Yugoslav Journal of Operations Research
5 (1995), Number 2, 221-232

A DISCRETE PROJECTION QUASI-NEWTON METHOD FOR LINEARLY CONSTRAINED PROBLEMS

Nada I. ĐURANOVIĆ-MILIČIĆ *)
Department of Mathematics, Faculty of Technology and Metallurgy,
University of Belgrade, Karnegijeva 4,
11000 Beograd, Yugoslavia

Abstract: In this paper we define a discrete quasi-Newton algorithm which uses only function values for finding an optimal solution to the problem $\min\{\ \varphi(x) \mid x \in X\ \}$, where X is a convex polytope. It is shown that using this algorithm one can reduce the initial problem to a finite number of subproblems of the type $\min\{\ \varphi(x) \mid x \in C\ \}$, where C is a linear manifold. It is also shown that each cluster point of the sequence generated by the algorithm is an optimal point of the considered optimization problem.

Keywords: Linear inequality-constrained minimization, Discrete quasi-Newton method, Projected gradient, Cholesky factorization.

1. INTRODUCTION

We consider the following minimization problem:

$$\min\{\ \varphi(x) \mid x \in X\ \},\quad X = \{\ x \in R^n \mid Ax \geq b\ \}, \tag{1}$$

where A is an $m \times n$ matrix, b an m-vector.

We use in the sequel the following notation. We define the index set

*) This research was supported by Science Fund of Serbia, grant number 0401F, through Mathematical Institute

$I(x) = \{ \ j \in \{ \ 1, \ldots, m \ \} \mid a_{jq}^T x = b_j \ \}$, the matrix $\hat{A} = (a_j), \ j \in I(x)$ associated with the corresponding working set; q denotes the number of constraints in the working set. By q Z we denote a matrix whose columns form a basis for the null space of $\hat{A}$, $g(x)$ denotes the gradient $\nabla\varphi(x)$ and $g_p(x) = Z^T g(x)$ denotes the projected gradient. The vector e_i denotes the i-th unit vector of the appropriate dimension.

Throughout the paper we make the following assumptions:

*A*1 • φ: $R^n \rightarrow R$ is a convex twice continuously differentiable function;

*A*2 • $X \neq 0$ and there exists a point $\overline{x} \in X$ such that the level set
$L = \{ \ x \in X \mid \varphi(x) \leq \varphi(\overline{x}) \ \}$ is bounded;

*A*3 • the matrix $\hat{A}_0$ associated with the initial working set has full row rank.

If (1) is solved by quasi-Newton method, the step s from the current iterate x can be defined as $s = Z s_p$, where

$$B_p s_p = -g_p(x). \tag{2}$$

Here B_p denotes the current quasi-Newton approximation to the projected Hessian. The next iterate is obtained as $x + \alpha s$, where $\alpha > 0$ is the steplength parameter.

The BFGS (Broyden-Fletcher-Goldfarb-Shanno) quasi-Newton formula for the updated projected matrix $\overline{B}_p$ is (see [7]):

$$\overline{B}_p = B_p + \frac{1}{g_p^T s_p} g_p g_p^T + \frac{1}{\alpha_p^T s_p} y_p y_p^T , \tag{3}$$

where $s_p = Z^T s, \ y_p = Z^T y, \ y = g(x + \alpha s) - g(x)$.

The earliest quasi-Newton method for linear inequality constraints was due to Goldfarb, [8], and was based on an extension of Davidon's method.

When a quasi-Newton method is implemented using updates to the Cholesky factors, it is possible to avoid losing positive-definiteness. The Cholesky factorization of the appropriate projected Hessian can be updated after each rank-one modification. This procedure ensures that the matrix B_p is always positive definite and hence that the computed search direction is always a descent direction. The use of the Cholesky factorization as an aid to the implementation of a numerically stable quasi-Newton method was suggested by Gill and Murray, [3], Fletcher and Powell, [2].

In the algorithm presented in this paper we use the finite difference approximation g_p to $g_p(x)$. Namely, since only the projected gradient is required to perform the update (3), the vector $g_p(x)$ can be approximated directly by taking finite-

-differences of φ along the $n-q$ columns of Z rather than with respect to each variable. Then, using (2) and (3) we obtain the search direction. The steplength parameter α along this direction is computed by the Armijo rule.

Since the direction s is defined by (2) this method falls into the category of null-space methods, which tend to be more efficient as the number of constraints in the working set increases. The larger the number of constraints active at x, the smaller the dimension of the system of equations (2) to be solved for s.

In such a kind of algorithms a good choice of Z plays a very important role. A full discussion of different formulations for Z is given in [1], [5], [10], [11], [8].

Now we shall present the algorithm. For simplicity, we drop the subscript k associated with the current iteration. It is assumed that an initial feasible point x_0 is known. A technique for computing such a point is discussed, for example, in [4]. Because of assumptions $A1$ and $A2$ it follows that the level set $L = \{ x \in X \mid \varphi(x) \le \varphi(x_0) \}$ is compact.

2. A DISCRETE PROJECTION QUASI-NEWTON METHOD

Firstly we shall present the complete algorithm. After that its constituent parts will be presented. The complete algorithm is as follows.

The Main Algorithm

```
$k \leftarrow 0$;  $p \leftarrow 0$;

Given a feasible point $x_0$, a full-rank working set $\hat{A}_0$ and scalars $\varepsilon_1, \varepsilon_2$,
$0 \le \varepsilon_2 < \varepsilon_1 < 1$, $\gamma \in (0,1)$; and a positive definite matrix $B_p^0$ ;
Determine $Z_0$ and a finite - difference interval- vector $h_0$;

repeat
   repeat
      near stationary point $\leftarrow \|g_p^k\| \le \gamma^p \varepsilon_1$ ;
      if not near stationary point then
         Apply Procedure $s$
      end if
   until near stationary point;
   Compute the Lagrange multipliers estimates $\lambda$
   if $\lambda \ge 0$ then
      converged $\leftarrow \gamma^p \varepsilon_1 \le \varepsilon_2$;
      if not converged then
```

```
          Apply Procedure $s$
      end if
    else
      $\nu$ – the most negative Lagrange multiplier estimate $\lambda_s$;
      if $\nu \le -\gamma^p \varepsilon_1$ then
        Delete constraint with multiplier estimate $\lambda_s$
        $q \leftarrow q-1$;
        $Z \leftarrow Z_d$;  $B_p \leftarrow B_p^d$;
        Determine a new interval - vector $h$;
    else
      if $\gamma^p \varepsilon_1 \le \varepsilon_2$ then
        Apply $\varepsilon$ - procedure
      else
        Apply Procedure $s$
      end if
    end if
  end if
until converged
```

Procedure s

From Main Algorithm available: a feasible point x_k, a full-rank working set $\hat{A}_k$, a matrix Z_k, an update B_p^k, an approximation to the projected gradient g_p^k and an interval-vector h_k:

```
$s_p \leftarrow$ a solution $\bar{s}$ of the system of equations $B_p s = -g_p$;
$s \leftarrow Z s_p$;
$\alpha \leftarrow$ an Armijo feasible step along $s$;
$\alpha_c \leftarrow$ the step to the nearest constraint;
$\bar{x} \leftarrow x + \alpha s$;  $k \leftarrow k+1$;
$\bar{g}_p \leftarrow$ an approximation to the projected gradient at $\bar{x}$;
if $\alpha < \alpha_c$ then
  $y_p \leftarrow \bar{g}_p - g_p$;
  $B_p \leftarrow B_p + \frac{1}{g_p^T s_p} g_p g_p^T + \frac{1}{\alpha y_p^T s_p} y_p y_p^T$;
  $x \leftarrow \bar{x}$;  $p \leftarrow p+1$;
else
  Add constraint to the working set
  $q \leftarrow q+1$;
  $Z \leftarrow Z_a$;  $B_p \leftarrow B_p^a$;  $x \leftarrow \bar{x}$;
```

```
    Determine a new interval- vector $h$;
    end if
```

ε - procedure

From Main Algorithm available: a feasible point x_k, a full-rank working set $\hat{A}_k$, a Lagrange multipliers estimate λ_k, a scalar ε_2, $0 < \varepsilon_2 < 1$, a set of indices $I_s = \{ i \mid -\varepsilon_2 < [\ \lambda_k\ ] \le 0 \}$, an update B_p^k; an interval-vector h_k; $\hat{A}^+ \leftarrow$ the pseudo inverse of $\hat{A}$:

```
$s(\varepsilon_2) \leftarrow \varepsilon_2 \sum_{m \in I_s} (\hat{A}^+)^T e_m$;
$x(\varepsilon_2) \leftarrow x + s(\varepsilon_2)$;
```

Compute the Lagrange multipliers estimate $\lambda^{\varepsilon 2}$ corresponding to the point $x(\varepsilon_2)$:

```
$(\lambda^{\varepsilon 2} - \lambda)_s \leftarrow \min\{ (\lambda^{\varepsilon 2} - \lambda_i),\ i \in I_s \}$;
converged $\leftarrow (\lambda^{\varepsilon 2} - \lambda)_s > 0$;
if not converged then
    Delete constraint with multiplier estimate $\lambda_s$
    $q \leftarrow q - 1$;
    $Z \leftarrow Z_d$; $B_p \leftarrow B_p^d$;
    Determine a new interval- vector $h$;
end if
```

The Algorithm for Computing the Approximation to the Projected Gradient

```
$i \leftarrow 1$;
Given a feasible point $x_k$, a full - rank working set $\hat{A}_k$;
Determine $Z_k$ and a finite - difference interval- vector $h_k$;
while $i \le n - q$ do
    $\bar{e}_i \leftarrow Z e_i$;
    $z_j \leftarrow$ the $j$ - th element of the vector $\bar{e}_i$;
    $\bar{h}_j \leftarrow \sum_{j=1}^{n} | h_j z_j |$;
    $x_i \leftarrow x + \bar{h}_i \bar{e}_i$;
    $[\ g_p\ ]_i \leftarrow \frac{1}{h_i}[\ \varphi(x_i) - \varphi(x)\ ]$;
    $i \leftarrow i+1$;
end do
```

(By $[\ g_p\ ]_i$ we denote the forward-difference approximation to an i-th element of the projected gradient $g_p(x_k)$).

Note 1. When we add or delete a constraint we must modify the representation of Z to correspond to the new working set. If we add a constraint, the relationship between the old Z and new Z_a may be expressed as:
$Z\bar{Q} = (Z_a\, \bar{z})$, where $\bar{z}$ is the q-th column of $Q\bar{Q}$ and $\bar{Q}$ denotes a sweep of plane rotations (see [8]). The corresponding relationship between the old B_p and the new B_p^a (see [8]) is:

$$\bar{Q}^T B_p \bar{Q} = \begin{vmatrix} B_p^a & \\ & 1 \end{vmatrix},$$

that is, we delete the last row and column from the matrix $\bar{Q}^T B_p \bar{Q}$ to get the matrix B_p^a (the dimension of B_p^a is smaller by one).

If we delete a constraint, the relationship between the old Z and new Z_d may be expressed as $Z_d = (Z\, \bar{z})$, where $\bar{z}$ is the $(q+1)$-th column of $Q\bar{Q}$ (see [8]). The corresponding relationship between the old B_p and the new B_p^d (see [8]) is:

$$Z_d^T B_p Z_d = B_p^d = \begin{vmatrix} B_p & \\ & 1 \end{vmatrix},$$

that is, the dimension of B_p^d is larger by one.

Note 2. We use the finite (forward)-difference approximation to the projected gradient. We wish to obtain a "sensible" interval-vector h that will produce reasonably accurate approximations to the projected gradient throughout the course of the minimization (see 4.6.1.3. and 8.6.25. in [7]). A vector h (that is, a set of intervals $\{h_i$, $i = 1,....,n\}$should be specified at the initial point x_0 . Unfortunately, since Z changes completely when the working set changes, there is no guarantee that the set of optimal intervals computed at one point has any straightforward relationship to the set of optimal intervals at another point. One strategy that often gives good results for lineraly constrained problems used in the presented method is discussed in [7].

Note 3. As we know, at the solution x^* to (1) it holds that

$$g(x^*) = \hat{A}^T \lambda^* \tag{4}$$

for some vector of Lagrange multipliers. Although there are no Lagrange multipliers at a non-stationary point, it is important to have some means of estimating Lagrange multipliers at points for wich (4) does not hold. Any such estimate should be consistent, i.e. λ_k should have the proprty that $x_k \to x^*$ implies $\lambda_k \to \lambda^*$.

In addition, a method for estimating Lagrange multipliers should use the same factorization involved in representing Z. Possible forms for a Lagrange multiplier estimate are given in [6], [12].

Note 4. The known property of a convex continuously differentiable function $\varphi: R^n \to$ $\to R$ is that $[\ g(x+\alpha s)\ ]^T s$ is a nondecreasing function of α.

Consequently, for any $\overline{\alpha} \geq 0$ we have $[\ g(x+\overline{\alpha}s)\ ]^T s \geq [\ g(x)\ ]^T s$. For some $\alpha > \overline{\alpha}$ it follows that $[\ g(x+\alpha s)\ ]^T s \geq [\ g(x+\overline{\alpha}s)\ ]^T s \geq [\ g(x)\ ]^T s$, that is, for such α we have $[\ g(x+\alpha s)\ ]^T s \geq [\ g(x)\ ]^T s$. Hence, the condition $y^T s = [\ g(x+\alpha s)\ ]^T s$ $-[\ g(x)\ ]^T s > 0$ is satisfied. (Since we know (see [7]) that the BFGS update used in the Projection Quasi-Newton Algorithm has the property of hereditary positive definiteness if and only if $y^T s > 0$).

3. CONVERGENCE

The following lemma shows that linear independence of the working sets is maintained.

Lemma 1. If the matrix $\hat{A}_0$ associated with the initial working set has full row rank, and if s_k satisfies $\hat{A}_k s_k = 0$ for every $k \geq 0$, then $\hat{A}_k$ will have full row rank for all k>0.

Proof. See [8].

In the following lemma we shall show that we may delete any constraint corresponding to a negative λ_i.

Lemma 2. Let x belong to the manifold defined by the index set $I = I(x)$ and suppose $g_p = 0$. If we delete from the working set the constraint corresponding to a negative λ_s then the vector $s = -ZB_p^{-1}\,\overline{g}_p$ in the subspace corresponding to the new index set $\overline{I} = I \setminus \{i_s\}$ is a feasible descent direction because the conditions

$$a_s^T(x+\alpha s) > b_s, \quad \alpha > 0 \text{ and } g(x)^{T_s} < 0$$

are satisfied, where $\overline{g}_p$ denotes the projection of g onto the new subspace.

Proof. We shall assume $g_p = 0$. By g we denote the finite-difference approximation to $g(x)$.

It follows that $g = \sum_{j\in I} \lambda_j a_j$. Consequently,

$$\bar{g}_p = \bar{Z}^T g(x) = \bar{Z}^T \sum_{j\in I} \lambda_j a_j = \lambda_s \bar{Z}^T a_s \neq 0 \tag{5}$$

since $\lambda_s < 0$ and the vectors a_j, $j \in I$ are, by assumption *A3*, linearly independent. By $\bar{Z}$ we denote a basis matrix for the null space of the new working set $\hat{A} = (a_j)$, $j \in \bar{I}$. We have

$$g^T s = -g^T \bar{Z} B_p^{-1} \bar{g}_p = -(\bar{Z}^T g)^T B_p^{-1} \bar{Z}^T g = -\bar{g}_p^T B_p^{-1} g_p < 0 \tag{5a}$$

since B_p is positive definite and $\bar{g}_p \neq 0$ by (5).

Since $\|g - g(x)\| = 0(h)$ (see [7]), by the continuity of the product $(\bullet)^T s$ it follows from (5a) that $g(x)^T s < 0$, too. Finally,

$$a_s^T s = -a_s^T \bar{Z} B_p^{-1} \bar{Z}^T \sum_{j\in I} \lambda_j a_j = -\lambda_s (\bar{Z}^T a_s)^T B_p^{-1} (\bar{Z}^T a_s) > 0 \tag{6}$$

because B^{-1} is positive definite, $\lambda_s < 0$ and $\bar{Z}^T a_s \neq 0$ by linear independence of vectors a_j, $j \in I$ (follows from assumption *A3* and Lemma 1). Since $a_s^T x = b_s$, we obtain, using (8) $a_s^T (x + \alpha s) = a_s^T x + \alpha a_s^T s > b_s$, $\alpha > 0$, and that is what we had to prove.

Note. However, when a constraint is deleted from the working set based on a negative multiplier estimate, it is not true that every definition of the search direction with the new working set will be feasible with respect to the deleted constraint. See [6] for a more detailed discussion of techniques for testing the accuracy of Lagrange multiplier estimates and of the conditions that must apply in order to guarantee a feasible descent direction for quasi-Newton methods.

In the following lemma near-zero Lagrange multipliers estimates are considered. By I_s we denote the set of indices of contraints with near-zero multipliers and the corresponding point by $\bar{x}$. If each constraint $a_j^T x = b_j$, $j \in I_s$ is perturbed by a small positive scalar ε and the Lagrange multipliers estimates are recomputed it is possible to determine the nature of the curvature in the neighborhood of $\bar{x}$.

Lemma 3. Let $x(\varepsilon)$ be such that

$$a_j^T x(\varepsilon) = \begin{cases} b_j + \varepsilon, & j \in I_s \\ b_j, & j \notin I_s \end{cases},$$

then $x(\varepsilon) = \bar{x} + s(\varepsilon)$, where $s(\varepsilon) = \varepsilon \sum_{j \in I_s} (\hat{A}^{+})^T e_j$, and $\hat{A}^{+} = (\hat{A}^T \hat{A})^{-1} \hat{A}^T$ is the pseudo-inverse of $\hat{A}$.

Proof. See [5].

This lemma implies that the decision to drop or retain a constraint with a near-zero multiplier will depend upon whether the sign of the corresponding element of $\lambda^{\varepsilon} - \lambda$ is negative or positive (λ^{ε} denotes the Lagrange multiplier estimate at the point $x(\varepsilon)$).

In the following theorem we shall give the proof of convergence.

Theorem. Let assumptions *A1-A3* be satisfied. Then each cluster point of the sequence generated by the algorithm presents an optimal point to the problem (1). This algorithm reduces the initial problem (1) to a finite number of subproblems of the type $\min\{ \varphi(x) \mid x \in C \}$, where *C* is a linear manifold.

Proof. By the Mean Value Theorem we have

$\varphi(x_k + \alpha s_k) = \varphi(x_k) + \alpha g(\xi_k)^T s_k, \quad \xi_k = x_k + \theta_k \alpha s_k, \quad \theta_k \in (0,1], \quad \alpha > 0.$

Since $g_k^T s_k = g_k^T Z(B_p^k)^{-1} g_p^k = -(Z^T g_k)^T (B_p^k)^{-1} g_p^k = -(g_p^k)^T (B_p^k)^{-1} g_p^k < 0$ if $g_p^k \neq 0$ by assumed positive definiteness of B_p^k, and since $\| g(x_k) - g_k \| = 0(h_k)$, it follows by the continuity of the product $(\bullet)^T s_k$ that $g(x)^T s_k < 0$ for x in some interval $(\underline{x}, \bar{x})$ about x_k. Hence, by the continuity of $g(x)$ we have that $g(\xi_k)^T s_k < 0$ for sufficiently small $\hat{\alpha}$, $0 < \alpha \leq \hat{\alpha}$. Hence, $\varphi(x_k + \alpha s_k) < \varphi(x_k)$, that is, $\{ x_k \} \subset L = \{ x \in X \mid \varphi(x) \leq \varphi(x_0) \}$, where L is by assumptions *A1* and *A2* a compact set. Consequently there exists a subsequence $\{ x_{k_j} \}$ such that $x_{k_j} \to x^* \in L$ as $j \to \infty$. By continuity of φ we have $\varphi(x_{k_j} \} \to \varphi(x^*)$ as $j \to \infty$. From the Algorithm it follows that $g_p^k \to 0$ as $k \to \infty$. By continuity of $g_p(x)$ it follows that $g_p(x_{k_j}) \to g_p(x^*) = 0$ as $j \to \infty$. Since from the Algorithm we have $\lambda(x^*) \geq 0$, it follows that x^* is an optimal point to the problem (1).

Let us prove the second part of the theorem. At a k-th iteration one of the two following cases can be realized. For simplicity, we drop the subscript k associated with the current iteration.

Case 1. x is a near optimal point of φ on the subspace corresponding to the index set $I = I(x)$; that is, we obtain

$$\| g_p \| \leq \gamma^p \varepsilon_1.$$

After some finite number of steps we shall ge

$$\| g_p \| \leq \varepsilon_2, \gamma^p \varepsilon_1 \leq \varepsilon_2 \text{ and } \lambda \geq 0.$$

that is, we find an optimal point $x = x_{opt}$ to the problem (1).

If $\| g_p \| \leq \gamma^p \varepsilon_1$ and λ_s- the most negative Lagrange multiplier estimate satisfies the condition $\lambda_s \leq -\gamma^p \varepsilon_1$ we delete the constraint corresponding to the index i_s and repeat the procedure of optimization on the new so defined subspace.

Or, after some finite number of steps we shall obtain

$$\| g_p \| \leq \varepsilon_2, \quad \lambda_s > -\varepsilon_2, \quad \gamma^p \varepsilon_1 \leq \varepsilon_2,$$

and we proceed to check optimality conditions on an ε_2 -active manifold, that is, to check the sign of

$$(\lambda^{\varepsilon 2} + \lambda)_i, \quad i \in I_s,$$

where I_s denotes the index set of the near-zero multipliers. If $\min \{ (\lambda^{\varepsilon 2} - \lambda)_i, \ i \in I_s \} = (\lambda^{\varepsilon 2} - \lambda)_s < 0$ we delete the constraint corresponding to the index i_s and repeat the procedure of minimization on the new so defined subspace; otherwise we have found an optimal point $x = x_{opt}$.

During the above described procedure we must stay in the feasible region X. Our step-size is $\alpha = \min \{ \overline{\alpha}, \alpha_c \}$, where $\overline{\alpha}$ denotes an Armijo step along the given direction and α_c denotes the step to the nearest inactive constraint. If $\alpha = \alpha_c$, the working set is modified to indicate that a new constraint is satisfied exactly. After the constraint defined by α_c is added to the working set, $\hat{A}$ and Z change accordingly and the next iteration proceeds in the same fashion.

In this way,either we find a near optimal point of φ on the subspace corresponding to $I(x)$, or the process is interrupted if $\overline{\alpha} = \alpha_c$. In both cases the last obtained point is used as a starting point and the minimization process proceeds. From Lemma 2 and the first part of Theorem it follows that the sequence of function values $\{ \varphi(x_k) \}$ is monotonically decreasing.

Case 2. x is not a near optimal point of φ on the subspace corresponding to the index set $I = I(x)$; that is, we obtain

$$\| g_p \| > \gamma^p \varepsilon_1.$$

In that case we proceed to look for the minimum of φ on the current subspace until we either find an optimal point on that subspace ($\| g_p \| \leq \gamma^p \varepsilon_2$) ,or we attach a new constraint if $\bar{\alpha} = \alpha_c$. In both cases the last obtained point is used as a starting point and the minimization proceeds.

In both cases we also obtain decreasing values of φ since we make a positive step $\alpha > 0$ along a descent feasible direction $s = -ZB_p^{-1}g_p$ (by positive definiteness of B_p). (If $\bar{\alpha} = \alpha_c$, we have $\alpha_c = \min\limits_j \{ \frac{b_j - a_j^T x}{a_j^T s}, \ a_j^T s < 0, j \notin I \}$. Since $b_j - a_j^T x < 0$, $j \notin I$, it follows that $\alpha_c > 0$).

From the above consideration of Case 1 and Case 2 it follows that either we find an optimal point on a subspace or we add a new constraint to the set of active constraints. Since, by assumption $A3$ and Lemma 1 the vectors a_i, $i \in I(x)$ are linearly independent, such an extension of the set of active constraints must be finished after a finite number of iterations, which cannot be larger than n, where n is the dimension of the problem (1). Since the sequence $\{\varphi(x_k)\}$ is monotonically decreasing, all the sets $I(x_k)$ are different. Since all index sets $I(x_k) \subset \{1, \ldots, m\}$ it follows that their number is finite (the number of subsets of a set consisting of m elements). Therefore the number of problems $\min\{\varphi(x) \mid x \in C\}$, where C is a manifold corresponding to the index set $I(x_k)$ must be finite.

4. CONCLUSION

We can conclude that the presented method has the following advantages:

1. This algorithm requires only evaluations of function values: namely a finite-difference quasi-Newton algorithm will differ in other respects from a method for exact gradients because of the $n - q$ function evaluations required to obtain a projected gradient approximation.

2. With the values of intervals $\{h_i, i = 1, \ldots\ldots, n\}$ determined according to Gill and Murray (1972, 1981) the cheaper forward-difference formula is used until x_k is in a neighbourhood where $\| g_p \|$ is small, in which event a switch to central differences is made. At present, quasi-Newton methods using these techniques are generally more effective than any other methods which utilize only function values.

3. The use of the Cholesky factorization allows us to avoid the loss (through rounding errors) of positive definiteness in the Hessian (or inverse Hessian) approximation.

4. When properly implemented, finite-difference quasi-Newton methods are extremely efficient and have the same rapid convergence as their counterparts with exact gradients.

REFERENCES

[1] Fletcher, R., "Minimizing general functions subject to linear constraints", in: F.A.Lootsma, (ed.) *Numerical Methods for Non-linear Optimization*, Academic Press, London and New York, 1972, 279-296.

[2] Fletcher, R. and Powell, M.J.D., "On the modification of LDL^T factorizations", *Mathematics of Computation*, 28 (1974) 1067-1087.

[3] Gill, P.E. and Murray, W., "Quasi-Newton methods for unconstrained optimization", *J.Inst. Maths. Applics*. 9 (1972), 91-108.

[4] Gill, P.E. and Murray, W., "A numerically stable form of the simplex algorithm", *Linear Algebra and its Applications*, 7 (1973) 99-138.

[5] Gill, P.E. and Murray, W., *Numerical Methods for Constrained Optimization*, Academic Press, London and New York, 1974.

[6] Gill, P.E. and Murray, W., "The computation of Lagrange multiplier estimates for constrained minimization", *Mathematical Programming*, 17 (1979) 32-60.

[7] Gill, P.E., Murray, W. and Wright, M.H., *Practical Optimization*, Academic Press, London and New York, 1981.

[8] Goldfarb, D., "Extension of Davidon's variable metric method to maximization under linear inequality and equality constraints", *SIAM Journal on Applied Mathematics*, 17 (1969) 739-764.

[9] McCormick, G.P., "The variable reduction method for nonlinear programming", *Management Science*, 17 (1970) 146-160.

[10] McCormick, G.P., "A second order method for the linearly constrained nonlinear programming problem", in: Rosen, J.B., Mangasarian, O.L. and Ritter, K., (eds) *Nonlinear Programming*, Academic Press, London and New York, 1970, 207-243.

[11] Peters, G. and Wilkinson, J.H., "The least squares problem and pseudoinverses", *Computer Journal* 13, (1970) 309-316.

[12] Pschenichniy, B.H. and Danilin, Y.M., *Chislennie metodi v ekstremalnih zadachah*, Nauka, Moskva, 1975.