SIAM J. CONTROL AND OPTIMIZATION
Vol. 17, No. 5, September 1979

0363-0129/79/1705-0005$01.00/0

A NEW CLASS OF AUGMENTED LAGRANGIANS IN NONLINEAR PROGRAMMING*

G. DI PILLO† AND L. GRIPPO‡

Abstract. In this paper a new class of augmented Lagrangians is introduced, for solving equality constrained problems via unconstrained minimization techniques. It is proved that a solution of the constrained problem and the corresponding values of the Lagrange multipliers can be found by performing a single unconstrained minimization of the augmented Lagrangian. In particular, in the linear quadratic case, the solution is obtained by minimizing a quadratic function. Numerical examples are reported.

1. Introduction.

During recent years, a number of research works in the area of nonlinear programming have been devoted to the study of methods for solving constrained problems of the form:

$$\text{minimize } f(x) \quad \text{subject to } g(x)=0,$$

via unconstrained minimization techniques.

The most recent results are concerned with the "method of multipliers", which was independently introduced in 1968 by Hestenes [1] and Powell [2]. We refer, e.g. to [3]–[6] for an introduction to this method and for an exposition of related refinements and extensions.

As it is known, the method of multipliers provides the solution of the constrained problem via the solution of a sequence of unconstrained problems of the form:

$$\min_x f(x)+[\lambda, g(x)]+c\|g(x)\|^2,$$

where $c>0$ is a penalty coefficient and λ is an approximation of the Lagrange multiplier.

The relevant feature of the method is that, under suitable assumptions, the solution of the constrained problem is obtained by recursively updating λ, without the need to increase c to infinity. Thus the ill-conditioning associated with the usual penalty methods can be avoided.

The main drawback of the multiplier method is that, in principle, it requires an infinite sequence of unconstrained minimization problems to be solved.

To overcome this difficulty, a further development of the method was proposed by Fletcher [7], [8], who introduced in the augmented Lagrangian a multiplier vector continuously dependent on x. In this way a single minimization is required, as opposed to a sequence of minimizations required in the multiplier method. A related algorithm was proposed and analyzed by Mukai and Polak [9]. These methods, however, require a matrix inversion at each function evaluation and this may limit somewhat their applicability.

A different possibility was considered by Wierzbicki [10], who devised several algorithms for locating directly the saddle-point of the augmented Lagrangian by simultaneous updating of x and λ and without resorting to matrix inversions.

* Received by the editors April 5, 1978 and in revised form January 17, 1979. This work was supported in part by the Consiglio Nazionale delle Ricerche.

† Istituto di Automatica—Università di Roma, Via Eudossiana, 18-00184 Roma, Italy.

‡ Centro di Studio dei Sistemi di Controllo e Calcolo Automatici del CNR, Via Eudossiana, 18-00184 Roma, Italy.

In this paper, we propose a different approach based on the consideration of a new class of augmented Lagrangians obtained by adding to the augmented Lagrangian of Hestenes a penalty term on the first order necessary condition $\nabla_x f+[\partial g/\partial x]^T\lambda=0$.

This leads to a function of the form:

$$S(x,\lambda;c)=f(x)+[\lambda,g(x)]+c\|g(x)\|^2+\left\|M(x)\left(\nabla_x f(x)+\frac{\partial g(x)^T}{\partial x}\lambda\right)\right\|^2,$$

where $M(x)$ is an appropriate weighting matrix.

It is shown that, under suitable hypotheses, a local solution of the constrained problem and the corresponding values of the Lagrange multipliers can be found by performing a single local unconstrained minimization of $S(x,\lambda;c)$ with respect to both x and λ, for finite values of the penalty coefficient and without requiring matrix inversions. In particular, in the linear quadratic case there exists a value of c for which $S(x,\lambda;c)$ is a positive definite quadratic function of (x,λ), so that the solution of a quadratic programming problem with equality constraints can be obtained in a finite number of iterations of a conjugate direction algorithm.

The proposed method has been tested on several problems, with quite satisfactory results. Although the primary emphasis of this paper is not on numerical aspects, we report here a set of numerical examples showing that the method appears to be promising.

2. Problem formulation. We consider the following minimization problem:

Problem P.

$$\text{(1)}\qquad \text{minimize } f(x), \qquad x\in R^n \quad \text{subject to } g(x)=0$$

where $f\colon R^n\to R^1$ and $g\colon R^n\to R^m$, with $m\leqq n$. We assume, unless otherwise stated, that the functions f and g are three times continuously differentiable on R^n.

The Lagrangian $L\colon R^n\times R^m\to R^1$ for problem P is defined by

$$L(x,\lambda)=f(x)+[\lambda,g(x)],$$

where $[\cdot,\cdot]$ denotes the Euclidean scalar product.

We introduce the following augmented Lagrangian function

$$\text{(2)}\qquad S(x,\lambda;c)=f(x)+[\lambda,g(x)]+c\|g(x)\|^2+\left\|M(x)\left(\nabla f(x)+\frac{\partial g(x)^T}{\partial x}\lambda\right)\right\|^2$$

where $c>0$, $M(x)$ is a $(p\times n)$ matrix with twice continuously differentiable elements and $m\leqq p\leqq n$.

To simplify notation, we shall denote by $\nabla_x L(x,\lambda)$ the gradient and by $\nabla_x^2 L(x,\lambda)$ the Hessian of $L(x,\lambda)$ with respect to x, i.e.:

$$\nabla_x L(x,\lambda)\triangleq\nabla f(x)+\frac{\partial g(x)^T}{\partial x}\lambda,$$

$$\nabla_x^2 L(x,\lambda)\triangleq\nabla^2 f(x)+\sum_{i=1}^m \lambda_i\nabla^2 g_i(x).$$

3. Preliminary results. In the sequel we shall make use of the following properties, which establish the relationships between stationary points of $L(x,\lambda)$ and stationary points of $S(x,\lambda;c)$, under the assumption that f, g are two times continuously differentiable and that $M(x)$ is a continuously differentiable matrix.

PROPOSITION 1. *Let $(\bar{x}, \bar{\lambda})$ be a stationary point for $L(x, \lambda)$, then $(\bar{x}, \bar{\lambda})$ is a stationary point for $S(x, \lambda; c)$.*

Proof. Employing a dyadic expansion for $M(x)$, that is

$$M(x) = \sum_{j=1}^{p} e_j m_j(x),$$

where e_j is the jth column of the $(p \times p)$ identity matrix and $m_j(x)$ is the jth row of $M(x)$, we obtain the following expressions for the components of the gradient of $S(x, \lambda; c)$ on $R^n \times R^m$:

$$\begin{aligned}\nabla_x S(x, \lambda; c) = \nabla_x L(x, \lambda) + 2c \frac{\partial g(x)^T}{\partial x} g(x) + 2\nabla_x^2 L(x, \lambda) M^T(x) M(x) \nabla_x L(x, \lambda)\\ + 2\left[\sum_{j=1}^{p} \left(\frac{\partial m_j^T(x)}{\partial x}\right)^T \nabla_x L(x, \lambda)\, e_j^T\right] M(x) \nabla_x L(x, \lambda),\end{aligned} \tag{3}$$

$$\nabla_\lambda S(x, \lambda; c) = g(x) + 2 \frac{\partial g(x)}{\partial x} M^T(x) M(x) \nabla_x L(x, \lambda). \tag{4}$$

Therefore, $\nabla L(\bar{x}, \bar{\lambda}) = 0$ and $g(\bar{x}) = 0$ imply $\nabla_x S(\bar{x}, \bar{\lambda}; c) = 0$, $\nabla_\lambda S(\bar{x}, \bar{\lambda}; c) = 0$. □

PROPOSITION 2. *Let $(\bar{x}, \bar{\lambda})$ be a stationary point for $S(x, \lambda; c)$ and assume that $g(\bar{x}) = 0$ and that $M(\bar{x})\, [\partial g(\bar{x})/\partial x]^T$ is an $(m \times m)$ nonsingular matrix. Then $(\bar{x}, \bar{\lambda})$ is a stationary point for $L(x, \lambda)$.*

Proof. Under the hypotheses stated, $\nabla_\lambda S(\bar{x}, \bar{\lambda}; c) = 0$ and $g(\bar{x}) = 0$ imply $M(\bar{x}) \nabla_x L(\bar{x}, \bar{\lambda}) = 0$ so that from $\nabla_x S(\bar{x}, \bar{\lambda}; c) = 0$ we get $\nabla_x L(\bar{x}, \bar{\lambda}) = 0$. □

PROPOSITION 3. *Let $X \times L$ be a compact subset of $R^n \times R^m$ and assume that $M(x)\, [\partial g(x)/\partial x]^T$ is an $(m \times m)$ nonsingular matrix for any $x \in X$. Then, there exists a $\bar{c} > 0$ such that for all $c \geqq \bar{c}$, if $(\bar{x}, \bar{\lambda}) \in X \times L$ is a stationary point of $S(x, \lambda; c)$, $(\bar{x}, \bar{\lambda})$ is also a stationary point of $L(x, \lambda)$.*

Proof. Let $(\bar{x}, \bar{\lambda}) \in X \times L$ be a stationary point of $S(x, \lambda; c)$. Then, by (4), $\nabla_\lambda S(\bar{x}, \bar{\lambda}; c) = 0$ implies

$$M(\bar{x}) \nabla_x L(\bar{x}, \bar{\lambda}) = -\frac{1}{2} \left[\frac{\partial g(\bar{x})}{\partial x} M^T(\bar{x})\right]^{-1} g(\bar{x}).$$

Therefore, making use of (3) and recalling that $\nabla_x S(\bar{x}, \bar{\lambda}; c) = 0$, we have:

$$\begin{aligned}0 = M(\bar{x}) \nabla_x S(\bar{x}, \bar{\lambda}; c) = \Bigg[-\frac{1}{2} \left[\frac{\partial g(\bar{x})}{\partial x} M^T(\bar{x})\right]^{-1} + 2c M(\bar{x}) \frac{\partial g(\bar{x})^T}{\partial x}\\ - M(\bar{x}) \left(\nabla_x^2 L(\bar{x}, \bar{\lambda}) M^T(\bar{x}) + \sum_{j=1}^{m} \left[\frac{\partial m_j^T(\bar{x})}{\partial x}\right]^T \nabla_x L(\bar{x}, \bar{\lambda})\, e_j^T \right) \left[\frac{\partial g(\bar{x})}{\partial x} M^T(\bar{x})\right]^{-1} \Bigg] g(\bar{x}).\end{aligned}$$

Hence, by the continuity assumptions and the compactness of $X \times L$, there exists a $\bar{c} > 0$ such that for all $c \geqq \bar{c}$ and any $(\bar{x}, \bar{\lambda}) \in X \times L$ the matrix multiplying $g(\bar{x})$ is nonsingular, so that for $c \geqq \bar{c}$, $g(\bar{x}) = 0$. Then, the proof can be completed as in Proposition 2. □

4. Local optimality results.

In order to establish a relationship between local minimum points of (1) and local unconstrained minimum points of S in $R^n \times R^m$ we need a known result on pairs of quadratic forms.

LEMMA 1. *Suppose that P and Q are quadratic forms with the property that $P(y) \leqq 0$ and $Q(y) \leqq 0$ only if $y = 0$. If one of them is nonnegative, then there is a number $c > 0$ such* *that*

$$P(y)+cQ(y)>0,$$

for all $y \neq 0$.

Proof. See [6, p. 113]. □

Then we can prove the following:

THEOREM 1. *Let* $(\bar{x}, \bar{\lambda})$ *be a stationary point for* $L(x, \lambda)$ *and assume that*

(i) $M(\bar{x})\,[\partial g(\bar{x})/\partial x]^T$ *has full rank*

(ii) $\bar{x}$ *is a local minimum point for Problem* P, *satisfying the second order sufficiency condition*:

$$[x, \nabla_x^2 L(\bar{x}, \bar{\lambda})x]>0 \qquad \forall x: x \neq 0,\ \frac{\partial g(\bar{x})}{\partial x}x=0.$$

Then there exists a $c^*>0$ *such that for any* $c \geqq c^*$, $(\bar{x}, \bar{\lambda})$ *is an isolated local minimum point for* $S(x, \lambda; c)$.

Proof. By Proposition 1, $(\bar{x}, \bar{\lambda})$ is a stationary point for $S(x, \lambda; c)$. Consider the Hessian matrix of $S(x, \lambda; c)$ evaluated at $(\bar{x}, \bar{\lambda})$. Since $\nabla_x L(\bar{x}, \bar{\lambda})=0$ and $g(\bar{x})=0$ we have:

$$\nabla_x^2 S(\bar{x}, \bar{\lambda}; c)=\nabla_x^2 L(\bar{x}, \bar{\lambda})+2c\frac{\partial g(\bar{x})^T}{\partial x}\frac{\partial g(\bar{x})}{\partial x}+2\nabla_x^2 L(\bar{x}, \bar{\lambda})M^T(\bar{x})M(\bar{x})\nabla_x^2 L(\bar{x}, \bar{\lambda}),$$

$$\nabla_\lambda^2 S(\bar{x}, \bar{\lambda}; c)=2\frac{\partial g(\bar{x})}{\partial x}M^T(\bar{x})M(\bar{x})\frac{\partial g(\bar{x})^T}{\partial x},$$

$$\nabla_{x\lambda}^2 S(\bar{x}, \bar{\lambda}; c)=\frac{\partial g(\bar{x})^T}{\partial x}+2\nabla_x^2 L(\bar{x}, \bar{\lambda})M^T(\bar{x})M(\bar{x})\frac{\partial g(\bar{x})^T}{\partial x}.$$

Introduce now the quadratic forms in (x, λ):

$$P(x, \lambda)=[x, \nabla_x^2 L(\bar{x}, \bar{\lambda})x]+2\left\|M(\bar{x})\nabla_x^2 L(\bar{x}, \bar{\lambda})x+M(\bar{x})\frac{\partial g(\bar{x})^T}{\partial x}\lambda\right\|^2+2\left[x, \frac{\partial g(\bar{x})^T}{\partial x}\lambda\right],$$

$$Q(x, \lambda)=2\left\|\frac{\partial g(\bar{x})}{\partial x}x\right\|^2.$$

It can be easily verified that

$$[x^T \quad \lambda^T]\begin{bmatrix}\nabla_x^2 S(\bar{x}, \bar{\lambda}; c) & \nabla_{x\lambda}^2 S(\bar{x}, \bar{\lambda}; c)\\ \nabla_{\lambda x}^2 S(\bar{x}, \bar{\lambda}; c) & \nabla_\lambda^2 S(\bar{x}, \bar{\lambda}; c)\end{bmatrix}\begin{bmatrix}x\\ \lambda\end{bmatrix}=P(x, \lambda)+cQ(x, \lambda),$$

so that $(\bar{x}, \bar{\lambda})$ is an isolated local minimum of $S(x, \lambda; c)$ if $P(x, \lambda)+cQ(x, \lambda)$ is positive definite.

We observe now that the assumptions of Lemma 1 are satisfied. In fact $Q(x, \lambda)\geqq 0$; moreover $Q(x, \lambda)=0$ implies $[\partial g(\bar{x})/\partial x]x=0$ so that, taking into account assumption (ii), $P(x, \lambda)\leqq 0$ implies $x=0$ and $M(\bar{x})\,[\partial g(\bar{x})/\partial x]^T\lambda=0$. Finally, by (i) the last equality gives $\lambda=0$.

Then, by Lemma 1, there exists a value $c^*>0$ such that $P(x, \lambda)+c^*Q(x, \lambda)>0$ $\forall (x, \lambda)\neq 0$ and being $Q(x, \lambda)\geqq 0$ the same is true for any $c \geqq c^*$. □

A converse result is given in the following theorem.

THEOREM 2. *Let* $(\bar{x}, \bar{\lambda})$ *be a local minimum point for* $S(x, \lambda; c)$ *and assume that*:

(i) $g(\bar{x})=0$,

(ii) $p=m$, $M(\bar{x})[\partial g(\bar{x})/\partial x]^T$ *is nonsingular*,

(iii) *the Hessian matrix* $\nabla^2 S(\bar{x}, \bar{\lambda}; c)$ *is positive definite.*

Then $\bar{x}$ is an isolated local minimum point for Problem P.

Proof. By Proposition 2, $(\bar{x}, \bar{\lambda})$ is a stationary point for $L(x, \lambda)$; therefore, since $\nabla^2 S(\bar{x}, \bar{\lambda}; c)$ is positive definite, we have:

$$\begin{aligned}&[x, \nabla_x^2 L(\bar{x}, \bar{\lambda})x] + 2\left[x, \frac{\partial g(\bar{x})^T}{\partial x}\lambda\right] + 2c\left\|\frac{\partial g(\bar{x})}{\partial x}x\right\|^2 \\ &+2\left\|M(\bar{x})\nabla_x^2 L(\bar{x}, \bar{\lambda})x + M(\bar{x})\frac{\partial g(\bar{x})^T}{\partial x}\lambda\right\|^2 > 0 \qquad \forall (x, \lambda) \neq 0.\end{aligned} \tag{5}$$

Now let $x \neq 0$ be such that $[\partial g(\bar{x})/\partial x]x = 0$ and take:

$$\lambda = -\left[M(\bar{x})\frac{\partial g(\bar{x})^T}{\partial x}\right]^{-1} M(\bar{x})\nabla_x^2 L(\bar{x}, \bar{\lambda})x.$$

Then we obtain from (5) that $(\bar{x}, \bar{\lambda})$ satisfies the second order sufficiency condition:

$$[x, \nabla_x^2 L(\bar{x}, \bar{\lambda})x] > 0 \qquad \forall x : x \neq 0, \frac{\partial g(\bar{x})}{\partial x}x = 0$$

so that $\bar{x}$ is an isolated local minimum point for Problem P. □

We note that a local result can also be stated under assumptions weaker than those employed in Theorem 2.

THEOREM 3. *Let f, g be two times continuously differentiable and let $(\bar{x}, \bar{\lambda})$ be a local minimum point for $S(x, \lambda; c)$. Assume that*

(i) $g(\bar{x}) = 0$;

(ii) *$M(x)$ is a continuously differentiable $(m \times n)$ matrix such that $M(\bar{x})[\partial g(\bar{x})/\partial x]^T$ is nonsingular.*

Then $\bar{x}$ is a local minimum point for problem P.

Proof. By Proposition 2, $(\bar{x}, \bar{\lambda})$ is a stationary point for $L(x, \lambda)$. This implies:

$$S(\bar{x}, \bar{\lambda}; c) = f(\bar{x}).$$

Therefore, since $(\bar{x}, \bar{\lambda})$ is a local minimum point for $S(x, \lambda; c)$, there exist neighborhoods Ω, Λ of $\bar{x}$, $\bar{\lambda}$, such that

$$f(\bar{x}) \leqq S(x, \lambda; c) \qquad \forall\, x \in \Omega, \quad \lambda \in \Lambda.$$

This yields

$$\begin{aligned}&f(\bar{x}) \leqq f(x) + \left\|M(x)\nabla f(x) + M(x)\frac{\partial g(x)^T}{\partial x}\lambda\right\|^2 \\ &\qquad \forall\, x \in \Omega \cap \{x : g(x) = 0\} \\ &\qquad \forall\, \lambda \in \Lambda.\end{aligned} \tag{6}$$

On the other hand, by the continuity assumptions, there exists a neighborhood Ω' of $\bar{x}$, $\Omega' \subseteq \Omega$, such that

$$\lambda = -\left[M(x)\frac{\partial g(x)^T}{\partial x}\right]^{-1} M(x)\nabla f(x) \in \Lambda, \qquad \forall\, x \in \Omega'. \tag{7}$$

By combining (6), (7) it can be concluded

$$f(\bar{x}) \leqq f(x) \qquad \forall\, x \in \Omega' \cap \{x : g(x) = 0\}. \qquad \square$$

Under the assumptions stated in Proposition 3 it is also possible to ensure that a local minimum point for $S(x, \lambda; c)$ is an admissible point for Problem P.

Therefore, we obtain the following theorem:

THEOREM 4. *Let f, g be two times continuously differentiable and let $X \times L$ be a compact subset of $R^n \times R^m$. Assume that $M(x)$ is a continuously differentiable $(m \times n)$ matrix such that $M(x)[\partial g(x)/\partial x]^T$ is nonsingular for any $x \in X$. Then, there exists a $\bar{c} > 0$ such that for all $c \geqq \bar{c}$, if $(\bar{x}, \bar{\lambda}) \in X \times L$ is a local minimum point of $S(x, \lambda; c)$, $\bar{x}$ is a local minimum point for problem* P.

Proof. The proof follows from Proposition 3 and Theorem 3. □

5. Choice of $M(x)$. In the preceding theorems an important role is played by the matrix $M(x)$. We indicate here some possible choices for $M(x)$ which ensure, under suitable assumptions on the originary problem, that the hypotheses made on $M(\bar{x})[\partial g(\bar{x})/\partial x]^T$ are satsified.

In particular, assume that $[\partial g(\bar{x})/\partial x]^T$ has full rank; then

(a) the choice:

$$M(x) = \mu \frac{\partial g(x)}{\partial x}, \qquad \mu > 0,$$

satisfies the hypotheses of Theorems 1, 2, 3, 4;

(b) the same is true for any choice of $M(x)$ such that $M(x)[\partial g(x)/\partial x]^T$ is an $m \times m$ invertible submatrix of $[\partial g(x)/\partial x]^T$;

(c) the choice:

$$M(x) = \mu I, \qquad \mu > 0,$$

satisfies the hypotheses of Theorem 1.

An important special case of (b) is when the vector x can be split into two subvectors, x_1, x_2, $x_1 \in R^m$, $x_2 \in R^{n-m}$, such that $\partial g(x)/\partial x_1$ is nonsingular for any x. This happens when there exists a set of independent or "decision" variables x_2 and a set of dependent or "state" variables x_1. In such a case a convenient choice for $M(x)$ could be:

$$M(x) = \mu[I_m \mid 0].$$

6. Global optimality results. The results given in § 4 are local in character. A global result is obtained when Problem P has a unique global solution $\bar{x}$ on a compact set X and $\bar{x}$ is an interior point of X:

THEOREM 5. *Let $(\bar{x}, \bar{\lambda})$ be a stationary point for $L(x, \lambda)$ and assume that:*

(i) *assumptions* (i) *and* (ii) *of Theorem* 1 *are satisfied*

(ii) *$\bar{x}$ is the unique global minimum point of Problem* P *on a compact set $X \subseteq R^n$ and $\bar{x} \in \operatorname{int}(X)$.*

Then, for every compact set $L \subseteq R^m$ such that $\bar{\lambda} \in \operatorname{int}(L)$, there exists a $c^*(L) > 0$ such that for any $c \geqq c^*(L)$, $(\bar{x}, \bar{\lambda})$ is the unique global minimum point of $S(x, \lambda; c)$ on $X \times L$.

Proof. Let $L \subseteq R^m$ be a compact set such that $\bar{\lambda} \in \operatorname{int}(L)$. By Theorem 1, there exists a $c_1^* > 0$ such that for $c \geqq c_1^*$, $(\bar{x}, \bar{\lambda})$ affords a local isolated minimum to $S(x, \lambda; c)$. Therefore, since $\bar{x} \in \operatorname{int}(X)$ and $\bar{\lambda} \in \operatorname{int}(L)$ there exist, for $c \geqq c_1^*$, spherical neighborhoods $\Omega(\bar{x}, \varepsilon_c)$, $\Lambda(\bar{\lambda}, \varepsilon_c)$ of $\bar{x}$ and $\bar{\lambda}$ such that $\Omega(\bar{x}, \varepsilon_c) \subseteq X$, $\Lambda(\bar{\lambda}, \varepsilon_c) \subseteq L$ and

$$S(x, \lambda; c) > S(\bar{x}, \bar{\lambda}; c) \qquad \forall (x, \lambda) \in \Omega(\bar{x}, \varepsilon_c) \times \Lambda(\bar{\lambda}, \varepsilon_c), \quad (x, \lambda) \neq (\bar{x}, \bar{\lambda}).$$

Assume now that the conclusion of the theorem is false. Then, for any integer $k \geqq c_1^*$ there exists $(x_k, \lambda_k) \in X \times L$ such that:

$$S(x_k, \lambda_k; k) \leqq S(\bar{x}, \bar{\lambda}; k) = f(\bar{x}).$$

Moreover, it can be easily verified that for $k \geqq c_1^*$,

$$\Omega(\bar{x}, \varepsilon_k) \times \Lambda(\bar{\lambda}, \varepsilon_k) \supseteq \Omega(\bar{x}, \varepsilon_{c_1^*}) \times \Lambda(\bar{\lambda}, \varepsilon_{c_1^*})$$

so that either $\|x_k - \bar{x}\| \geqq \varepsilon_{c_1^*}$ or $\|x_k - \bar{x}\| < \varepsilon_{c_1^*}$ and $\|\lambda_k - \bar{\lambda}\| \geqq \varepsilon_{c_1^*}$. Now, since $X \times L$ is compact, the sequence $\{(x_k, \lambda_k)\}$ admits a convergent subsequence $\{(x_{k_j}, \lambda_{k_j})\}$ for $x_{k_j} \to \hat{x} \in X$, $\lambda_{k_j} \to \hat{\lambda} \in L$ and

$$S(x_{k_j}, \lambda_{k_j}; k_j) \leqq f(\hat{x}).$$

It follows:

$$\limsup_{j\to\infty} S(x_{k_j}, \lambda_{k_j}; k_j) \leqq f(\bar{x}),$$

from which we obtain

$$f(\hat{x}) + [\hat{\lambda}, g(\hat{x})] + \limsup_{j\to\infty} k_j \|g(x_{k_j})\|^2 + \left\| M(\hat{x}) \nabla f(\hat{x}) + M(\hat{x}) \frac{\partial g(\hat{x})^T}{\partial x} \hat{\lambda} \right\|^2 \leqq f(\bar{x}).$$

This implies

$$g(\hat{x}) = 0, \qquad f(\hat{x}) \leqq f(\bar{x}).$$

Then, by assumption (ii), we have $\hat{x} = \bar{x}$ and, by (i) of Theorem 1, $\hat{\lambda} = \bar{\lambda}$. Therefore we get a contradiction either with $\|\hat{x} - \bar{x}\| \geqq \varepsilon_{c_1^*}$ or with $\|\hat{\lambda} - \bar{\lambda}\| \geqq \varepsilon_{c_1^*}$. It can be concluded that there exists a value $c^*(L)$ such that for $c \geqq c^*(L)$, $(\bar{x}, \bar{\lambda})$ is the unique global minimum point of $S(x, \lambda; c)$ on $X \times L$. □

A converse result can easily be stated if it is assumed that any global minimum point of Problem P on X is a stationary point of the Lagrangian $L(x, \lambda)$:

THEOREM 6. *Let f, g be differentiable, let $X \times L$ be a given subset of $R^n \times R^m$ and let $(\bar{x}, \bar{\lambda})$ be a global minimum point for $S(x, \lambda; c)$ on $X \times L$. Assume that*

(i) $g(\bar{x}) = 0$

(ii) *for any global minimum point $\hat{x}$ of Problem* P *on X there exists a multiplier $\hat{\lambda} \in L$ such that $\nabla_x L(\hat{x}, \hat{\lambda}) = 0$.*

Then $\bar{x}$ is a global minimum point of Problem P *on X.*

Proof. By (i) we obtain:

$$S(\bar{x}, \bar{\lambda}; c) = f(\bar{x}) + \left\| M(\bar{x}) \nabla f(\bar{x}) + M(\bar{x}) \frac{\partial g(\bar{x})^T}{\partial x} \bar{\lambda} \right\|^2 \leqq S(x, \lambda; c) \quad \forall (x, \lambda) \in X \times L.$$

Therefore we obtain, in particular,

$$f(\bar{x}) \leqq f(x) \qquad \forall (x, \lambda) \in X \times L : g(x) = 0, \quad \nabla f(x) + \frac{\partial g(x)^T}{\partial x} \lambda = 0.$$

This implies by (ii) that $\bar{x}$ is a global minimum point for Problem P on X. □

Making use of the results given in Proposition 3 and in Theorem 6, we can also state the following:

THEOREM 7. *Let f, g be two times continuously differentiable and let $X \times L$ be a compact subset of $R^n \times R^m$. Assume that $M(x)$ is a continuously differentiable $(m \times n)$ matrix such that $M(x)[\partial g(x)/\partial g]^T$ is nonsingular for any $x \in X$; assume further that* (ii) *of Theorem 6 holds. Then, there exists a $c^* > 0$ such that for all $c \geqq c^*$, if $(\bar{x}, \bar{\lambda}) \in \operatorname{int}(X \times L)$ is a global minimum point of $S(x, \lambda; c)$ on $X \times L$, $\bar{x}$ is a global minimum point for Problem* P *on X.* □

An important special case in which a global property holds is that of quadratic problems with linear equality constraints:

Problem QP.

$$\text{minimize } f(x)=[x, Ax]+[a, x]$$

subject to:

$$Bx=b,$$

where:

(i) $[x, Ax]>0\ \forall x: x\neq 0, Bx=0$

(ii) B has full rank.

In this case we can take for $M(x)$ any constant matrix M such that MB^T has rank m.

We have the following:

THEOREM 8. *Under the assumptions stated for Problem* QP, *there exists a* $c^*>0$ *such that for* $c\geqq c^*$ *the function* $S(x, \lambda; c)$ *is a positive definite quadratic function whose global minimum is the unique solution of Problem* QP.

Proof. By Proposition 1, the optimal solution of Problem QP is a stationary point of $S(x, \lambda; c)$ for any c.

On the other hand, noting that the second order homogeneous part of $S(x, \lambda; c)$ is given by $P(x, \lambda)+cQ(x, \lambda)$ where

$$P(x, \lambda)=[x, Ax]+\|M(2Ax+B^T\lambda)\|^2+[x, B^T\lambda],$$

$$Q(x, \lambda)=\|Bx\|^2$$

and making use of Lemma 1 it can be proved, as in Theorem 1, that there exists a $c^*>0$ such that for $c\geqq c^*$ the quadratic function $S(x, \lambda; c)$ is positive definite. □

7. Numerical examples. In order to evaluate the theory, several numerical examples were explored.

We report here the results obtained for the same set of test problems considered in [11].

The unconstrained minimization of $S(x, \lambda; c)$ with respect to x and λ was performed by the Fletcher–Reeves conjugate gradient method assuming as starting point

$$x_i=2, \qquad i=1, \cdots, n,$$

$$\lambda_i=0, \qquad i=1, \cdots, m$$

and taking for the penalty coefficient the value $c=10$ whenever it worked; only in Example 6 it was necessary to increase c, and the value $c=100$ was used.

For each example we indicate the matrix $M(x)$ employed, and the number N of iterations needed to reach a local minimum point (x^*, λ^*) with the given significant figures.

Example 1.

Minimize

$$f(x)=(x_1-x_2)^2+(x_2+x_3-2)^2+(x_4-1)^2+(x_5-1)^2$$

subject to

$$x_1+3x_2=0,$$
$$x_3+x_4-2x_5=0,$$
$$x_2-x_5=0,$$

$M(x)=I$
$x^*=(-0.7674,\ 0.2558,\ 0.6279,\ -0.1162,\ 0.2558)$
$\lambda^*=(2.0465,\ 2.2325,\ -5.9534)$
$N=8$.

Example 2.
Minimize

$$f(x)=(x_1-1)^2+(x_1-x_2)^2+(x_2-x_3)^4$$

subject to:

$$x_1(1+x_2^2)+x_3^4-4-3\sqrt{2}=0,$$

$M(x)=I$
$x^*=(1.1048,\ 1.1966,\ 1.5352)$
$\lambda^*=-0.1072\times 10^{-1}$
$N=32$.

Example 3.
Minimize

$$f(x)=(x_1-1)^2+(x_1-x_2)^2+(x_3-1)^2+(x_4-1)^4+(x_5-1)^6$$

subject to:

$$x_1^2x_4+\sin(x_4-x_5)-2\sqrt{2}=0,$$
$$x_2+x_3^4x_4^2-8-\sqrt{2}=0,$$

$M(x)=I$
$x^*=(1.1661,\ 1.1821,\ 1.3802,\ 1.5060,\ 0.6109)$
$\lambda^*=(-0.8553\times 10^{-1},\ -0.3187\times 10^{-1})$
$N=189$.

Example 4.
Minimize

$$f(x)=(x_1-1)^2+(x_1-x_2)^2+(x_2-x_3)^2+(x_3-x_4)^4+(x_4-x_5)^4$$

subject to:

$$x_1+x_2^2+x_3^3-2-3\sqrt{2}=0,$$
$$x_2-x_3^2+x_4+2-2\sqrt{2}=0,$$
$$x_1x_5-2=0,$$

$M(x)=\partial g(x)/\partial x$
$x^*=(1.1911,\ 1.3626,\ 1.4728,\ 1.6350,\ 1.6790)$
$\lambda^*=(-0.3882\times 10^{-1},\ -0.1674\times 10^{-1},\ -0.2898\times 10^{-3})$
$N=80$.

Example 5.
Minimize

$$f(x)=0.01(x_1-1)^2+(x_2-x_1^2)^2$$

subject to:

$$x_1+x_3^2+1=0,$$

$M(x)=[1\quad 0\quad 0]$
$x^*=(-1.0000,\ 1.0000,\ 0.2294\times 10^{-5})$
$\lambda^*=0.3999\times 10^{-1}$
$N=52.$

Example 6.
Minimize

$$f(x)=-x_1$$

subject to:

$$x_2-x_1^3-x_3^2=0,$$

$$x_1^2-x_2-x_4^2=0,$$

$M(x)=10^3\, I$
$x^*=(1.0000,\ 1.0000,\ 0.0000,\ 0.0000)$
$\lambda^*=(-1.0000,\ 1.0000)$
$N=90.$

Example 7.
Minimize

$$f(x)=\log(1+x_1^2)-x_2$$

subject to:

$$(1+x_1^2)^2+x_2^2-4=0$$

$M(x)=\partial g(x)/\partial x$
$x^*=(0.0000,\ 1.7320)$
$\lambda^*=0.2867$
$N=15.$

8. Concluding remarks. From a theoretical standpoint, the method proposed in this paper combines several advantages of existing techniques for the solution of constrained problems via unconstrained minimization. On the other hand, possible disadvantages are the increase in dimensionality of the minimization problem, the presence of first order derivatives in the augmented Lagrangian and the fact that $S(x, \lambda; c)$ may be unbounded with respect to λ. This latter difficulty, however, can be overcome in many instances by employing suitable transformations. Another point where attention is needed is the threshold value of the penalty coefficient. Actually it happens that in the convex case the threshold value c^* for $S(x, \lambda; c)$ is larger than the threshold value of the penalty coefficient in the method of multipliers, where c can be given, in principle, any positive value.

The implementation of a computing procedure which makes the best use of the results given here will be considered in the future. There seems to be particular interest in the extension to the proposed method of the results on the automatic selection of the penalty parameter already established for algorithms based on other Lagrangians [9], [12]. Moreover, further investigations will be devoted to the extension of the results considered here to inequality constrained problems and to optimal control problems.

Acknowledgments. The authors would like to thank the Control Theory Center of the University of Warwick, G.B., where much of this work was carried out.

REFERENCES

[1] M. R. HESTENES, *Multipliers and gradient methods*, Computing Methods in Optimization Problems, vol. 2, L. A. Zadeh, L. W. Neustadt, A. V. Balakrishnan, eds., Academic Press, New York, 1969, pp. 143–163.

[2] M. J. D. POWELL, *A method for nonlinear constraints in minimization problems*, Optimization, R. Fletcher, ed., Academic Press, New York, 1969, pp. 283–298.

[3] R. T. ROCKAFELLAR, *Penalty methods and augmented Lagrangians in nonlinear programming*, 5th IFIP Conference on Optimization Techniques, Part I, R. Conti, A. Ruberti, eds., Springer–Verlag, Berlin, 1973, pp. 418–425.

[4] D. P. BERTSEKAS, *Multiplier methods: a survey*, Automatica, 12 (1976), pp. 133–145.

[5] D. A. PIERRE AND M. J. LOWE, *Mathematical Programming via Augmented Lagrangians: an Introduction with Computer Programs*, Addison-Wesley, Reading, MA, 1975.

[6] M. R. HESTENES, *Optimization Theory. The Finite Dimensional Case*, John Wiley, New York, 1975.

[7] R. FLETCHER, *An exact penalty function for nonlinear programming with inequalities*, Math. Programming, 5 (1973), pp. 129–150.

[8] ———, *Methods related to Lagrangian functions*, Numerical Methods for Constrained Optimization, P. E. Gill, W. Murray, eds., Academic Press, New York, 1974, pp. 219–239.

[9] H. MUKAI AND E. POLAK, *A quadratically convergent primal-dual algorithm with global convergence properties*. Memo No. ERL-M455, College of Engineering, Univ. of California, Berkeley, June 1974.

[10] A. P. WIERZBICKI, *A primal-dual large scale optimization method based on augmented Lagrange functions and interaction shift prediction*, Ricerche di Automatica, 7 (1976), pp. 34–58.

[11] A. MIELE, P. E. MOSELEY, A. V. LEVY AND G. M. COGGINS, *On the method of multipliers for mathematical programming problems*, J. Optimization Theory Appl., 10 (1972), pp. 1–33.

[12] E. POLAK, *On the stabilization of locally convergent algorithms for optimization and root finding*, Automatica, 12 (1976), pp. 337–342.