Acta Ciencia Indica, Vol. XLI M, No. 3, 207 (2015)

A GLOBALLY CONVERGENT ALGORITHM FOR CONSTRAINED OPTIMIZATION USING QUADRATIC APPROXIMATION

SAURABH SRIVASTAVA

Department of Mathematics, Hindustan College of Science & Technology, Farah, Mathura (U.P.), India

RECEIVED : 15 July, 2015

REVISED : 31 July, 2015

"This paper presents an algorithm utilizing a quadratic approximation for determining the search direction and an exact penalty function for choosing the step length. The algorithm is globally convergent. The algorithm provides a rule for choosing the penalty parameter near a solution employs a search are rather than a search direction to avoid truncation of the step length".

KEYWORDS : Globally convergence, Exact penalty function, penalty parameter.

Introduction

We consider the following problem

$$\text{Min } \{f(x : h(x) = 0, g(x) \leq 0\} \quad (1)$$

where $f : R^n \to R, g : R^n \to R^m, h : R^n \to R^r$.

There are several algorithm which possess a superliner rate of convergence but are only locally convergent given by Chamberlen [2], Fletecher [5], Giull and Murray [6] and Powell [11], [12], Rios LM, Sahinidis NV, [15]. To stabilize globally these algorithms, the main difficulty is that they generate sequences which are not compulsorily feasible. It makes a confusion to decide whether the next iteration is an improvement or not. In this class of algorithm we find search direction by solving a first or second order approximation to the original problem and the step length is selected by approximately to the original problem and the step length is selected by approximately minimizing an exact penalty function. The concept of application of penalty function in constrained minimization problem was studied and given in detail by Conn [3], Daniel [4] and Robinson [13]. Global convergence for constrained optimization problem was proved by Han [7], [8], Wang, Huang and He [14]. Modified Armijo procedure for penalty function approach was used by Audet, Dennis [1], Levenberg [9] and Marquardt [10].

This paper presents an algorithms utilizing a quadratic approximation for determining the search direction and an exact penalty function for choosing the step length, which overcomes above obtained difficulties.

The algorithm :

Superlinear convergence is obtained by using a search direction which is obtained by solving a quadratic approximation to the original problem. Let L denotes the lagrangian function as :

$$L(x, \lambda/\mu) = f(x) + <\lambda_/, g(x)> + <\mu, h(x)> \quad \dots (2)$$

Let $H \in R^{n \times n}$ is an estimate of the Hessian $L_{xx}(x, \lambda, \mu)$. The quadratic approximation to P, given x and H, is defined to be the program:

$$QP(x,H): \quad \min \{f_x(x)p + (\tfrac{1}{2}p^T HP \mid p \in \hat{F}(x)\} \quad \dots (3)$$

where $$F(x): \{p \in R^n \mid g(x) + g_x(x)p \le 0, h(x) + h_x(x)p = 0\} \quad \dots (4)$$

Solving $QP(x_1 H)$ yields a search direction which is certainly satisfactory near a solution point, but which may not be satisfactory.

The exact penalty function employed in this paper is $r : R^n \times R \to R$ defined by

$$r(x, c) = f(x)(+ c\psi : (x) \quad \dots (5)$$

where $\psi : R^n \to R$ is defined by

$$\psi(x) = \max \{g^j(x), j \in m; |h^i(x)|, j \in r; 0\} \quad \dots (6)$$

where m, r denote, respectively the sets $\{1_1 2_1 \dots, m\}$ and $\{1, 2, \dots, r\}$. The penalty parameter is c.

Obviously $\psi(x) \ge 0$ and $\psi(x) \ge 0$ and $\psi(x) = 0$ iff x is feasible. To make certain whether a given search direction p is a descent direction for $\gamma(x, c)$, we require first order estimates and $\hat{\gamma}(r, p, c)$, $\psi(x, +p)$; these are obtained by replacing $g^j(x+p)$ by $g^j(x) + g^j_x(x)p$ and $h^j(x+p)$ by $h^j(\mathrm{I}) + h^i_x(x)p$ in the appropriate definitions yielding.

$$\hat{\gamma}(x_1, p, c) \; f(x) + f_x(x) \; p + c\psi x, \; p) \quad \dots (7)$$

$$\hat{\psi}(x, p) = \max\left\{g^j(x)p, j \in m, \left|h^j(x) + h^j{}_x(x)p : j \in r; 0\right|\right\} \quad \dots (8)$$

It is supposed that f, g, h are continuously differentiable,

Let $\theta : R^n \times R^n \times R \to R$ is defined by

$$\theta(x, p, c) = \hat{\gamma}(x, p, c) - \gamma(x, c) \quad \dots (9)$$

so that $\theta(x, p, c)$ is a first order estimate of $\gamma(x+p, c) - \gamma(x, c)$; p is a descent direction for $\gamma(x, c)$ if $\theta(x, p, c) < 0$. If p is a solution of $QP(x, H)$, then, under certain condition, c can be chosen so that p is a descent direction γ (x, c).

Proposition 1 : Let $\{p, \lambda, \mu\}$ be a Kuhn-Tucker triple for QP (x, H). Then p is a descent direction for $\gamma(x,c)$ if

(i) H is a positive definite.

(ii) $$c \ge \sum_{j=1}^{m} \lambda^j + \sum_{j=1}^{m} |\mu^j| \quad \dots (10)$$

This result is followed from the fact that if $\{p, \lambda, \mu\}$ is a Kuhn-Tucker triple for $QP(x, {}_H)$, then

$$\theta(x, p, c) \le -p^T hp - \left[c - \sum_{j=1}^{m} \lambda^j + \sum_{j=1}^{m} |\mu^j| \psi(x)\right] \quad \dots (11)$$

Algorithm Model

Data : $x, \in R^n$, $c_0 > 0$, $\delta > 1$

Step 0 : Set $i = 1$

Step 1 : If $C_i - 1 \ge \bar{c}(x_i)$ set $c_i = c_{i-1}$.

if $c_{i-1} < \bar{c}(x_i)$ Set $c_i = \max\{\delta C_{i-1}, \bar{c}(x_i)\}$

Step 2: Compute any $x_{i+1} \in A\ (x_i, c_i)$
Set $i = i + 1$ and go to step 1.

The procedure for choosing the penalty parameter is given in step 1. If x_i is the current value of x and c_i the current value of the penalty parameter, then $A\ (x_i, c_i)$ is the set of all possible successors to x_i that can be generated by the algorithm so that $x_{i+1} \in A\ (x_i, c_i)$. Let D and $d\ D_c$ respectively, denote the set of points satisfying the necessary conditions for P and P_c min $\{r(x, c)\}$. Now sufficient conditions for the convergence of D may be stated as in the form suppose that for all $c > 0$, $A\ (\bullet, c)$ has the of theorem the following property.

Theorem-1 :

(i) if $\{x_i\}$ is any infinite sequence such that $x_{i+1} \in A\ (x_i, c)$, $c \geq \bar{c}\ (x_i)$, for all i, than any accumulation point x^* of $\{x_i\}$ satisfying $x^* \in D_c$. Suppose $\bar{c}\ : R^n \to R$ has the following properties.

(ii) $x \in D_c$ and $c \geq \bar{c}\ (x) \Rightarrow x \in D$.

(iii) $\bar{c}$ is continuous.

Then any sequence $\{x_i\}$ generated by the algorithm model has the following properties;

(a) if c_{i-1} is increased finitely often any accumulation point of $\{x_i\}$ *i.e.* x^* satisfies $x^* \in D$.

(b) if c_{i-1} is increased finitely often, for $i \in K$ say then the infinite sequence $\{x_i\}$ $i \in k$ has no accumulation points.

An immediate consequence of above stated theorem is as follows.

Corollary : If the sequence $\{x_i\}$ generated by the algorithm model is bounded. then c_i is increased only finitely often, and any accumulation x^* of $\{x_i\}$ is desirable ($x^* \in D$).

Penalty Parameter : A possible formula for choosing c is suggested by (10). We replace $(\lambda_1 \mu)$ in (10) by continuous first order estimates $\{\bar{\lambda}(x), \bar{\mu}(x)\}$ where $\bar{\lambda} : R^n \to R^m$ and $\bar{\mu} : R^n \to R^m$ are defined by.

$$(\bar{\lambda}(x), \bar{\mu}(x)) = \arg\min \Big\{ \left\| \nabla f\ (x) + gx\ (x)^T \lambda + h_x\ (x)^T .\mu \right\|^2 + \sum_{j=1}^{m} \left[(\psi\ (x) - g^j\ (x))^2\ (\lambda^j)^2 \right] + \sum_{j=1}^{r} \left[\psi\ (x) - \left| h^j\ (x) \}\right)^2\ (\mu^j)^2 \Big| \right] . \lambda \in R^m \mu \in R^r \Big\} \qquad \ldots (12)$$

The first term in the RHS of (12) ensures that $\bar{\lambda}$ and $\bar{\mu}$ are estimates of the multipliers and the second and third terms ensure their continuity if $\{\hat{x}, \hat{\lambda}, \hat{\mu}\}$ is a Kuhn-Tucker triple for P, then $\bar{\lambda}(\hat{x}) = \hat{\lambda}$ and $\bar{\mu}(\hat{x}) = \hat{\mu}$ Our test function $\bar{c} : R^n \to R$ is now defined by

$$\bar{c}\ (x) = \max \left\{ \sum_{j=1}^{m} \bar{\lambda}_j\ (x) + \sum_{j=1}^{r} \left| \bar{\mu}^j\ (x) \right| + b,\ b \right\} \qquad \ldots (13)$$

where b is an arbitrary small constant.

Search Direction : Let $p(x_1 H)$ denote any solution of $QP\ (x, H)$. The algorithm selects for the search direction if this is consistent with convergence suitable conditions for acceptance are :

(i) A solution $\hat{p}$ (x, H) of $QP\,(x, H)$ exists.

(ii) $\| \hat{p}\ (x, H) \| \leq L.$ … (14)

(iii) $\theta\,(x, \hat{p}\ (x, H)\,, c) \leq -T\,(x)$

where L is a large positive constant and T is a continuous function satisfying $T(x) \geq 0$ and $T(x) = 0$ iff $x \in D$. A suitable function is

$$T\,(x) = \min\left\{\in \left[\psi\,(x) + \left\| \nabla\,(x) + g_i\,(x)^T\,\overline{\lambda}\,(x)^T\,\overline{\mu}\,(x)\right\|^2\right]^2\right\} \qquad \text{… (15)}$$

where $\in$ is small positive constant and $\overline{\lambda}\,(x)^T$ denotes the vector whose i^{th} component is $\overline{\lambda^i}\,(x)^+$.

A convergent algorithm $\hat{p}$ (x, H) as a search direction if conditions (14) are satisfied and $\overline{p}\,(x, c)$ otherwise, can be constructed. The test condition (14) such that the former requirement is satisfied.

Step Length : Since γ is not continuously differentiable the standard Armijo test, which employs the gradient of γ, cannot be employed. This can be modified using our estimate $q\,(x, \alpha p, c)$ of $\gamma\,(x + \alpha p, c) - \gamma\,(x, c)$. It is easily shown that $\theta\,(x, \alpha p, c) \leq \alpha\theta\,(x, p, c)$ for all $\alpha \in [0, 1]$. So that the modified Armijo procedure is, choose the largest $\alpha \in A = \{1, \beta, \beta^2 .....\}$. where $\beta \in (0,1)$ such that

$$\gamma\,(x + px, c) - \gamma\,(x) \leq \overline{\delta}\,\alpha\theta\,(r, p_1 c). \qquad \text{… (16)}$$

for some $\overline{\delta} \in (0,1)$.

We now have all the necessary ingredients to state the algorithm.

Main Algorithm

Data : $x, H, c_0 \geq 0.\ \delta > 1, L \in (0, \infty), \in < 1, \beta \in (0, 1).$

Step 0 : Set $i = 0$.

Step 1 : If $c_{i-1} \geq \overline{c}\ (x_i)$ set $c_i = c_{i-1}$.

If $c_{i-1} < \overline{c}\ (x_i)$, set $c_i = \max\,\{\delta\, c_{i-1}, \overline{c}\ (x_i)\}$

Step 2 : If

(α) A minimum norm solution $\overline{p}\ (x_i, H_i)$ of $QP\,(x; H)$ exists.

(β) $\| \overline{p}\ (x_i, H_i) \| \leq L.$

(γ) $\theta\,(x_i, \overline{p}\ (x_i, H_i), c_i) \leq -T\,(x_i)$

then compute $\overline{p}\ (x_i, H_i)$ and

set $p_i = (x_i, H_i)$, $\overline{p_i} = \overline{p}\ (x_i, H_i)$ Else set $p_i = \overline{p}\ (x_i, c_i)$.

Step 3 : If $p_i = \hat{p}\ (x_i H_i)$. Compute the largest $\alpha_i \in A$ such that

$\gamma\,(x_i + \alpha_i p_i + \alpha_i^2\ \overline{p_i}) - \gamma\ (x_i, c_i) \leq 1/8\ \alpha_i\ \theta\,(x_i, p_i, c_i)$

and set $x_{i+1} = x_i + \alpha_i p_i + \alpha_i^2\ \overline{p_i}$.

If $p_i = \overline{p}\ (x_i, c_i)$ compute the largest $\alpha_i \in A$ such that

$\gamma\,(x_i + \alpha_i p_i, c_i) - \gamma\ (x_i, c_i) \leq \dfrac{\alpha_I}{4} Q\,(x_i, p, c_i)$ and set $x_{i+1} = x_i + \alpha_i\, p_i$.

Step-4 : Update H_i to H_{i+1}.

Step-5 : Set $i = I + 1$ and go to step 1.

Now we shall discuss global convergence of this algorithm.

Global Convergence : For all $x \in R^n$ let $I(x)$ index the most active inequality constraints and $E(x)$ the most active equality constraints *i.e.*

$$I(x) = \{j \in m \mid g^j(x) = \Psi(x)\} \quad \dots (17)$$

and

$$E(x) = \{j \in r \mid h^j(x) = \psi(x)\} \quad \dots (18)$$

Either $I(x)$ or $E(x)$ may be empty indeed for almost all x in R^n there will exist only one integer in $I(x) \cup E(x)$, which is therefore a small subset of the active constraints. Here we make following assumptions.

(H1) The functions $f_i\, g$, h are continuously differentiable

(H2) For all x the vectors $\{\nabla g^j(x),\ j \in I(x)\ ;\ \nabla h^j(x), j \in E(x)\}$ are linearly independent To establish global convergence we prove that $\bar{c}$ and A. satisfy the hypothesis of theorem 1. To proceed, we need to define D, D_c more precisely as

$$D = \{ x \in R^n \mid (x, \lambda, \mu) \text{ is a Kuhn Tucker triple for } p \} \quad \dots (18)$$

$$D_c = \{x \in R^n \mid Q^1 (x_1 C) = Q (x_1\ \bar{p}\ (x, c)\}, c) = 0\} \quad \dots (19)$$

It follows that $(\bar{\lambda}(x), \bar{\mu}(x))$ to (12) always exists and is unique and that $\bar{\lambda}, \bar{\mu}$ are continuos. Also if $(\hat{x}, \hat{\lambda}, \hat{\mu})$ is a Kuhn-Tucker triple for P_i then $\hat{\lambda} = \hat{\lambda}(\hat{x}),\ \hat{\mu} = \hat{\mu}(\hat{x})$.

Hence we have

Proposition-2 $\bar{c}\ : R^n \to R$ is continuous.

Next we know that $\theta^1 (x, c) \leq 0$ and that $\bar{p}\ (x, c)$ is descent direction for $\gamma\ (x, c)$ if $\theta^1 (x, c) \leq \theta(x,\ \bar{p}\ (x, c), c) < 0$ so that $\theta^1 (x, c) = 0$ is a necessary condition for the unconstrained minimization problem *i.e.*

$$\min \{\gamma (x, c) \mid x \in R^n\}$$

Hence D_c is a set of desirable points for P_c using the dual from of $Q\ P_c\ (x_1)$, the following results can be established.

Proposition-3 : Let $c \geq\ \bar{c}\ (x)$. Then $x \in D_c$ iff $x \in D$ it is a typical exact penalty function results, establishing equivalence of P and P_c.

Recalling that $\bar{p}\ (x,c)$ is a solution to $Qp_c\,(x)$ always exists and is unique. It can also be established given $\bar{c} > 0$, that $\bar{p}\ (\bullet, c)$ is continuous. Moreover the continuity of $\bar{\lambda}$ and $\bar{\mu}$ imply that T is continuous also $T(x) > 0$ if $x \in D$. Hence $T : R^n \times R \to R$ defined by $\pi (x, c) = \max \{- T(x)\ ,\ \theta^1 (x, c)\}$ is continuous in x. Since $\theta (x,\ p\ (x, H)\ , c) \leq - T(x) \leq \pi (x, c)$ and $\theta\ \bar{p}\ (x_i, (x_1 c), c) < \theta^1 (x, c) \leq \pi (x, c)$, it follows that

$$\theta (x, p\,(x, H, c), c) \leq \pi (x, c) \quad \dots (20)$$

for all x and all H. Since $\theta (x, \alpha p\ (x, H, c), c)$ is a first order estimate of $\gamma\ (x + \alpha p\ (x, H, c) + \alpha^2\, \bar{p}\ \ (x, c), c) - \chi\ (x, c)$ and since $\pi (\bullet, c)$ is continuos, it follows that for all (x, c) s.t. $c \geq \bar{c}\ (x)$ and $x \in D_c$.

There exists an $\in > 0\ \delta > 0$ s.t. $\gamma (x'', c) - \gamma (x^l, c) \leq - \delta$ $\quad \dots (21)$

for all $x^1 \in \beta (x, \in)$ all $x^n \in A\ (x; H, c)$ and all symmetric H. Hence if x^* be any accumulation point of an infinite sequence $\{x_i\}$ satisfying $x_{i+!} \in A\ (x_i; H, c)$ and $c > \bar{c}\ (x_i)$ for all i; then $x^* \in D_c$. Thus $\bar{c}$ and A. satisfy the hypothesis (i) to (iii) of theorem 1 yielding let $\{x_i\}$ be an infinite sequence generated by the main algorithm. Then $\{x_i\}$ has the convergence properties specified in conclusions (a) and (b) of theorem 1 and the corollary to theorem 1.

Conclusion

Thus we have derived an algorithm which is implacable and is a useful workhouse which can be modified in many ways to improve performance without affecting its asymptotic properties.

References

1. Audet, C., Dennis, Jr J.E., Mesh adaptive diresct search algorithms for constrained optimization, *SIAM Journal on Optimization*, **17(1)**, 188-217 (2006).
2. Chamberlan, R.M., "Some examples of cycling in variable metric methods for constrained minimization", *Mathematical Programming*, **6,** 378-383 (1979).
3. Conn, A.R., "Constrained optimization using a non-differentiable penalty function", *SIAM Jounral on Numerical Analysis,* **10,** 760-779 (1973).
4. Daniel, J.W., "Stability of the solution of definite quadratic programs", *Mathematical Programming*, **5,** 41-53 (1973).
5. Fletehcer, R., "Minimizing general functions subject to linear constraints", In *F.A. Lootsmar ed. Numerical methods for nonlinear optimization*, Academic press, London, pp. 279-296 (1972).
6. Giull, P.E. and Murray, W., "The computation of Lagrange multiplier estimates for optimization. *Mathematical Programming*, **17**, 32-60 (1979).
7. Han, S.P., "Super linearly convergent variable metric algorithms for general nonlinear programming problems", *Mathematical Programming*, **11,** 263-282 (1976).
8. Han, S.P., "A globally convergent method for nonlinear programming", *Journal of Optimization Theory and Applications*, **22,** 297-309 (1977).
9. Levenberg, K., "A method for the solution of certain problems in least squares", *Quarterly of Applied Mathematics*, **2**, 164-168 (1994).
10. Marquardt, D., "An algorithm for least squares estimation of nonlinear parameters", *SIAM Journal on Applied Mathematics*, **11,** 431-441 (1963).
11. Powell, M.J.D., "Variable metric methods for constrained optimization", *Math. Prog., Springer*, 288-311 (1983).
12. Powell, M.J.D., Yuan, T,, "A trust region algorithm for equality constrained optimization", *Math. Prog.*, **49**, 189-211 (1991).
13. Robinson, S.M., "A quadratically convergent algorithm for general nonlinear programming problems", *Mathematical Programming*, **3,** 145-156 (1972).
14. Wang, L., Huang, F. HeQ, "An effective co-evalutionary differential evaluation for constrained optimization", *Applied Mathematics and Computation*, Elsevier, **186**, 340-356 (2007).
15. Rios, L.M., Sahinidis, N.V., " Derivative- free optimization : a review of algorithms and comparison of software implementations", *Journal of Global Optimization*, **2013**, 917-941(2000).

❑