z,y,\alpha)$ becomes then

$f(x) + \sum_{i=1}^{m} (\psi(\alpha g_i(x)+y_i)_+ - \psi(y_i))$, which is what is given in 2.1 for the inequality constraints.

Because $\psi'(0) = 0$, we have that $(\psi(\zeta)_+)' = \psi'(\zeta)_+$ for all ζ in R. On the other hand $(\psi(\zeta)_+)'' = \psi''(\zeta)_+$ only for $\zeta \neq 0$ in R, with equality holding for $\zeta = 0$ if we assume in addition that $\psi''(0) = 0$. This extra assumption will be explicitly made where needed. For notational simplicity here and elsewhere we have used the same ψ function for both inequality and equality constraint functions $g_i, i=1,\ldots,k$. In fact different ψ functions may be used for each constraint function $g_i, i=1,\ldots,k$. Occasionally we shall write ψ_i to denote the ψ function used with a specific constraint function $g_i, i=1,\ldots,k$. Typical ψ functions which satisfy all of conditions 2.2 are

2.3 (a) $\psi(\zeta) = \frac{1}{\alpha t}|\zeta|^t \quad \alpha \in R, \ \alpha > 0, \ t \geqq 2$

(b) $\psi(\zeta) = \cosh \zeta - \frac{\zeta^2}{2} - 1$

(c) $\psi(\zeta) = \frac{1}{2}(\cosh \zeta - 1)^2$

If the ψ function of 2.3a with $t = 2$ is used for both inequality and equality constraints we would obtain Rockafellar's Lagrangian [21-24] which is not twice differentiable globally because $\psi''(0) > 0$. However, every other ψ function given in 2.3 has the property that $\psi''(0) = 0$ and hence $(\psi(\zeta)_+)'' = \psi''(\zeta)_+$ for <u>all</u> ζ in R . This property that $\psi''(0) = 0$ that leads to global twice differentiability will be exploited in some of the subsequent results such as theorems 3.4, 3.10 and algorithms 4.7 and 4.9 . A more general Lagrangian formulation is given in [13] .

For the sake of explicitness we give below a Lagrangian based on the ψ function of 2.3a with $t = 4$ for the inequality constraints and $t = 2$ for equality constraints

2.4 $$L(x,y,\alpha) = f(x) + \frac{1}{4\alpha}\sum_{i=1}^{m}((\alpha g_i(x)+y_i)_+^4 - y_i^4) + \frac{1}{2\alpha}\sum_{i=m+1}^{k}((\alpha g_i(x)+y_i)^2 - y_i^2)$$

$$= f(x) + \frac{1}{4\alpha}\sum_{i=1}^{m}((\alpha g_i(x)+y_i)_+^4 - y_i^4) + \sum_{i=m+1}^{k}(\frac{\alpha}{2}g_i(x)^2 + y_i g_i(x))$$

Every single result obtained in this paper applies but is not limited to this specific Lagrangian which is twice (thrice) differentiable globally if the functions $f, g_i, i=1,\ldots,k$, are also twice (thrice) differentiable globally.

We begin by relating Kuhn-Tucker points of problem 1.1 and stationary points of L, that is (x,y) such that $\nabla L(x,y,\alpha) = 0$.

2.5 <u>Equivalence theorem</u> Let f and $g_i, i=1,\ldots,k$ be differentiable at $\bar{x}$ and let α be <u>any</u> positive number. If $(\bar{x},\bar{u})$ is a Kuhn-Tucker point of 1.1 then $\bar{x}$ and $\bar{y}$ defined by 2.6 below constitute a stationary point of L. Conversely if $(\bar{x},\bar{y})$ is a stationary point of L, then $\bar{x}$ and $\bar{u}$ defined by 2.8 below, constitute a Kuhn-Tucker point of 1.1. ▪

<u>Proof</u> Suppose that $(\bar{x},\bar{u})$ satisfies the Kuhn-Tucker conditions of problem 1.1. Define $\bar{y}$ in R^k as follows

2.6 $$\psi'(\bar{y}_i) = \frac{\bar{u}_i}{\alpha} \qquad i = 1,\ldots,k$$

The existence of a unique $\bar{y}$ satisfying 2.6 is assured by assumption 2.2b. Hence

$$\nabla_1 L(\bar{x},\bar{y},\alpha) = \nabla f(\bar{x}) + \sum_{i\in I} \alpha\psi'(\bar{y}_i)_+ \nabla g_i(\bar{x}) + \sum_{i\in J} \alpha\psi'(\alpha g_i(\bar{x})+\bar{y}_i)_+ \nabla g_i(\bar{x})$$

$$+ \sum_{i\in E} \alpha\psi'(\bar{y}_i)\nabla g_i(\bar{x}) = \nabla f(\bar{x}) + \sum_{i=1}^{k} \bar{u}_i \nabla g_i(\bar{x}) = 0$$

$$\frac{\partial L}{\partial y_i}(\bar{x},\bar{y},\alpha) = \psi'(\alpha g_i(\bar{x})+\bar{y}_i)_+ - \psi'(\bar{y}_i) = \psi'(\bar{y}_i) - \psi'(\bar{y}_i) = 0 \quad i \in I$$

$$\frac{\partial L}{\partial y_i}(\bar{x},\bar{y},\alpha) = \psi'(\alpha g_i(\bar{x}))_+ - \psi'(0) = 0 \quad i \in J$$

$$\frac{\partial L}{\partial y_i}(\bar{x},\bar{y},\alpha) = \psi'(\alpha g_i(\bar{x})+\bar{y}_i) - \psi'(\bar{y}_i) = \psi'(\bar{y}_i) - \psi'(\bar{y}_i) = 0 \quad i \in E$$

Hence $\nabla_2 L(\bar{x},\bar{y},\alpha) = 0$ and $(\bar{x},\bar{y})$ is a stationary point of L . To prove the converse it will be convenient to establish the following key lemma first.

2.7 <u>Lemma</u> For any $\alpha > 0$ and L defined by 2.1 we have that

$$\left\langle \frac{\partial L}{\partial y_i}(x,y,\alpha) = 0, i=1,\ldots,m \right\rangle \Longleftrightarrow \left\langle g_i(x) \leqq 0,\ y_i \geqq 0,\ y_i g_i(x) = 0, i=1,\ldots,m \right\rangle$$

$$\left\langle \frac{\partial L}{\partial y_i}(x,y,\alpha) = 0, i=m+1,\ldots,k \right\rangle \Longleftrightarrow \left\langle g_i(x) = 0, i=m+1,\ldots,k \right\rangle$$

<u>Proof</u> For $i=1,\ldots,m$

$$\left\langle \frac{\partial L}{\partial y_i}(x,y,\alpha) = \psi'(\alpha g_i(x) + y_i)_+ - \psi'(y_i) = 0 \right\rangle \Longleftrightarrow$$

$$\left\langle \begin{array}{c} \alpha g_i(x) + y_i \geqq 0,\ \alpha g_i(x) + y_i = y_i \\ \text{or} \\ \alpha g_i(x) + y_i < 0,\ y_i = 0 \end{array} \right\rangle \Longleftrightarrow \left\langle \begin{array}{c} g_i(x) = 0,\ y_i \geqq 0 \\ \text{or} \\ g_i(x) < 0,\ y_i = 0 \end{array} \right\rangle \Longleftrightarrow$$

$$\left\langle \begin{array}{c} g_i(x) \leqq 0 \\ y_i \geqq 0 \\ y_i g_i(x) = 0 \end{array} \right\rangle$$

For $i = m+1, \ldots, k$

$$\left\langle \frac{\partial L}{\partial y_i}(x,y,\alpha) = \psi'(\alpha g_i(x)+y_i) - \psi'(y_i) = 0 \right\rangle \Longleftrightarrow \left\langle \alpha g_i(x)+y_i = y_i \right\rangle \Longleftrightarrow \left\langle g_i(x) = 0 \right\rangle \quad \blacksquare$$

To complete the proof of theorem 2.5 now, suppose that $(\bar{x},\bar{y})$ is a stationary point of L. Define $\bar{u} \in R^k$ as follows

2.8 $$\bar{u}_i = \alpha\psi'(\alpha g_i(\bar{x})+\bar{y}_i)_+ \qquad i=1,\ldots,m$$

$$\bar{u}_i = \alpha\psi'(\alpha g_i(\bar{x})+\bar{y}_i) \qquad i=m+1,\ldots,k$$

Hence $\bar{u}_i \geqq 0, i=1,\ldots,m$ and

$$\nabla f(\bar{x}) + \sum_{i=1}^{k} \bar{u}_i \nabla g_i(\bar{x}) = \nabla_1 L(\bar{x},\bar{y},\alpha) = 0$$

By lemma 2.7 we have that, since $\nabla_2 L(\bar{x},\bar{y},\alpha) = 0$, $g_i(\bar{x}) \leqq 0$, $y_i \geqq 0$, $y_i g_i(\bar{x}) = 0$, $i=1,\ldots,m$ and $g_i(\bar{x}) = 0, i=m+1,\ldots,k$. Hence for $i \in J$, $\bar{y}_i = 0$, $\bar{u}_i = \alpha\psi'(\bar{y}_i) = 0$, and so $\bar{u}_i g_i(\bar{x}) = 0, i=1,\ldots,m$. Hence $(\bar{x},\bar{u})$ satisfies the Kuhn-Tucker conditions for problem 1.1. $\blacksquare$

The significance of theorem 2.5 lies in the fact that the problem of finding a Kuhn-Tucker point of a nonlinear programming problem has been reduced to that of finding solutions of the nonlinear equations $\nabla L(x,y,\alpha) = 0$ for any positive α. In section 4 we shall describe two computational methods for finding Kuhn-Tucker points of problem 1.1 based on solving these nonlinear equations.

We establish a result now which is essentially due to Arrow, Gould and Howe [1], but under different assumptions from theirs. This result is important in establishing some of the duality and computational results to follow. We shall need the <u>second order sufficiency</u> conditions for problem 1.1 [5]. A point $(\bar{x},\bar{u}) \in R^n \times R^k$ is said to satisfy the second order sufficiency conditions for problem 1.1 if it satisfies the Kuhn-Tucker conditions of problem 1.1 and if $x\nabla^2 L^0(\bar{x},\bar{u})x > 0$ for each nonzero x in R^n satisfying $\nabla g_i(\bar{x})x = 0$ for $i \in \{i \mid \bar{u}_i > 0, g_i(\bar{x})=0, i=1,\ldots,m\} \cup E$, and $\nabla g_i(\bar{x})x \leqq 0$ for $i \in \{i \mid \bar{u}_i = 0, g_i(\bar{x}) = 0, i=1,\ldots,m\}$, where L^0 is the classical Lagrangian defined by $L^0(x,u) = f(x) + \sum_{i=1}^{k} u_i g_i(x)$. We shall also use the concept of <u>strict complementarity</u> at $(\bar{x},\bar{u})$, that is $\bar{u}_i \neq 0$ for each $i=1,\ldots,k$ for which $g_i(\bar{x}) = 0$.

2.9 <u>Theorem</u> (Positive definiteness of $\nabla_{11}L(\bar{x},\bar{y},\alpha)$ and saddlepoint result)

(a) Let $f, g_i, i=1,\ldots,k$ be twice differentiable at $\bar{x}$, let $(\bar{x},\bar{u})$ satisfy the second order sufficiency conditions for problem 1.1, and let strict complementarity hold at $(\bar{x},\bar{u})$. Then for $\bar{x}$ and $\bar{y}$ defined by 2.6 and for some large enough but finite α, $\nabla_{11}L(\bar{x},\bar{y},\alpha)$ is positive definite and

2.10 $$L(\bar{x},y,\alpha) \leqq L(\bar{x},\bar{y},\alpha) < L(x,\bar{y},\alpha) \qquad \forall y \in R^k, \forall x \in N(\bar{x}),\ x \neq \bar{x}$$

where $N(\bar{x})$ is some open neighborhood of $\bar{x}$. If $f, g_i, i=1,\ldots,m$ are convex, and $g_i, i=m+1,\ldots,k$ are affine, then $N(\bar{x}) = R^n$.

(b) Conversely, if 2.10 holds, with $<$ possibly replaced by $\leqq$, for some $\alpha > 0$, then $\bar{x}$ is a solution of 1.1 subject to the extra restriction that $x \in N(\bar{x})$. ∎

<u>Proof</u>: (a) By strict complementarity we have that for $i \in I$, $\bar{u}_i > 0$, and hence by 2.6a $\bar{y}_i > 0$ and $\alpha g_i(\bar{x}) + \bar{y}_i > 0$. So $(\psi(\zeta)_+)'' = \psi''(\zeta)$

for $\zeta = \alpha g_i(\bar{x}) + \bar{y}_i$, $i \in I$. From 2.6b we have that for $i \in J$, $\bar{y}_i = 0$ and hence $\alpha g_i(\bar{x}) + \bar{y}_i < 0$. So $\psi(\zeta)_+ = (\psi(\zeta)_+)' = (\psi(\zeta)_+)'' = 0$ for $\zeta = \alpha g_i(\bar{x}) + \bar{y}_i$, $i \in J$. Thus

2.11 $$\nabla_{11}L(\bar{x},\bar{y},\alpha) = \nabla^2 f(\bar{x}) + \sum_{i\in I\cup E} \alpha^2\psi''(\alpha g_i(\bar{x})+\bar{y}_i)\nabla g_i(\bar{x})\nabla g_i(\bar{x})^T$$

$$+ \sum_{i=1}^{k} \alpha\psi'(\alpha g_i(\bar{x})+\bar{y}_i)\nabla^2 g_i(\bar{x})$$

$$= [\nabla_{11}L^0(\bar{x},\bar{u}) + \sum_{i\in I\cup E} \alpha^2\psi''(\bar{y}_i)\nabla g_i(\bar{x})\nabla g_i(\bar{x})^T] \text{ (by 2.6)}$$

where $L^0(x,u)$ is the standard Lagrangian. Note that by 2.2a and strict complementarity we have that $\psi''(\bar{y}_i) > 0$ for $i\in I\cup E$. Hence by Debreu's theorem [4, theorem 3] which states that

$$\left\langle x \neq 0,\ Mx = 0 \Longrightarrow xLx > 0 \right\rangle \Longleftrightarrow \left\langle \begin{matrix} L + \gamma M^T M \text{ is positive} \\ \text{definite for } \gamma \text{ sufficiently large} \end{matrix} \right\rangle,$$

and the second order sufficiency conditions, it follows that the term in the square bracket above is positive definite for α sufficiently large. Hence for α large enough $\nabla_{11}L(\bar{x},\bar{y},\alpha)$ is positive definite and the second inequality of 2.10 holds for x , different from $\bar{x}$, in some open neighborhood $N(\bar{x})$ of $\bar{x}$. To establish the first inequality of 2.10 we have from 2.6a and 2.6b that $\bar{y}_i \geqq 0$ for $i \in I$ and $\bar{y}_i = 0$ for $i \in J$ and hence

$$L(\bar{x},y,\alpha) - L(\bar{x},\bar{y},\alpha) = \sum_{i=1}^{m} (\psi(\alpha g_i(\bar{x})+y_i)_+ - \psi(y_i))$$

$$+ \sum_{i=m+1}^{k} (\psi(\alpha g_i(\bar{x})+y_i)-\psi(y_i)) \leqq 0$$

where the last inequality follows from the fact that $\psi(\alpha g_i(\bar{x})+y_i)_+ \leqq \psi(y_i)_+ \leqq \psi(y_i)$

for $i=1,\ldots,m$, since $g_i(\bar{x}) \leqq 0$, ψ is nonnegative and $\psi(0) = 0$, and from the fact that $g_i(\bar{x}) = 0$ for $i = m+1,\ldots,k$. Hence $L(\bar{x},y,\alpha) \leqq L(\bar{x},\bar{y},\alpha)$ for all y in R^k. If in addition $f, g_i, i=1,\ldots,m$ are convex, and $g_i, i=m+1,\ldots,k$ are affine then it follows from the convexity and monotonicity of $\psi(\ )_+$, the convexity of ψ and affineness of $g_i, i=m+1,\ldots,k$, that $L(x,y,\alpha)$ is convex in x for each fixed y and α. Since $\nabla_1 L(\bar{x},\bar{y},\alpha) = 0$, it follows that $L(\bar{x},\bar{y},\alpha) \leq L(x,\bar{y},\alpha)$ for all x in R^n.

(b) Suppose now that 2.10 holds. From the second inequality of 2.10 we get that $\nabla_2 L(\bar{x},\bar{y},\alpha) = 0$ and from lemma 2.7 we obtain that $\bar{y}_i \geqq 0$, $g_i(\bar{x}) \leqq 0$, $\bar{y}_i g_i(\bar{x}) = 0$ for $i=1,\ldots,m$, and $g_i(\bar{x}) = 0$ for $i=m+1,\ldots,k$. Hence $\bar{x}$ is feasible. For any other feasible point x which is also in $N(\bar{x})$ we have that

$$0 \leqq L(x,\bar{y},\alpha) - L(\bar{x},\bar{y},\alpha) \qquad \text{(by 2.10)}$$

$$= f(x) - f(\bar{x}) + \sum_{i=1}^{m} (\psi(\alpha g_i(x)+\bar{y}_i)_+ - \psi(\alpha g_i(\bar{x})+\bar{y}_i)_+$$

$$+ \sum_{i=m+1}^{k} (\psi(\alpha g_i(x)+\bar{y}_i) - \psi(\alpha g_i(\bar{x})+\bar{y}_i))$$

$$= f(x) - f(\bar{x}) + \sum_{i\in I} (\psi(\alpha g_i(x)+\bar{y}_i)_+ - \psi(\bar{y}_i)_+) + \sum_{i\in J} (\psi(\alpha g_i(x))_+ - \psi(\alpha g_i(\bar{x}))_+)$$

$$+ \sum_{i\in E} (\psi(\bar{y}_i) - \psi(\bar{y}_i)) \leqq f(x) - f(\bar{x}) .$$

Hence $f(\bar{x}) \leqq f(x)$ for all $x \in N(\bar{x})$ which are feasible. ∎

2.12 Remark If it is further assumed that $\psi''(0) > 0$, which is the case in Rockafellar's Lagrangian, then the strict complementarity requirement at $(\bar{x},\bar{u})$ for theorem 2.9 above can be slightly weakened to the following: $\bar{u}_i > 0$ for $i=1,\ldots,m$ for which $g_i(\bar{x}) = 0$.

2.13 Remark We observe that $L(x,y,\alpha)$ is concave in y for each fixed α and fixed feasible x if we assume that $\psi''(0) = 0$ and $\psi''(\zeta)$ is non-decreasing for $\zeta \geqq 0$. This follows from the facts that

$\frac{\partial^2 L}{\partial y_i \partial y_j}(x,y,\alpha) = 0$ for $i \neq j$, and that for $g_i(x) \leqq 0, i=1,\ldots,m$

$$\begin{aligned}\frac{\partial^2}{\partial y_i^2}L(x,y,\alpha) &= \psi''(\alpha g_i(x)+y_i)_+ - \psi''(y_i) \\ &\leqq \psi''(\alpha g_i(x)+y_i)_+ - \psi''(y_i)_+ \leqq 0\end{aligned}$$

and for $g_i(x) = 0, i=m+1,\ldots,k$

$$\frac{\partial^2}{\partial y_i^2}L(x,y,\alpha) = \psi''(\alpha g_i(x)+y_i) - \psi''(y_i) = 0$$

It was also shown in the proof of theorem 2.9a above that if $f, g_i, i=1,\ldots,m$ are convex and $g_i, i=m+1,\ldots,k$ are affine then L is convex in x for each fixed y and α.

3. DUALITY

We observe first that as a consequence of lemma 2.7 the primal problem 1.1 is equivalent to

3.1
$$\underset{x,y}{\text{minimize}} \quad L(x,y)$$
$$\text{subject to} \quad \nabla_2 L(x,y,\alpha) = 0$$

where L is defined by 2.1 or more specifically by 2.4 and α is any positive number. We shall associate with this problem the following dual problem

3.2
$$\underset{x,y}{\text{maximize}} \quad L(x,y)$$
$$\text{subject to} \quad \nabla_1 L(x,y,\alpha) = 0$$

We shall assume no convexity in many of the following results, and hence the standard techniques of deriving duality results such as the use of minmax theorem [26,10,27] will not apply, nor will the elegant conjugate function theory of Rockafellar [20] apply directly, however see [24] .

The results of this section consist of a weak duality-theorem 3.3 (for which convexity is needed), a duality theorem 3.4 which relates a Kuhn-Tucker point of 1.1 to a Kuhn-Tucker point of the dual problem 3.2 and to a second order maximum of 3.2 under no convexity assumptions and finally to a global solution of 3.2 under convexity. The converse duality theorem 3.10 similarly relates a local solution of the dual problem 3.2 to a Kuhn-Tucker point of the primal problem 1.1 and to a second order minimum under no convexity assumptions and finally to a global solution of 1.1 under convexity.

Probably the most important features of these duality theorems are the absence of inequality constraints from the dual problem 3.2 and the relations between second order optima of the dual problems obtained in theorem 3.4 and 3.10 without any convexity assumptions. Related local results for a specific L have also been given by Buys [3].

3.3 Weak duality theorem Let $\hat{x}$ satisfy the constraints of the primal problem 1.1, or equivalently let $(\hat{x},\hat{y})$ satisfy the constraints of 3.1, let (x,y) satisfy the constraints of the dual problem 3.2, and let $f, g_i, i=1,\ldots,m$ be differentiable and convex on R^n, and let $g_i, i=m+1,\ldots,k$, be affine functions. Then $f(\hat{x}) \geqq L(x,y)$. ∎

Proof

$$f(\hat{x}) \geqq f(x) + \nabla f(x)(\hat{x}-x) \qquad \text{(by convexity of } f\text{)}$$

$$= f(x) - \sum_{i=1}^{m} \alpha\psi'(\alpha g_i(x)+y_i)_+ \nabla g_i(x)(\hat{x}-x) - \sum_{i=m+1}^{k} \alpha\psi'(\alpha g_i(x)+y_i)\nabla g_i(x)(\hat{x}-x)$$

$$\text{(since } \nabla_1 L(x,y,\alpha) = 0\text{)}$$

$$\geqq f(x) + \sum_{i=1}^{m} \alpha\psi'(\alpha g_i(x)+y_i)_+ (g_i(x)-g_i(\hat{x})) + \sum_{i=m+1}^{k} \alpha\psi'(\alpha g_i(x)+y_i)(g_i(x)-g_i(\hat{x}))$$

(by convexity of $g_i, i=1,\ldots,m$, and affineness of $g_i, i=m+1,\ldots,k$)

$$\geqq f(x) + \sum_{i=1}^{m} \alpha\psi'(\alpha g_i(x)+y_i)_+ g_i(x) + \sum_{i=m+1}^{k} \alpha\psi'(\alpha g_i(x)+y_i)g_i(x)$$

(by primal feasibility of $\hat{x}$)

$$\geqq f(x) + \sum_{i=1}^{m} (\psi(\alpha g_i(x)+y_i)_+ - \psi(y_i)_+) + \sum_{i=m+1}^{k} (\psi(\alpha g_i(x)+y_i) - \psi(y_i))$$

(by convexity of $\psi(\ )_+$ and ψ)

$$= L(x,y,\alpha) .$$ ∎

3.4 <u>Duality theorem</u> Let $f, g_i, i=1,\dots,k$ be differentiable at $\bar{x}$.

(a) If $(\bar{x},\bar{u})$ is a Kuhn-Tucker point of 1.1, and if either strict complementarity holds at $(\bar{x},\bar{u})$ or $\psi_i''(0) = 0$ for $i \in I$, then $(\bar{x},\bar{y})$ defined by 2.6 satisfies the following Kuhn-Tucker conditions of the dual problem 3.2

3.5
$$\begin{cases} \nabla_2 L(\bar{x},\bar{y},\alpha) + \bar{v}\nabla_{12} L(\bar{x},\bar{y},\alpha) = 0 \\ \nabla_1 L(\bar{x},\bar{y},\alpha) + \bar{v}\nabla_{11} L(\bar{x},\bar{y},\alpha) = 0 \\ \nabla_1 L(\bar{x},\bar{y},\alpha) = 0 \end{cases}$$

with $\bar{v} = 0$.

(b) If $f, g_i, i=1,\dots,k$ are twice differentiable at $\bar{x}$, if the second order sufficiency conditions and strict complementarity hold at $(\bar{x},\bar{u})$, if $\nabla g_i(\bar{x})$, $i \in I \cup E$ are linearly independent, then for sufficiently large α, $(\bar{x},\bar{y})$ determined by 2.6 forms an isolated local maximum of the dual problems 3.2 if $\psi_i''(0) > 0$ for $i \in I \cup J$. If $\psi_i''(0) = 0$ for $i \in I \cup J$ then $(\bar{x},\bar{y})$ forms an isolated local maximum of 3.2 subject to the additional constraints that $y_i = 0$, $i \in J$.

(c) If in addition to the assumptions of part a above, $g_i, i=1,\dots,m$ are differentiable and convex on R^n and $g_i, i=m+1,\dots,k$ are affine, then $(\bar{x},\bar{y})$ solves the dual problem 3.2 and the extrema $f(\bar{x})$ and $L(\bar{x},\bar{y},\alpha)$ are equal. ∎

<u>Proof</u>:

(a) By theorem 2.5 we have that $\nabla_1 L(\bar{x},\bar{y},\alpha) = 0$ and $\nabla_2 L(\bar{x},\bar{y},\alpha) = 0$ which are the Kuhn-Tucker conditions 3.5 with $\bar{v} = 0$. Strict complementarity or $\psi_i''(0) = 0$ for $i \in I$ are imposed here so that the second derivatives of 3.5 are well defined.

(b) By part a of this theorem, $(\bar{x},\bar{y})$ satisfies the Kuhn-Tucker conditions 3.5 of the dual problem 3.2 with $\bar{v} = 0$. To show that $(\bar{x},\bar{y})$ is an isolated local maximum of 3.2 we need to show that the second order sufficiency conditions for 3.2 are satisfied at $(\bar{x},\bar{y})$, that is

$$3.6 \qquad (\nabla_{11}L(\bar{x},\bar{y},\alpha) \quad \nabla_{12}L(\bar{x},\bar{y},\alpha)) \begin{pmatrix} x \\ y \end{pmatrix} = 0 \qquad (x \quad y) \neq 0$$

implies that

$$3.7 \qquad (x \quad y)\ \nabla^2 L(\bar{x},\bar{y},\alpha) \begin{pmatrix} x \\ y \end{pmatrix} < 0$$

where

$$3.8 \quad \nabla^2 L(\bar{x},\bar{y},\alpha) = \begin{bmatrix} \nabla_{11}L(\bar{x},\bar{y},\alpha) & \nabla_{12}L(\bar{x},\bar{y},\alpha) \\ \nabla_{21}L(\bar{x},\bar{y},\alpha) & \nabla_{22}L(\bar{x},\bar{y},\alpha) \end{bmatrix} = \begin{bmatrix} \nabla_{11}L(\bar{x},\bar{y},\alpha) & \underset{i \in I \cup E}{\alpha\psi''(\bar{y}_i)\nabla g_i(\bar{x})} & 0 \\ \underset{i \in I \cup E}{\alpha\psi''(\bar{y}_i)\nabla g_i(\bar{x})} & 0 & 0 \\ 0 & 0 & \underset{i \in J}{-\psi_i''(0)} \end{bmatrix}$$

But for (x,y) satisfying 3.6 we have that

$$3.9 \qquad (x \quad y)\nabla^2 L(\bar{x},\bar{y},\alpha) \begin{pmatrix} x \\ y \end{pmatrix} = -x\nabla_{11}L(\bar{x},\bar{y},\alpha)x + y\nabla_{22}L(\bar{x},\bar{y},\alpha)y$$

For the case where $\psi_i''(0) > 0$ for $i \in I \cup J$, we obtain the negativity of 3.9 for $(x,y_J) \neq 0$ from the positive definiteness of $\nabla_{11}L(\bar{x},\bar{y},\alpha)$ for large α (by theorem 2.9a) and from $-\psi_J''(0) < 0$. The case $(x,y_J) = 0$, is excluded because by 3.6 and 3.8, $y_{IUE} \neq 0$ and

$$\sum_{i\in I\cup E} \alpha\psi''(\bar{y}_i)\nabla g_i(\bar{x})y_i = 0$$

which contradicts the linear independence of $\nabla g_i(\bar{x})$, $i \in I \cup E$, since by strict complementarity $\psi''(\bar{y}_i) > 0$ for $i \in I \cup E$. We have thus established that 3.6 implies 3.7 if $\psi_i''(0) > 0$ for $i \in I \cup J$. When $\psi_i''(0) = 0$ for $i \in I \cup J$ we establish that 3.6 implies 3.7 under the added assumption that $y_J = 0$. For this case 3.9 becomes

$$(x \quad y)\nabla^2 L(\bar{x},\bar{y},\alpha)\begin{pmatrix} x \\ y \end{pmatrix} = -x\nabla_{11}L(\bar{x},\bar{y},\alpha)x$$

which is negative for $x \neq 0$ by the positive definiteness of $\nabla_{11}L(\bar{x},\bar{y},\alpha)$ for large α. The case $x = 0$ is excluded because then $y_{IUE} \neq 0$ and by 3.6 and 3.8

$$\sum_{i\in I\cup E} \alpha\psi''(\bar{y}_i)\nabla g_i(\bar{x})y_i = 0$$

which contradicts the linear independence of $\nabla g_i(\bar{x})$, $i \in I \cup E$.

(c) By part a of this theorem $(\bar{x},\bar{y})$ determined from 2.6 satisfies 3.5 with $\bar{v} = 0$. Hence $\nabla_1 L(\bar{x},\bar{y},\alpha) = 0$ and $(\bar{x},\bar{y})$ is a feasible point for the dual problem 3.2. For any dual feasible point (x,y) we have by the weak duality theorem 3.3 that $f(\bar{x}) \geqq L(x,y,\alpha)$. But

$$L(\bar{x},\bar{y},\alpha) = f(\bar{x}) + \sum_{i\in I}(\psi(\bar{y}_i)_+ - \psi(\bar{y}_i)) + \sum_{i\in J}(\psi(\alpha g_i(\bar{x}))_+ - \psi(0)) + \sum_{i=m+1}^{k}(\psi(\bar{y}_i) - \psi(\bar{y}_i))$$

$$= f(\bar{x}) + \sum_{i\in I}(\psi(\bar{y}_i) - \psi(\bar{y}_i)) \qquad (\text{since } \bar{y}_i \geqq 0,\ i\in I, \text{ and } g_i(\bar{x}) < 0,\ i \in J)$$

$$= f(\bar{x})$$

Hence $L(\bar{x},\bar{y},\alpha) = f(\bar{x}) \geqq L(x,y,\alpha)$ for any dual feasible point (x,y), and $(\bar{x},\bar{y})$ is a global solution of 3.2. ∎

3.10 Converse duality theorem Let $(\bar{x},\bar{y})$ be a local or global solution of the dual problem 3.2, let $f, g_i, i=1,\ldots,k$ be twice continuously differentiable at $\bar{x}$, and let either $\bar{y}_i > 0$ for $i \in I$ or $\psi_i''(0) = 0$ for $i \in I$.

(a) If the matrix $\nabla_{11}L(\bar{x},\bar{y},\alpha)$ is nonsingular then $\bar{x}$ and $\bar{u} \in R^k$ determined by 2.8 satisfy the Kuhn-Tucker conditions for the primal problem 1.1.

(b) If in addition $\nabla_{11}L(\bar{x},\bar{y},\alpha)$ is positive definite and $\bar{y}_i > 0$ for $i \in I$, then $\bar{x}$ and $\bar{u} \in R^k$ determined by 2.8 satisfy the second order sufficient optimality conditions for the primal problem 1.1 .

(c) If in addition to the assumption of part a above, f is convex or pseudoconvex at $\bar{x}$, $g_i, i=1,\ldots,m$, are convex or quasiconvex at $\bar{x}$, and $g_i, i=m+1,\ldots,k$, are affine or simultaneously quasiconvex and quasiconcave at $\bar{x}$, then $\bar{x}$ is a global solution of the primal problem 1.1 . ■

Proof:

(a) Since $(\bar{x},\bar{y})$ is a solution of the dual problem 3.2, $(\bar{x},\bar{y})$ and some $(\bar{v}_0,\bar{v}) \in R \times R^n$ satisfy the following Fritz John conditions [12, p. 170]

3.11
$$\begin{cases} \bar{v}_0\nabla_1 L(\bar{x},\bar{y},\alpha) + \bar{v}\nabla_{11}L(\bar{x},\bar{y},\alpha) = 0 \\ \bar{v}_0\nabla_2 L(\bar{x},\bar{y},\alpha) + \bar{v}\nabla_{12}L(\bar{x},\bar{y},\alpha) = 0 \\ \nabla_1 L(\bar{x},\bar{y},\alpha) = 0 \\ (\bar{v}_0,\bar{v}) \neq 0 \end{cases}$$

From the first and third equations above and the nonsingularity of $\nabla_{11}L(\bar{x},\bar{y},\alpha)$ it follows that $\bar{v} = 0$ and hence $\bar{v}_0 \neq 0$. So $\nabla_1 L(\bar{x},\bar{y},\alpha) = 0$ and $\nabla_2 L(\bar{x},\bar{y},\alpha) = 0$ and by theorem 2.5, $\bar{x}$ and $\bar{u}$ defined by 2.8 satisfy the Kuhn-Tucker conditions for 1.1 .

(b) By part a above $\bar{x}$ and $\bar{u}$ determined by 2.8 satisfy the Kuhn-Tucker conditions for 1.1. As in the proof theorem 2.9 we have since $\bar{y}_i > 0$ for $i \in I$ that

3.12 $$\nabla_{11}L(\bar{x},\bar{y},\alpha) = \nabla_{11}L^0(\bar{x},\bar{u}) + \sum_{i\in I\cup E} \alpha^2\psi''(\bar{y}_i)\nabla g_i(\bar{x})\nabla g_i(\bar{x})^T$$

where L^0 is the standard Lagrangian. We establish now the implication

3.13 $$\langle\nabla g_i(\bar{x}) = 0, x\neq 0, i\in I\cup E\rangle \implies x\nabla_{11}L^0(\bar{x},\bar{u})x > 0$$

For if not, then for some $\hat{x} \neq 0$, $\nabla g_i(\bar{x})\hat{x} = 0$, $i \in I\cup E$ and $\hat{x}\nabla_{11}L^0(\bar{x},\bar{u})\hat{x} \leqq 0$ which by 3.12 gives that $\hat{x}\nabla_{11}L(\bar{x},\bar{y},\alpha)\hat{x} \leqq 0$ which contradicts the positive definiteness of $\nabla_{11}L(\bar{x},\bar{y},\alpha)$. Implication 3.13, which because of 2.8 and $\bar{y}_i > 0$ for $i \in I$, is the second order sufficient optimality condition for 1.1.

(c) This part follows from the sufficiency theorem of the Kuhn-Tucker conditions [12, theorem 2, p. 162]. ∎

4. LOCAL COMPUTATIONAL ALGORITHMS

We shall present in this section two local algorithms for the solution of problem 1.1 which are based on reducing the problem 1.1 to that of finding solutions of the $n + k$ nonlinear equations $\nabla L(x,y,\alpha) = 0$. The first algorithm 4.7 is a Newton algorithm for which we establish, under suitable conditions, local superlinear or quadratic convergence rates. The second method is an extension of the method of multipliers investigated by Arrow-Solow [2], Hestenes [8,9] Powell [19] Haarhoff and Buys [7] and Miele, Moseley and Cragg [14,15] for the case of equality constraints. Our extension is to inequality constraints and to a general Lagrangian. We establish linear convergence for the algorithm and indicate under what sort of conditions we may expect fast or slow convergence of the method. In [3] Buys gives, for a specific Lagrangian, a dual algorithm which is related to our stationary point problem. One specific implementation of his algorithm, for equality constraints only, turns out to be the method of multipliers [8,9] and for which he establishes local convergence. For inequalities however, a particular case of his algorithm gives a relative to a special case of our algorithm 4.9. He does not however establish convergence nor a rate of convergence for that algorithm.

In both of the computational methods to be considered here, in order to establish convergence we need to have a nonsingular Hessian at the solution point. For that purpose we shall employ another Lagrangian $M(x,y,\alpha)$ which is obtained by augmenting $L(x,y,\alpha)$ in such a way that $\nabla^2 M(\bar{x},\bar{y},\alpha)$ is nonsingular and $\nabla M(\bar{x},\bar{y},\alpha) = 0$ is equivalent to $\nabla L(\bar{x},\bar{y},\alpha) = 0$. (The feasibility of augmenting L to obtain another Lagrangian M which has identical stationary points as L, presents the intriguing possibility of generating a still wider class of unconstrained Lagrangians with possibly better properties than L. This possibility has not been investigated in depth here.) We define

4.1 $$M(x,y,\alpha) = L(x,y,\alpha) - \sum_{i=1}^{m} y_i^2 \, \psi(-g_i(x))_+$$
$$= f(x) + \sum_{i=1}^{m} (\psi(\alpha g_i(x)+y_i)_+ - \psi(y_i) - y_i^2 \, \psi(-g_i(x))_+)$$
$$+ \sum_{i=m+1}^{k} (\psi(\alpha g_i(x)+y_i) - \psi(y_i))$$

We establish now immediately the equivalence of stationary points of L and M.

4.2 Lemma $\nabla L(\bar{x},\bar{y},\alpha) = 0 \iff \nabla M(\bar{x},\bar{y},\alpha) = 0$ ∎

Proof ($\Longrightarrow$) By lemma 2.7 we have that $\bar{y}_i g_i(\bar{x}) = 0$, $i=1,\ldots,m$. Hence

$$\nabla_1 M(\bar{x},\bar{y},\alpha) = \nabla_1 L(\bar{x},\bar{y},\alpha) + \sum_{i=1}^{m} \bar{y}_i^2 \, \psi'(-g_i(\bar{x}))_+ \nabla g_i(\bar{x}) = 0$$

$$\frac{\partial}{\partial y_i} M(\bar{x},\bar{y},\alpha) = \frac{\partial}{\partial y_i} L(\bar{x},\bar{y},\alpha) - 2\bar{y}_i \, \psi(-g_i(\bar{x}))_+ = 0, \quad i=1,\ldots,m$$

$$\frac{\partial}{\partial y_i} M(\bar{x},\bar{y},\alpha) = \frac{\partial}{\partial y_i} L(\bar{x},\bar{y},\alpha) = 0, \qquad i=1,\ldots,k$$

($\Longleftarrow$) We will first show that $\nabla_2 L(\bar{x},\bar{y},\alpha) = 0$.

$$\frac{\partial M}{\partial y_i}(\bar{x},\bar{y},\alpha) = \frac{\partial L}{\partial y_i}(\bar{x},\bar{y},\alpha) - 2\bar{y}_i\psi(-g_i(\bar{x}))_+$$

$$= \psi'(\alpha g_i(\bar{x})+\bar{y}_i)_+ - \psi'(\bar{y}_i) - 2\bar{y}_i\psi(-g_i(\bar{x}))_+ = 0 \ , \ i=1,\ldots,m$$

We will now show that $\bar{y}_i\psi(-g_i(\bar{x}))_+ = 0$ for $i=1,\ldots,m$, and hence $\nabla_2 L(\bar{x},\bar{y},\alpha) = 0$. Suppose not, then for some i , $\bar{y}_i \neq 0$ and $g_i(\bar{x}) < 0$. Two cases can arise.

Case 1: $\left\langle \bar{y}_i < 0 \ , \ g_i(\bar{x}) < 0 \right\rangle \implies \left\langle \begin{matrix} \psi'(\alpha g_i(\bar{x}) + \bar{y}_i)_+ = 0 \\ \psi'(\bar{y}_i) < 0 \\ 2\bar{y}_i\psi(-g_i(\bar{x}))_+ < 0 \end{matrix} \right\rangle$

$$\implies \left\langle \psi'(\alpha g_i(\bar{x}) + \bar{y}_i)_+ - \psi'(\bar{y}_i) - 2\bar{y}_i\psi(-g_i(\bar{x}))_+ > 0 \right\rangle \implies$$

Contradicts $\frac{\partial M}{\partial y_i}(\bar{x},\bar{y},\alpha) = 0$

Case 2: $\left\langle \bar{y}_i > 0 \ , \ g_i(\bar{x}) < 0 \right\rangle \implies \left\langle \begin{matrix} \psi'(\alpha g_i(\bar{x}) + \bar{y}_i)_+ < \psi'(\bar{y}_i) \\ 2\bar{y}_i \ \psi(-g_i(\bar{x}))_+ > 0 \end{matrix} \right\rangle$

$$\implies \left\langle \psi'(\alpha g_i(\bar{x})+\bar{y}_i)_+ - \psi'(\bar{y}_i) - 2\bar{y}_i\psi(-g_i(\bar{x}))_+ < 0 \right\rangle \implies$$

Contradicts $\frac{\partial M}{\partial y_i}(\bar{x},\bar{y},\alpha) = 0$

Hence $\nabla_2 L(\bar{x},\bar{y},\alpha) = 0$. By lemma 2.7 we have that $\bar{y}_i g_i(\bar{x}) = 0$, $i=1,\ldots,m$, and hence

$$\nabla_1 L(\bar{x},\bar{y},\alpha) = \nabla_1 M(\bar{x},\bar{y},\alpha) + \sum_{i=1}^{m} \bar{y}_i^{\,2}\psi'(-g_i(\bar{x}))_+\nabla g_i(\bar{x}) = 0 \quad \blacksquare$$

We establish next the positive definiteness of $\nabla_{11}M(\bar{x},\bar{y},\alpha)$, the nonsingularity of $\nabla^2 M(\bar{x},\bar{y},\alpha)$ and the saddlepoint property of M.

4.3 Lemma (a) Let the assumptions of theorem 2.9a hold. Then for sufficiently large but finite α, $\nabla_{11}M(\bar{x},\bar{y},\alpha)$ is positive definite, $\nabla^2 M(\bar{x},\bar{y},\alpha)$ is nonsingular and

4.4 $$M(\bar{x},y,\alpha) \leqq M(\bar{x},\bar{y},\alpha) < M(x,\bar{y},\alpha) \qquad \forall y \in R^k, \ \forall x \in N(\bar{x}), \ x \neq \bar{x}$$

where $N(\bar{x})$ is some open neighborhood of $\bar{x}$.

(b) Conversely, if 4.4 holds, with $<$ possibly replaced by $\leqq$, then $\bar{x}$ is a solution of 1.1 subject to the extra restriction that $x \in N(\bar{x})$. ∎

Proof (a) As in the proof 2.9a we have that

$$\nabla_{11}M(\bar{x},\bar{y},\alpha) = \nabla_{11}L(\bar{x},\bar{y},\alpha) - \sum_{i\in I} \bar{y}_i^{\,2}\,\psi''(0)\nabla g_i(\bar{x})\nabla g_i(\bar{x})^T$$
$$= \nabla_{11}L^0(\bar{x},\bar{u}) + \sum_{i\in I} (\alpha^2\psi''(\bar{y}_i) - \bar{y}_i^{\,2}\psi''(0))\nabla g_i(\bar{x})\nabla g_i(\bar{x})^T$$
$$+ \sum_{i\in E} \alpha^2\psi''(\bar{y}_i)\nabla g_i(\bar{x})\nabla g_i(\bar{x})^T$$

It follows again by strict complementarity, the second order sufficiency conditions and Debreu's theorem that for large enough α, $\nabla_{11}M(\bar{x},\bar{y},\alpha)$ is positive definite and hence the second inequality of 4.4 holds. The first inequality of 4.4 holds because

$$M(\bar{x},y,\alpha) \leqq L(\bar{x},y,\alpha) \leqq L(\bar{x},\bar{y},\alpha) = M(\bar{x},\bar{y},\alpha)$$

To show that $\nabla^2 M(\bar{x},\bar{y},\alpha)$ is nonsingular we observe that

$$4.5 \quad \nabla^2 M(\bar{x},\bar{y},\alpha) = \begin{bmatrix} \nabla_{11} M(\bar{x},\bar{y},\alpha) & \underset{i\in I\cup E}{\alpha\psi''(\bar{y}_i)\nabla g_i(\bar{x})} & 0 \\ \underset{i\in I\cup E}{\alpha\psi''(\bar{y}_i)\nabla g_i(\bar{x})} & 0 & 0 \\ 0 & 0 & \underset{i\in J}{-\psi_i''(0)-2\psi_i(-g_i(\bar{x}))_+} \end{bmatrix}$$

The nonsingularity of $\nabla^2 M(\bar{x},\bar{y},\alpha)$ for large α follows from the positive definiteness of $\nabla_{11} M(\bar{x},\bar{y},\alpha)$, the linear independence of $\nabla g_i(\bar{x})$, $\psi''(\bar{y}_i) > 0$, $i\in I\cup E$, and $-\psi_i''(0) - 2\psi_i(-g_i(\bar{x}))_+ < 0$, $i \in J$.

(b) From 4.4 we have that $\nabla M(\bar{x},\bar{y},\alpha) = 0$. By lemma 4.2 it follows that $\nabla L(\bar{x},\bar{y},\alpha) = 0$ and by lemma 2.7 we have that $\bar{y}_i g_i(\bar{x}) = 0$, $i=1,\ldots,m$. Hence $M(\bar{x},\bar{y},\alpha) = L(\bar{x},\bar{y},\alpha)$. From 4.4 we also have that for $x\in N(\bar{x})$

$$0 \leq M(x,\bar{y},\alpha) - M(\bar{x},\bar{y},\alpha) \leq L(x,\bar{y},\alpha) - L(\bar{x},\bar{y},\alpha)$$

The rest of the proof is identical to that of 2.9b. ∎

We observe that if in 4.1, $\psi''(0) = 0$ for the ψ function explicitly stated therein and for the ψ_i, $i \in \{1,\ldots,m\}$ associated with the inequality constraints then M is globally twice differentiable provided that f and g_i, $i=1,\ldots,k$ are also twice differentiable. To be specific we state explicitly a recommended globally twice differentiable M function associated with problem 1.1

$$4.6 \qquad M(x,y,\alpha) = f(x)+\frac{1}{4\alpha}\sum_{i=1}^{m}\left((\alpha g_i(x)+y_i)_+^4 - y_i^4 - y_i^2(-g_i(x))_+^4\right) + \sum_{i=m+1}^{k}\left(\frac{\alpha}{2}g_i(x)^2 + y_i g_i(x)\right)$$

We are prepared now to state and establish the convergence and rates of convergence of our algorithms

4.7 Newton Algorithm Choose $\alpha > 0$ and $(x^0, y^0) \in R^n \times R^k$. Determine (x^{j+1}, y^{j+1}) from (x^j, y^j) as follows: Linearize $\nabla M(x,y,\alpha) = 0$ around the point (x^j, y^j) and solve for (x^{j+1}, y^{j+1}). ▮

4.8 Local convergence and rate of convergence of the Newton algorithm

(a) Let $(\bar{x}, \bar{u}) \in R^n \times R^k$ satisfy the Kuhn-Tucker conditions of 1.1, let f and g_i, $i=1,\ldots,k$, be twice continuously differentiable at each point of an open neighborhood of $\bar{x}$, let $\nabla g_i(\bar{x})$, $i \in I \cup E$ be linearly independent and let the assumptions of theorem 2.9a hold. Then for large enough but finite α there exists an open neighborhood $N(\bar{x}, \bar{y})$ of $(\bar{x}, \bar{y})$ in $R^n \times R^k$, where $\bar{y}$ is determined from $\bar{u}$ by 2.6, such that for every $(x^0, y^0) \in N(\bar{x}, \bar{y})$, the Newton algorithm iterates of 4.2 are well defined and converge superlinearly to $(\bar{x}, \bar{y})$ in the sense that

$$\lim_{j \to \infty} \frac{\|z^{j+1} - \bar{z}\|}{\|z^j - \bar{z}\|} = 0$$

where $z = (x \;\; y)$.

(b) If in addition f and g_i, $i=1,\ldots,k$ are three times differentiable on $N(\bar{x}, \bar{y})$, ψ is three times differentiable on R and $\psi_i'''(0) = 0$ for $i \in \{1,\ldots,m\}$, then the Newton iterates converge quadratically to $(\bar{x}, \bar{y})$, that is for some constant d and some integer j_0 depending on z^0

$$\|z^{j+1} - \bar{z}\| \leq d\|z^j - \bar{z}\|^2 \quad \text{for } j \geq j_0$$ ▮

Proof Since $(\bar{x},\bar{y}) \in R^n \times R^k$ is obtained from $(\bar{x},\bar{u})$ by 2.6 it follows by theorem 2.5 that $\nabla L(\bar{x},\bar{y},\alpha) = 0$, by lemma 4.2 that $\nabla M(\bar{x},\bar{y},\alpha) = 0$, and by lemma 4.3 a that $\nabla^2 M(\bar{x},\bar{y},\alpha)$ is nonsingular. The convergence properties stated in the theorem follow then from the local convergence theorem of Newton's method [16, p. 148]. ∎

We present now a second method which is an extension to inequality constraints and to more general Lagrangians of the method of multipliers. Originally this method was proposed for equality constraints by Arrow and Solow [2] by using differential equations to determine a small stepsize algorithm. Later and independently of Arrow and Solow and of each other Hestenes [8,9], Powell [19] and Haarhoff and Buys [7] used a similar Lagrangian approach for equality constraints and proposed a large stepsize method. Miele, Moseley and Cragg [14,15] made numerical tests of the algorithm and variants of it. More recently Buys [3] and Wierzbicki [28] considered extensions to inequality constraints. Buys suggested a dual problem approach for a specific Lagrangian function but did not give any convergence rates. Wierzbicki considers another specific but different Lagrangian.

4.9 Method of Lagrange multipliers Choose $\alpha > 0$, $\beta > 0$, $y^0 \in R^k$ and $x^0 \in R^n$ satisfying $\nabla_1 M(x^0, y^0, \alpha) = 0$. Having (x^j, y^j) determine (x^{j+1}, y^{j+1}) as follows

(a) Determine x^{j+1} such that $M(x^{j+1}, y^j, \alpha) = \underset{x \in R^n}{\text{minimum}}\ M(x, y^j, \alpha)$ or $\nabla_1 M(x^{j+1}, y^j, \alpha) = 0$. If x^{j+1} is not unique, take a closest x^{j+1} , in any norm, to x^j .

(b) $y^{j+1} = y^j + \beta\nabla_2 M(x^{j+1}, y^j, \alpha)$ ■

4.10 Local convergence and rate of convergence of the method of Lagrange multipliers Let the assumptions of 4.8a hold, and let $\bar{y}$ be determined from $\bar{u}$ by 2.6. Then for large but finite α, there exist open neighborhoods $N_0(\bar{x})$ and $N_0(\bar{y})$ in R^n and R^k respectively, such that for each $y^0 \in N_0(\bar{y})$ there exists a unique x^0 in the closure $\bar{N}_0(\bar{x})$ of $N_0(\bar{x})$ satisfying $\nabla_1 M(x^0, y^0, \alpha) = 0$ and such that the iterates of 4.9 are well defined and converge linearly to $(\bar{x}, \bar{y})$ for $\beta \in (0, \bar{\beta})$ for some $\bar{\beta} > 0$. ■

Proof As in the proof of theorem 4.8 we have that $\nabla M(\bar{x}, \bar{y}, \alpha) = 0$ by theorem 2.5 and lemma 4.2. By lemma 4.3a $\nabla_{11} M(\bar{x}, \bar{y}, \alpha)$ is positive definite for sufficiently large but finite α. It follows by the implicit function theorem [17, p. 128] that for some open neighborhoods $N(\bar{y})$ in R^k and $N(\bar{x})$ in R^n, there exists a function $e: R^k \to R^n$ which is continuously differentiable on $N(\bar{y})$ and such that

4.11 For $y \in \bar{N}(\bar{y})$, $x = e(y)$ is a unique solution of $\nabla_1 M(x, y, \alpha) = 0$ in $\bar{N}(\bar{x})$; $\bar{x} = e(\bar{y})$ and $e(y) \in \bar{N}(\bar{x})$

Define

4.12 $$N_0(\bar{x}) = N(\bar{x})$$

For $y^j \in \bar{N}(\bar{y})$ algorithm 4.9 is well defined and is equivalent to

4.13 $$y^{j+1} = y^j + \beta \nabla_2 M(e(y^j), y^j, \alpha)$$

if we assume for the time being that x^{j+1} of step 4.9a is unique.

Consider now the mapping $G:R^k \to R^k$ underlying the iteration 4.13 and defined by

4.14 $$G(y) = y + \beta \nabla_2 M(e(y), y, \alpha)$$

and its gradient evaluated at $\bar{y}$

4.15 $$\nabla G(\bar{y}) = I + \beta\nabla_{21} M(\bar{x},\bar{y},\alpha)\nabla e(\bar{y}) + \beta\nabla_{22} M(\bar{x},\bar{y},\alpha)$$

Differentiating $\nabla_1 M(e(y), y, \alpha) = 0$ with respect to y and evaluating at $\bar{y}$ gives

$$\nabla_{11} M(\bar{x},\bar{y},\alpha)\nabla e(\bar{y}) + \nabla_{12} M(\bar{x},\bar{y},\alpha) = 0$$

and hence

$$\nabla e(\bar{y}) = \nabla_{11} M(\bar{x},\bar{y},\alpha)^{-1}\nabla_{12} M(\bar{x},\bar{y},\alpha)$$

Substitution in 4.15 gives

4.16 $$\nabla G(\bar{y}) = I - \beta[\nabla_{21} M(\bar{x},\bar{y},\alpha)\nabla_{11} M(\bar{x},\bar{y},\alpha)^{-1}\nabla_{12} M(\bar{x},\bar{y},\alpha) - \nabla_{22} M(\bar{x},\bar{y},\alpha)]$$

By referring to 4.5 this expression can be rewritten as

4.17 $$\nabla G(\bar{y}) = I - \beta \begin{bmatrix} \underset{i \in I \cup E}{\alpha^2 \psi''(\bar{y}_i)^2 \nabla g_i(\bar{x}) \nabla_{11} M(\bar{x},\bar{y},\alpha)^{-1} \nabla g_i(\bar{x})} & 0 \\ 0 & \underset{i \in J}{\psi_i''(0) + 2\psi_i(-g_i(\bar{x}))_+} \end{bmatrix}$$

It follows from the linear independence of $\nabla g_i(\bar{x})$, $i \in I \cup E$, the positive definiteness of $\nabla_{11} M(\bar{x},\bar{y},\alpha)$ and $\psi_i''(0) + 2\psi_i(-g_i(\bar{x}))_+ > 0$, $i \in J$, that the matrix in the square brackets is positive definite. Hence for some $\bar{\beta} > 0$, the eigenvalues of $\nabla G(\bar{y})$ are less than one in magnitude for $\beta \in (0,\bar{\beta})$ and hence the spectral radius $\rho(\nabla G(\bar{y})) < 1$. It follows by Ostrowki's point of attraction [16, p. 145] that there exist open neighborhoods $N_1(\bar{y})$, $N_2(\bar{y})$ with $N_1(\bar{y}) \subset N_2(\bar{y}) \subset N(\bar{y})$ and such that when $y^0 \in N_1(\bar{y})$ the iterates of 4.13 remain in $N_2(\bar{y})$ and converge linearly to $\bar{y}$. Since e is differentiable on $N_2(\bar{y})$ the iterates $\{x^j\}$ defined by $x^j = e(y^{j-1})$ are also well defined and converge linearly to $\bar{x} = e(\bar{y})$ if we choose any $N_0(\bar{y}) \subset N_1(\bar{y})$. Hence the sequence $\{(x^j,y^j)\}$ converges linearly to $(\bar{x},\bar{y})$ and the theorem is established for the case when x^{j+1} of 4.9a is unique.

Suppose now that x^{j+1} of step 4.9a is not unique and that there also exists an $\hat{x}^{j+1} \neq x^{j+1} = e(y^j)$ such that $\nabla_1 L(\hat{x}^{j+1},y^j,\alpha) = 0$. We will show that x^{j+1} is closer than $\hat{x}^{j+1}$ to x^j and hence $\hat{x}^{j+1}$ will not appear in the sequence $\{x^j\}$ generated by the algorithm. We have from 4.11 that $\hat{x}^{j+1} \notin \bar{N}(\bar{x})$ and hence

$$4.18 \qquad ||\hat{x}^{j+1} - \bar{x}|| > \delta$$

where δ is the radius of some open ball $B_\delta(\bar{x})$ around $\bar{x}$ which is contained in $\bar{N}(\bar{x})$. It follows again by Ostrowski's point of attraction theorem that for the sequence $\{y^j\}$ obtained from 4.13 and starting with any $y^0 \in N_1(\bar{y})$.

$$4.19 \qquad \left\langle \begin{array}{l} ||y^j - \bar{y}|| \leq c\gamma^j ||y^0 - \bar{y}|| \\ ||y^{j+1} - y^j|| \leq 2c\gamma^j ||y^0 - \bar{y}|| \end{array} \right\rangle$$

where $\gamma = \rho(\nabla G(\bar{y})) + 2\varepsilon < 1$ for some $\varepsilon > 0$ and c is a positive constant depending on ε and the norm employed. Also since e is differentiable we also have for $x^j = e(y^{j-1})$ and $x^{j+1} = e(y^j)$ that

4.20 $$\left\langle \begin{array}{l} \|x^j-\bar{x}\| \leqq c'\gamma^j \|y^0-\bar{y}\| \\ \|x^{j+1}-x^j\| \leqq 2c'\gamma^j \|y^0-\bar{y}\| \end{array} \right\rangle$$

where c' is some positive constant. Define now

$$N_0(\bar{y}) = \{y \mid y \in N_1(\bar{y}),\ \ \|y-\bar{y}\| < \frac{\delta}{4c'}\} .$$

Hence for $y^0 \in N_0(\bar{y})$

4.21 $$\|x^j-\bar{x}\| < \frac{\delta}{2} \qquad \forall j \geqq 1$$

and

4.22 $$\|x^{j+1}-x^j\| < \frac{\delta}{2} \qquad \forall j \geq 1$$

It follows from 4.18 , 4.21 and 4.22 that

$$\|\hat{x}^{j+1}-x^j\| \geqq \|\hat{x}^{j+1}-\bar{x}\| - \|x^j-\bar{x}\| > \delta-\frac{\delta}{2} = \frac{\delta}{2} > \|x^{j+1}-x^j\|$$

Hence x^{j+1} is closer than $\hat{x}^{j+1}$ to x^j. So $x^{j+1} = e(y^j)$ will be picked up in step 4.9a rather than $\hat{x}^{j+1}$ and iteration 4.13 will again represent algorithm 4.9 for the nonunique case also. The remainder of the proof is the same as for the unique case. ∎

We make some remarks here about the relation between the size of β and the speed of convergence. Since the size of β was determined from the requirement that some norm of $\nabla G(\bar{y})$ as given by 4.17 is less than one, it follows from 4.17 and this requirement that for some $\delta \in (0,1)$

$$\delta \geqq \| \nabla G(\bar{y}) \| \geqq 1 - \beta \nu$$

and

$$\delta \geqq \| \nabla G(\bar{y}) \| \geqq \beta \nu - 1$$

where ν is the norm of the matrix in the square brackets of 4.17. Hence

$$\frac{2}{\nu} > \frac{1+\delta}{\nu} \geqq \beta \geqq \frac{1-\delta}{\nu} \qquad \text{for some } \delta \in (0,1)$$

Fast convergence is obtained when δ is close to zero and hence $\beta \cong \frac{1}{\nu}$. On the other hand when $\beta \cong 0$ or $\beta \cong \frac{2}{\nu}$, δ will be close to 1 and hence slow convergence is to be expected.

Another possible source of slow convergence is the condition number of $\nabla_{11} M(\bar{x},\bar{y},\alpha)$ which affects step a of algorithm 4.9. It can be shown that if $\bar{k} < n$ where $\bar{k}$ is the number of active inequality and equality constraints (that is the solution $\bar{x}$ of 1.1 lies on a manifold) then the condition number of $\nabla_{11} M(\bar{x},\bar{y},\alpha)$ approaches ∞ as a α approaches ∞. However for the special case when $\bar{k} = n$ (that is the solution $\bar{x}$ lies on a "vertex") then the condition number of $\nabla_{11}M(\bar{x},\bar{y},\alpha)$ remains finite when α approaches ∞ . On the other hand when α is not large enough, $\nabla_{11}M(\bar{x},\bar{y},\alpha)$ may not be positive definite then and again it may be difficult to implement step a of algorithm 4.9. In summary it may be stated that in general convergence problems may be expected for small values of either α and β and also for large values of either α and β . Best results should be at intermediate values of α and β . Numerical results of Miele, Mosley and Cragg [14, table 2, examples 6.2 and 6.3] where $\beta = \alpha$ (k in their notation) slow convergence occurred for both small and large values of β and fast convergence occurred for intermediate values of β .

In conclusion we make some remarks about possible enlargement of the region of convergence. Basically we are finding an unconstrained saddlepoint of $M(x,y,\alpha)$ or a root of $\nabla M(x,y,\alpha) = 0$. Since there are no global methods for solving these problems in the absence of convexity, there is little hope in the present state of the art for a foolproof global algorithm for solving our problem here. However practical improvements may be achieved by introducing a stepsize which is determined by either minimizing the function $||\nabla M(x,y,\alpha)||^2$ or using an Armijo procedure [17, p. 491] on the same function along the direction $(x^{j+1}-x^j, y^{j+1}-y^j)$ or along a related direction such that $||\nabla M(x,y,\alpha)||^2$ decreases sufficiently along that direction. Such a procedure would lead to a point satisfying $\nabla^2 M(\bar{x},\bar{y},\alpha)\nabla M(\bar{x},\bar{y},\alpha) = 0$. If $\nabla M(\bar{x},\bar{y},\alpha) = 0$ we are done, if not the procedure would have to be restarted again either at $(\bar{x},\bar{y})$ or elsewhere.

Another acceleration procedure is to ignore all ε-inactive inequality constraints, that is for some $\varepsilon > 0$ delete at iteration j the inequality constraints such that $g_i(x^j) < -\varepsilon$, $i \in \{1,\ldots,m\}$, from consideration. It can be shown [13, lemma A.17] that for some neighborhood of $\bar{x}$ and for some ε such a procedure would not remove any of the active constraints g_i, $i \in I$ from the problem.

ACKNOWLEDGEMENT

I am indebted to my student S. P. Han for valuable discussion and for his reading of the manuscript.

REFERENCES

1. K. J. Arrow, F. J. Gould & S. M. Howe: "A general saddle point result for constrained optimization," Department of Statistics, Chapel Hill, North Carolina, Mimeo Series No. 774, September 1971, Mathematical Programming 5, 1973, 225-234.

2. K. J. Arrow & R. M. Solow: "Gradient methods for constrained maxima with weakened assumptions," in "Studies in linear and nonlinear programming," K. J. Arrow, L. Hurwicz & H. Uzawa (ed.), Stanford University Press, Stanford, California, 1958.

3. J. D. Buys: "Dual algorithms for constrained optimization problems," Doctorate Dissertation, University of Leiden, June 1972.

4. G. Debreu: "Definite and semidefinite quadratic forms," Econometrica 20, 1952, 295-300.

5. A. V. Fiacco & G. P. McCormick: "Nonlinear programming: Sequential unconstrained minimization techniques," Wiley, New York, 1968.

6. T. N. E. Greville (ed.): "Theory and applications of spline functions," Academic Press, New York, 1969.

7. P. C. Haarhoff & J. D. Buys: "A new method for the optimization of a nonlinear function subject to nonlinear constraints," Computer J. 13, 1970, 178-184.

8. M. R. Hestenes: "Multiplier and gradient methods," in "Computing methods in optimization problems," Volume 2, L. A. Zadeh, L. W. Neustadt & A. V. Balakrishnan (eds.), Academic Press, New York, 1969, 143-164.

9. M. R. Hestenes: "Multiplier and gradient methods," J. Optimization Theory and Applications, 4, 1969, 303-320.

10. S. Karamardian: "Strictly quasi-convex (concave) functions and duality in mathematical programming," J. Math. Anal. Appl. 20, 1967, 344-358.

11. H. W. Kuhn & A. W. Tucker: "Nonlinear programming," in "Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability," J. Neyman (ed.), University of California Press, Berkeley, California, 1951, 481-492.

12. O. L. Mangasarian: "Nonlinear programming," McGraw-Hill, New York, 1969.

13. O. L. Mangasarian: "Unconstrained Lagrangians in Nonlinear Programming," University of Wisconsin, Computer Sciences Department Report No. 174, March 1973.

14. A. Miele, P. E. Moseley & E. E. Cragg: "Numerical experiments on Hestenes' method of multipliers for mathematical programming problems," Rice University, Aero-Astronautics Report No. 85, 1971.

15. A. Miele, P. E. Moseley & E. E. Cragg: "A modification of the method of multipliers for mathematical programming problems," in "Techniques of optimization," A. V. Balakrishnan (ed.), Academic Press, New York, 1972, 247-260.

16. J. M. Ortega: "Numerical analysis a second course," Academic Press, New York, 1972.

17. J. M. Ortega & W. C. Rheinboldt: "Iterative solution of nonlinear equations in several variables," Academic Press, New York, 1970.

18. A. Ostrowski: "Solution of equations and systems of equations," Academic Press, New York, 1960, second edition, 1966.

19. M. J. D. Powell: "A method for nonlinear constraints in minimization problems," in "Optimization," R. Fletcher (ed.), Academic Press, New York, 1969, 283-298.

20. R. T. Rockafellar: "Convex analysis," Princeton University Press, Princeton, New Jersey, 1970.

21. R. T. Rockafellar: "New applications of duality in convex programming," written version of talk presented at the Seventh International Symposium on Mathematical Programming, The Hague 1970, published in the Proceedings of the Fourth Conference on Probability, Brasov, Romania, 1971.

22. R. T. Rockafellar: "A dual approach to solving nonlinear programming problems by unconstrained optimization," Mathematics Department, Washington University, Seattle, December 1972.

23. R. T. Rockafellar: "the multiplier method of Hestenes and Powell applied to convex programming," Mathematics Department, Washington University, Seattle, December 1972.

24. R. T. Rockafellar: "Augmented Lagrange multiplier functions and duality in nonconvex programming," Mathematics Department, Washington University, Seattle, January 1973.

25. I. J. Schoenberg (ed): "Approximations with special emphasis on spline functions," Academic Press, New York, 1969.

26. J. Stoer: "Duality in nonlinear programming and the minimax theorem," Numerische Mathematik, 5, 1963, 371-379.

27. J. Stoer & C. Witzgall: "Convexity and optimization in finite dimensions I," Springer-Verlag, Heidelberg, 1970.

28. A. P. Wierzbicki: "A penalty function shifting method in constrained static optimization and its convergence properties," Archiwum Automatyki i Telemechaniki, 16, 1971, 395-416.

BIBLIOGRAPHIC DATA SHEET	1. Report No. WIS-CS-74-201	2.	3. Recipient's Accession No.
4. Title and Subtitle Unconstrained Lagrangians in Nonlinear Programming			5. Report Date December 1973 6.
7. Author(s) O. L. Mangasarian			8. Performing Organization Rept. No. TR#201
9. Performing Organization Name and Address The University of Wisconsin, Computer Sciences Department, 1210 W. Dayton Street, Madison, Wisconsin 53706			10. Project/Task/Work Unit No. 11. Contract/Grant No. GJ-35292
12. Sponsoring Organization Name and Address National Science Foundation Washington, D. C.			13. Type of Report & Period Covered 14.

15. Supplementary Notes

16. Abstracts

The main purpose of this work is to associate a wide class of Lagrangian functions with a nonconvex, inequality and equality constrained optimization problem in such a way that unconstrained stationary points and local saddlepoints of each Lagrangian are related to Kuhn-Tucker points or local or global solutions of the optimization problem. As a consequence of this we are able to obtain duality results and two computational algorithms for solving the optimization problem. One algorithm is a Newton algorithm which has a local superlinear or quadratic rate of convergence. The other method is a locally linearly convergent method for finding stationary points of the Lagrangian and is an extension of the method of multipliers of Hestenes and Powell to inequalities.

17. Key Words and Document Analysis. 17a. Descriptors

Optimization
Nonlinear Programming
Mathematical Programming
Lagrangian Methods

17b. Identifiers/Open-Ended Terms

17c. COSATI Field/Group

18. Availability Statement	19. Security Class (This Report) UNCLASSIFIED	21. No. of Pages 42
Available to public.	20. Security Class (This Page UNCLASSIFIED	22. Price