RAIRO. RECHERCHE OPÉRATIONNELLE

FRANK KÖRNER

Remarks on the Newton method for solving nonlinear equality constrained optimization problems

RAIRO. Recherche opérationnelle, tome 24, n° 3 (1990), p. 287-294

<http://www.numdam.org/item?id=RO_1990__24_3_287_0>

Recherche opérationnelle/Operations Research
(vol. 24, n° 3, 1990, p. 287 à 294)

REMARKS ON THE NEWTON METHOD FOR SOLVING NONLINEAR EQUALITY CONSTRAINED OPTIMIZATION PROBLEMS (*)

by Frank KÖRNER [1]

Abstract. — *We discuss the Newton method in connection with the Lagrangian function, the Lagrangian dual problem and decomposition. In case the Hessian of the Lagrangian is regular we show that all second-order informations are contained in this matrix.*

Keywords : Nonlinear programming; Newton's method; linearization techniques; second-order approximation.

Résumé. — *Nous examinons la méthode de Newton en relation avec la fonction lagrangienne, le problème du dual lagrangien et la décomposition. Dans le cas où la matrice hessienne du lagrangien est régulière, nous montrons qu'elle contient toutes les informations du deuxième ordre.*

1. INTRODUCTION

We consider problems of the following form:

$$\text{(P)} \qquad f(x) \to \min \quad \text{subject to } g(x)=0, \qquad x \in R^m,$$

with $g(x) := (g_1(x), \ldots, g_m(x))^T$. The Lagrangian to (P) becomes:

$$L(x, u) := f(x) + u^T g(x).$$

Many algorithms have been developed for determining a local minimizer of (P). We discuss Newton's method in connection with the Lagrangian function, the Lagrange dual problem and decomposition.

If the Hessian of the Lagrangian is regular, then this matrix contains much information on the dual problem and decomposition. The aim of this paper is to discuss some relationships in this sense.

(*) Received February 1989, revised in August 1989.
[1] Mauthestraße 13, VS-Schwenningen, D-7730, R.F.A.

Recherche opérationnelle/Operations Research, 0399-0559/90/03 287 08/$ 2.80

The linear systems occurring in the iteration process are solved with a special pivoting procedure. We obtain, without expensive efforts, the information if the second-order Kuhn-Tucker conditions are fultilled or not.

2. ON THE NEWTON METHODS FOR SOLVING (P)

We say that (x^*, u^*) fulfils the Kuhn-Tucker conditions (KT) if the following statement is true:

$$\nabla L(x^*, u^*)=\begin{bmatrix} \nabla f(x^*)+\nabla g(x^*)^T u^* \\ g(x^*) \end{bmatrix}=0, \tag{1}$$

with $\nabla g(x)=(\nabla g_1(x), \ldots, \nabla g_m(x))$.

The point (x^*, u^*) fulfils the second-order Kuhn-Tucker conditions (SKT) if the following two statements (i) and (ii) hold.

(i) Rank $(\nabla g(x^*))=m$.

(ii) The relation $\nabla g(x^*)^T s=0$, $s\neq 0$ implies $s^T L_{xx}(x^*, u^*)s>0$.

A direction s' with $\nabla g(x^*)^T s'=0$ and $s'^T L_{xx}(x^*, u^*)s'<0$ is called a direction of descent. If (x^*, u^*) fulfils SKT, then x^* is a local minimizer (*cf.* e. g. [3]), and if (x', u') fulfils the conditions (i) and (ii), then $\nabla^2 L(x', u')$ is regular (*cf.* e. g. [3]). For solving (1), we use the Newton method. With $z=(x, u)$, we obtain:

$$\nabla L(z^k)+\nabla^2 L(z^k)s^k=0 \tag{2}$$

and

$$z^{k+1}:=z^k+t_k s^k$$

with $s=(sx, su)$, where t_k denotes the step length, or we use two step sizes t_k' and t_k'' with:

$$x^{k+1}=x^k+t_k' sx^k \qquad \text{and} \qquad u^{k+1}=u^k+t_k'' su^k.$$

Now we consider the Lagrangian dual problem of (P) in a slightly different form.

$$\text{(L)} \qquad \varphi(u)\to \sup, \qquad u\in R^m$$

with

$$\varphi(u):=\operatorname{loc\,min}(L(x, u): x\in R^n).$$

We define $h(x')=\operatorname{loc\,min} h(x)$ as follows: there exists an open neighbourhood X of x' in such a way that $h(x')\leqq h(x)$ holds for all $x\in X$. Let $\operatorname{dom}\varphi:=(u:\varphi(u)>-\infty)$. If we use the dual problem, then we assume dom $\varphi\neq\varnothing$.

If we solve (L) with Newton's method, then we obtain:

$$\nabla\varphi(u^k)+\nabla^2\varphi(u^k)\,r^k=0, \tag{3}$$

and $u^{k+1}:=u^k+t_k\,r^k$.

Now we use the decomposition for solving (P). If we split the vector x into:

$$x=(x_1, x_2), \qquad x_1\in R^{n1}, \quad x_2\in R^{n2} \quad \text{with } n1+n2=n,$$

and $n2\geqq m$, then we have:

$$\text{(D)} \qquad F(x_1)\to\min, \qquad x_1\in R^{n1},$$

where

$$F(x_1):=\operatorname{loc\,min}\{f(x_1, x_2): g(x_1, x_2)=0,\ x_2\in R^{n2}\}.$$

The function F is discussed in [2] and [5]. Here, we consider only the regular case. The Newton method for (D) takes the following form:

$$\nabla F(x_1^k)+\nabla^2 F(x_1^k)\,v^k=0 \tag{4}$$

and

$$x_1^{k+1}:=x_1^k+t_k\,v^k.$$

We ask for a Kuhn-Tucker point (x^*, u^*) which satisfies *SKT*, *i.e.* we want to have a local minimizer x^*.

3. THE DERIVATIONS

Now we ask how we can calculate the first and second derivations in (2), (3) and (4) efficiently.

Let

$$\nabla^2 L(x, u)=\begin{bmatrix} L_{xx}(x, u) & \nabla g(x) \\ \nabla g(x)^T & 0 \end{bmatrix}=:\begin{bmatrix} C & A^T \\ A & 0 \end{bmatrix}=K. \tag{5}$$

If the matrices C and K are regular, then we obtain:

$$K^{-1}=\begin{bmatrix} C^{-1}-C^{-1}A^T PAC^{-1} & C^{-1}A^T P \\ PAC^{-1} & -P \end{bmatrix}$$

with $P=(AC^{-1}A^T)^{-1}$.

Denote

$$D:=K^{-1}=\begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}, \tag{6}$$

where D_{11} is of dimension (n, n).

We now have the following assertions:

THEOREM 1 (*cf.* e.g. [3]): *Let* $u' \in \operatorname{dom} \varphi$ *and let* (x', u') *be a point with* $\varphi(u')=L(x', u')$ *further let* $L_{xx}(x', u')$ *be regular. If matrix* D_{22} [*via* (6)] *is regular, then the following statements hold:*

$$\nabla \varphi(u')=g(x')$$

and

$$(\nabla^2 \varphi(u'))^{-1}=D_{22}.$$

Now we split matrix D_{11} via (6) into

$$D_{11}=\begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix},$$

where H_{11} is of dimension $(n1, n1)$.

Let (x'_2, u') be a Kuhn-Tucker point of the following problem:

$$f(x_1, x_2) \to \min \text{ subject to } g(x_1, x_2)=0 \; (x_1 \text{ constant}). \tag{7}$$

THEOREM 2 (*cf.* [11]): *Let matrix* H_{11} *be regular. Let* (x'_2, u') *be defined via* (7) *and let* $F(x'_1)=f(x'_1, x'_2)$ *then the next statements are true:*

$$\nabla F(x'_1)=f_{x1}(x')+g_{x1}(x')u'$$

and

$$(\nabla^2 F(x'_1))^{-1}=H_{11}.$$

The Theorems 1 and 2 show that we have only to calculate the matrix D as (5). From this matrix we can obtain the other second order derivatives efficiently.

4. ON THE SOLUTION OF THE LINEARIZED KUHN-TUCKER EQUATIONS

To solve the linear equation system (2) we use the following pivoting procedure (*cf.* [7]). We compute the matrix K^{-1} [K via (5)]. Let K be split in the following way:

$$K := \begin{pmatrix} C_{11} & C_{12}^T & A_1^T \\ C_{12} & C_{22} & A_2^T \\ A_1 & A_2 & 0 \end{pmatrix}.$$

We assume that A_2 is regular.

Pivoting procedure

S1: Compute A_2^{-1}.

S2: Calculate

$$P = \begin{bmatrix} C_{22} & A_2^T \\ A_2 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & A_2^{-1} \\ A_2^{-T} & -A_2^{-T} C_{22} A_2^{-1} \end{bmatrix}.$$

S3: Evaluate

$$U = \begin{pmatrix} T & (C_{21}^T, A_1^T)\, P \\ -P \begin{bmatrix} C_{21} \\ A_1 \end{bmatrix} & P \end{pmatrix}$$

with

$$T = C_{11} - \begin{bmatrix} C_{21} \\ A_1 \end{bmatrix}^T P \begin{bmatrix} C_{21} \\ A_1 \end{bmatrix}.$$

S4: Invert matrix T. Simultaneously convert all the remaining parts of the matrix U. We obtain:

$$\begin{pmatrix} T^{-1} & -T^{-1} (C_{21}^T, A_1^T)\, P \\ -P \begin{bmatrix} C_{21} \\ A_1 \end{bmatrix} T^{-1} & P \begin{bmatrix} C_{21} \\ A_1 \end{bmatrix} T^{-1} (C_{21}^T, A_1^T)\, P + P \end{pmatrix}$$

It is not hard to see that we have calculated K^{-1}.

Now we can state the following theorem:

THÉORÈME: 3 (*cf*. [7]) *Let* (x^*, u^*) *be a Kuhn-Tucker point of problem* (P), *and let* K *be regular. Then we obtain:*

(a) x^* *is a local minimizer iff* T *is positive definite.*

(b) x^* *is a local maximizer iff* T *is negative definite.*

(c) x^* *is a "saddle" point in the remaining case.*

In case K is regular, then T is regular.

We can check the positive definiteness of T by the pivoting procedure. We can choose all pivots on the main diagonal and all pivots are positive iff T is positive definite. This procedure also works if we use the LU-decomposition.

Now we try to find a vector r_1' with $r_1'^T T r_1' < 0$. If T is regular and not positive definite, then there exists such a vector. Now we consider matrix U with:

$$U = \begin{pmatrix} T^* \\ T1^* \\ T2^* \end{pmatrix},$$

and calculate $r_2 = T1\, r_1'$. The vector $s = (r_1', r_2)$ fulfils $(A_1, A_2)s = 0$ and is a direction of descent (*cf*. [7]).

We now extend the pivoting procedure. Sort K corresponding (7) and divide this matrix into the following blocks:

$$\begin{bmatrix} K_1 & K_2 \\ K_3 & K_4 \end{bmatrix},$$

where K_4 is a $(n2, n2)$ matrix which is the matrix K for problem (7).

Modified pivoting procedure

S1′: We use the pivoting procedure for K_4 and convert all the remaining parts of K. We obtain a matrix $T2$ ($T2 := T$ in step S3).

S2′: After step S4 for matrix K_4 we consider the resulting matrix K_1^* (K_1^* is the transformed matrix K_1). Now we use the pivoting procedure for this matrix and obtain a matrix $T1$.

Now we state the following theorem:

THEOREM 4: *Let matrix T be defined by the pivoting procedure and let the matrices $T1$ and $T2$ be determined by the modified pivoting procedure. Let K be regular.*

(*a*) *T is positive definite iff $T1$ and $T2$ are both positive definite.*

(*b*) *T is negative definite iff $T1$ and $T2$ are both negative definite.*

(*c*) *In the remaining cases is T indefinite.*

The proof of Theorem 4 is obvious and should be omitted. From this theorem we obtain

COROLLARY 5: *In order to obtain a local minimizer (maximizer) for (P) with the decomposition, we need a local optimum of problem* (7) *as well as of problem (D).*

5. REMARKS ON THE LINEARIZATION

The Hessian $\nabla^2 L$ contains a lot of information. If we require that the corresponding matrices be regular, then the different methods are equivalent in the sense that all information is contained in $\nabla^2 L$.

For this purpose, we use the inverse matrix. All our statements are related only to pivots and their signs, respectively. Consequently, it is possible to connect the solution process with the numerically stable pivot rules from the LU-decomposition (*cf.* e. g. [6]).

ACKNOWLEDGMENTS

I would like to express my thanks to the amonymous referees for his valuable suggestions.

REFERENCES

1 A. R. CONN and N. I. M. GOULD, On the Location of Indefinite Descent for Nonlinear Programming Algorithm, *SIAM J. Num. Anal.*, 1984, *21*pp. 1162-1179.

2. A. L. DONTCHEV and H. Th. JONGEN, On the Regularity of the Kuhn-Tucker Curve, *SIAM J. Control Optimization 1986, 24,* pp. 169-176.

3. R. FLETCHER, Practical Methods of Optimization, *John Wiley*, Chichester, 1981.

4. S. P. HAN, A Hybrid Method for Nonlinear Programming, Nonlinear Programming 3, *Academic Press*, New York, 1978, pp. 65-95.

5. H. Th. JONGEN, T. MÖBERT and K. TAMMER, On Iterated Minimization in Nonconvex Optimization, Report 488, Twente University of Technology, N. L.

6. A. KIELBASINSKI and H. SCHWETLICK, Numerische Lineare Algebra, *Dt. Verlag d. Wiss.*, Berlin, 1988.
7. F. KÖRNER and B. LUDERER, A Simultaneous Method for Checking Second-Order Kuhn-Tucker Conditions for Equality Constrained Nonlinear Programming Problems, *Syst. Anal. Sim. Mod.*, 1988, *5*, pp. 51-57.
8. F. KÖRNER, Remarks on Second-Order Conditions in Connection with the Algorithm of Beale for Quadratic Programming, *Europ. J. Oper. Res.*, 1989, *40*, pp. 85-89.
9. G. P. MCCORMICK, Nonlinear Programming. Theory, Algorithms, and Applications, *John Wiley*, Chichester, 1983.
10. H. SCHWETLICK, Numerische Lösung nichtlinearer Gleichungssysteme, *Dt. Verlag d. Wiss.*, Berlin, 1979.
11. A. SHAPIRO, Second-Order Derivatives of Extremal-Value Functions and Optimality Conditions for Semi-Infinite Programs, *Math. Oper. Res.*, 1985, *10*, pp. 207-219.
12. A. SHAPIRO, Second-Order Sensitivity Analysis and Asymptotic Theory of Parametrized Nonlinear Programs, *Math. Prog.*, 1985, *33*, pp. 280-299.
13. J. E. SPINGARN and R. T. ROCKAFELLAR, The Generic Nature of Optimality Conditions in Nonlinear Programming, *Math. Oper. Res.*, 1979, *40*, pp. 425-430.