J. Indian Inst. Sci., 66, Feb. 1986, pp. 89-94

Short Communication

On problems of augmentation of Lagrangians in penalty functions of nonlinear programming

SHIV PRASAD AND R.N. MUKHERJEE
Department of Applied Mathematics, Institute of Technology, Banaras Hindu University, Varanasi 221 005, India.

Received on July 28, 1984; Revised on May 19, 1985.

Abstract

We modify Pillow and Grippo's work in which a minimization problem with equality constraints is treated *via* an unconstrained minimization technique in which a new class of augmented Lagrangians is introduced and make a suitable use of a method analogous to the penalty function used by Mangasarian *via* the Wolfe dual of the original minimization problem. For simplicity we treat the case with equality constraints. We believe that this method can be generalized for inequality constrained problems as well. The advantage of the new method is that the penalty parameter remains finite as observed by Mangasarian.

Key words: Penalty functions, nonlinear programming, minimization problem.

1. Introduction

In Pillow and Grippo's work[1], a particular class of Lagrangian is augmented for reducing a minimization problem of nonlinear programming as described below to the case of an unconstrained minimization problem

$$\min_x f(x) \quad \text{subject to } g(x) = 0 \tag{1.1}$$

The Wolfe dual[2] corresponding to problem (1.1) is given as follows:

$$\max_{(x,\lambda)} L(x,\lambda) \quad \text{subject to } \nabla_x L(x,\lambda) = 0,\ \lambda \geq 0, \tag{1.2}$$

where

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

We modify a penalty function given by Mangasarian[2] in our case to give the following unconstrained maximization problem.

$$S(x,\lambda,\gamma) = f(x) + [\lambda, g(x)] - \frac{\gamma}{2} \| M(x) (\nabla_x f(x) + \frac{\partial g(x)^T}{\partial x} \lambda) \|^2 \tag{1.3}$$

$$\max_{(x,\lambda)} S(x,\lambda,\gamma) \tag{1.4}$$

We claim that for a constrained problem with inequality constraints the penalty functions introduced as in (1.3) *via* the Wolfe dual problem (1.2) is more appropriate than the one taken in Pillow and Grippo's work[1]. We obtain results on the lines of Mangasarian[2]. One can extend these results using K-K-T conditions of nonlinear programming for minimization with inequality constraints as well. In the next paragraph we explain certain notations used in relations (1.3) and (1.1).

2. Problem formulation

In formulation (1.1), $f: R^n \to R^1$ and $g: R^n \to R^m$, with $m \leq n$. We assume unless otherwise stated that f and g are three times continuously differentiable functions on R^n. $M(x)$ is a $p \times n$ matrix with twice continuously differentiable elements and $m \leq p \leq n$. γ is a positive real parameter.

To simplify the notation we shall denote by $\nabla_x L(x, \lambda)$ the *Gradient* and by $\nabla_x^2 L(x, \lambda)$ the *Hessian* of $L(x, \lambda)$.

3. Preliminary results

To establish the relationship between stationary points of $L(x, \lambda)$ and stationary points of $S(x, \lambda, \gamma)$ we obtain some preliminary results as follows. Underlying assumptions are that f and g are twice 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 $S(x, \lambda, \gamma)$.

Proof: Employing 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 components of gradient of $S(x, \lambda, \gamma)$ on $R^n x \ R^m$:

$$\nabla_x S(x, \lambda, \gamma) = \nabla_x L(x, \lambda) - \gamma \nabla_x^2 L(x, \lambda) M^T(x) M(x) \nabla_x L(x, \lambda)$$

$$- \gamma \left[\sum_{j=1}^{p} \left(\frac{\partial m_j^T(x)}{\partial x} \right)^T \nabla_x L(x, \lambda) e_j^T \right] M(x) \nabla_x L(x, \lambda). \tag{3.1}$$

$$\nabla_\lambda S(x, \lambda, \gamma) = g(x) - \gamma \frac{\partial g(x)}{\partial x} M^T(x) M(x) \nabla_x L(x, \lambda). \tag{3.2}$$

Therefore,

$$\nabla_x L(\bar{x}, \bar{\lambda}) = 0 \text{ and } g(\bar{x}) = 0 \text{ imply that}$$

$$\nabla_x S(\bar{x}, \bar{\lambda}, \gamma) = 0 \text{ and } \nabla_\lambda S(\bar{x}, \bar{\lambda}, \gamma) = 0.$$

Proposition 2: Let $(\bar{x}, \bar{\lambda})$ be a stationary point for $S(x, \lambda, \gamma)$ 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: Since $(\bar{x},\bar{\lambda})$ is a stationary point for $S(x,\lambda,\gamma)$, we have

$$\nabla_x S(\bar{x},\bar{\lambda},\gamma) = 0 \tag{3.3}$$

and

$$\nabla_\lambda S(\bar{x},\bar{\lambda},\gamma) = 0 \tag{3.4}$$

Also, we have

$$g(\bar{x}) = 0 \tag{3.5}$$

Therefore substituting (3.4) and (3.5) in (3.2), we have, because of the nonsingularity of $M(\bar{x})[\partial g(\bar{x})/\partial x]^T$,

$$M(\bar{x})\,\nabla_x L(\bar{x},\bar{\lambda}) = 0. \tag{3.6}$$

Again using (3.3) and (3.6) in relation (3.1), we have

$$\nabla_x L(\bar{x},\bar{\lambda}) = 0. \tag{3.7}$$

Further, we already have

$$\nabla_\lambda L(\bar{x},\bar{\lambda}) = g(\bar{x}) = 0. \tag{3.8}$$

The relations (3.7) and (3.8) together imply that $(\bar{x},\bar{\lambda})$ is a stationary point for $L(x,\lambda)$.

The proof of the next proposition follows an analogous course as that of proposition 3 in Pillow and Grippo[1].

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 is a $\bar{\gamma} > 0$ such that for all $\gamma \geq \bar{\gamma}$, if $(\bar{x},\bar{\lambda}) \in X \times L$ is a stationary point of $S(x,\lambda,\gamma)$, $(\bar{x},\bar{\lambda})$ is also a stationary point of $L(x,\lambda)$.

4. Local optimality results

The results derived next differ from that of Pillow and Grippo[1] are in tune with similar results in Mangasarian[2].

Theorem 1: Let $(\bar{x},\bar{\lambda})$ be a stationary point of $L(x,\lambda)$. Let $\nabla_x^2 L(\bar{x},\bar{\lambda})$ be positive definite with minimum eigenvalue $\bar{\eta} > 0$. Then for γ such that $\gamma D(\bar{x}) \geq 1/\bar{\eta}$, where $D(x) = M^T(x) M(x)$, $(\bar{x},\bar{\lambda})$ is a stationary point of $S(x,\lambda,\gamma)$ and the Hessian $\nabla^2 S(x,\lambda,\gamma)$ with respect to $(\bar{x},\bar{\lambda})$ is negative semi-definite. If in addition $\gamma D[\bar{x}] \geq 1/\bar{\eta}$ and $\nabla g(\bar{x})$ has linearly independent rows, then $\nabla^2 S(\bar{x},\bar{\lambda},\gamma)$ is negative definite and hence $(\bar{x},\bar{\lambda})$ is a strict local maximum of $S(x,\lambda,\gamma)$.

Proof: By proposition 1, $(\bar{x},\bar{\lambda})$ is a stationary point of $S(x,\lambda,\gamma)$. Then we have

$$\nabla S(x,\lambda,\gamma) = \begin{array}{l} \nabla_x L(x,\lambda) - \gamma \nabla_x^2 L(x,\lambda)\, M^T(x) M(x) \nabla_x L(x,\lambda) \\ \gamma\,[\sum_{j=1}^{p} \frac{\partial m_j^T(x)}{\partial x})^T \nabla_x L(x,\lambda)\, e_j^T]\, M(x) \nabla_x L(x,\lambda) \\ g(x) - \gamma \frac{\partial g(x)}{\partial x} M^T(x) M(x) \nabla_x L(x,\lambda) \end{array}$$

where the matrix is a 2×1 matrix.

Recalling that $\nabla_x L(\bar{x},\bar{\lambda})=0$ we have that,

$$\nabla^2 S(\bar{x},\bar{\lambda},\gamma)=\begin{bmatrix} \nabla_x^2 L(\bar{x},\bar{\lambda})\{1-\gamma M^T(\bar{x})\,M(\bar{x})\,\nabla_x^2 L(\bar{x},\bar{\lambda})\} & \{1-\gamma\nabla_x^2 L(\bar{x},\bar{\lambda})\,D(\bar{x})\}\dfrac{\partial g(\bar{x})^T}{\partial x} \\ \dfrac{\partial g(\bar{x})}{\partial x}\{1-\gamma\nabla_x^2 L(\bar{x},\bar{\lambda})\,D(\bar{x})\} & -\gamma\dfrac{\partial g(\bar{x})}{\partial x}D(\bar{x})\dfrac{\partial g(\bar{x})^T}{\partial x}\end{bmatrix}$$

where the above matrix is a 2×2 matrix.

Define

$$C=\nabla_x^2 L(\bar{x},\bar{\lambda}) \text{ and } A=\frac{\partial g(\bar{x})}{\partial x}$$

Then,

$$\nabla^2 S(\bar{x},\bar{\lambda}\gamma)=\begin{bmatrix} C(1-\gamma D(\bar{x})C) & (1-\gamma D(\bar{x})C)A^T \\ A(1-\gamma D(\bar{x})C) & -\gamma D(\bar{x})A\,A^T\end{bmatrix}$$

$$=\begin{bmatrix} C & A^T \\ A & 0\end{bmatrix}-\gamma D(\bar{x})\begin{bmatrix} C \\ A\end{bmatrix}\begin{bmatrix} C & A^T\end{bmatrix}$$

Now,

$$(x^T\lambda^T)\nabla^2 S(\bar{x},\bar{\lambda},\gamma)\binom{x}{\lambda}$$

$$\begin{aligned}&=x^TCx+2x^TA\lambda-\gamma D(\bar{x})\,\|Cx+A^T\lambda\|^2\\ &\leq-\bar{\eta}\,\|x\|^2+2\,\|x\|\,\|Cx+A^T\lambda\|-\gamma D(\bar{x})\,\|Cx+A^T\lambda\|^2\\ &=-\bar{\eta}\,(\|x\|-1/\bar{\eta}\,\|Cx+A^T\lambda\|)^2\\ &\quad-(\gamma D(\bar{x})-1/\bar{\eta})\,\|Cx+A^T\lambda\|^2.\end{aligned}$$

Hence because of the hypothesis, $\nabla^2 S(\bar{x},\bar{\lambda},\gamma)$ is negative semi-definite for $\gamma D(\bar{x})\geq 1/\bar{\eta}$.

Case I: $Cx+A^T\lambda\neq 0$. For this case it follows from

$\gamma D(\bar{x})\geq 1/\bar{\eta}$ that $(x^T\lambda^T)\,\nabla^2 S\,[\bar{x},\bar{\lambda},\gamma)\;\binom{x}{\lambda}<0$.

Case II: $Cx + A^T\lambda = 0$ and $\binom{x}{\lambda} \neq 0$. For this case we have that $x \neq 0$, else $\lambda^T A = 0$, $\lambda \neq 0$, which contradicts the assumption that the rows of A are linearly independent.

Hence,

$$(x^T \lambda^T)\nabla^2 S(\bar{x}, \bar{\lambda}, \gamma) \binom{x}{\lambda} = -x^T Cx < 0,$$

where the last inequality follows from the assumption that C is positive semi-definite.

5. Global optimality results

With no convexity assumption all the results derived so far are local results. One can globalize some of these results by assuming uniform strict convexity of f and convexity of g on R^n. In fact one can show that for each local solution $(x(\gamma), \lambda(\gamma))$ of (1.4) $x(\gamma)$ is the unique global solution of (1.1).

Theorem 2: Let f and g be convex and twice continuously differentiable functions on R^n with f being uniformly strictly convex, $(\bar{x}, \bar{\lambda})$ be a stationary point of (1.4), $g_i(\bar{x}) = 0$ and $M(\bar{x})[\partial g(\bar{x}))/(\partial g)]^T$ is an $m \times m$ nonsingular matrix. Then $(\bar{x}, \bar{y})$ is a stationary point of $L(x, \lambda)$ and x is the unique global solution of (1.1).

Proof of the above result is on similar lines as that of proposition 2 of section 3.

Remark 1: For a possible potential application of our result one can choose a quadratic programming problem of the following type:

$$\min f(x) = [x, Ax] + [a, x]$$

$$\text{subject to } Bx = b,$$

where (i) $[x, Ax] > 0 \ \forall x : x = 0, \ Bx = 0$
(ii) B has full rank.

We can take for $M(x)$ any constant matrix M such that MB^T has rank m. This example is taken from Pillow and Grippo [1].

Remark 2: The suitable choices of $M(x)$, as discussed in reference 1, will be valid here too.

Remark 3: For a comprehensive treatment of penalty method and barrier method in nonlinear programming reader is referred to Zangwill [3].

Acknowledgement

The authors are thankful to the referee for valuable criticism which enabled the authors to improve the earlier version of the manuscript.

References

1. Di Pillow, G. and Grippo, L. A new class of augmented Lagrangians in nonlinear programming, *SIAM J. Control Optimization*, 1979, **17**(5), 618.

2. MANGASARIAN, O.L. A dual differentiable exact penalty function, Computer Science Technical Report # 434, 1-22, University of Wisconsin, Madison, Wisconsin, 1981.

3. ZANGWILL, W.I. *Nonlinear programming,* Prentice-Hall Inc., New Jersey, 1969.