Aplikace matematiky

Fridrich Sloboda

Nonlinear iterative methods and parallel computation

Aplikace matematiky, Vol. 21 (1976), No. 4, 252–262

Persistent URL: http://dml.cz/dmlcz/103645

NONLINEAR ITERATIVE METHODS AND PARALLEL COMPUTATION

FRIDRICH SLOBODA

(Received July 1, 1975)

1. INTRODUCTION

In this article we will consider methods for linear systems of equations, which can be applied to the solution of nonlinear systems and optimization problems. The class of these methods includes the nonlinear Gauss-Seidel iteration. Let $F : R_n \to R_n$ have components $f_1, f_2, \ldots, f_n$, then the generalized nonlinear Gauss-Seidel iteration is described as follows:

(1.1)
(i) Choose an index $i \in \{1, 2, \ldots, n\}$ and a vector $p_k \in R_n$;
(ii) solve $f_i(x_k - \alpha p_k) = 0$ for $\alpha = \alpha_k$;
(iii) set $x_{k+1} = x_k - \alpha_k p_k$.

For $p_k = e_{k(\mathrm{mod}\, n)+1}$ this reduces to the nonlinear Gauss-Seidel method. Similarly, the generalized SOR-Newton iteration is defined if α_k is one Newton step toward the solution of (1.1):

$$x_{k+1} = x_k - [f_i(x_k)/f_i'(x_k)p_k]\, p_k \,. \tag{1.2}$$

In (1.2) we follow the notation used in [1]. These methods exhibit only linear rate of convergence and the structure of the algorithms is purely sequential. In [13] a projection method for linear systems of equations is suggested. Let us consider the system

$$Ax = b \tag{1.3}$$

where A is a regular n by n matrix and b is an n-vector. Let $x_0^{(0)}, x_0^{(1)}, \ldots, x_0^{(n)}$ be $n + 1$ linearly independent points of the space R_n. Then the algorithm [13] is described by the recurrent relation

$$x_i^{(k)} = x_{i-1}^{(k)} + \frac{b_i - (a_i, x_{i-1}^{(k)})}{(a_i, v_{i-1}^{(i)})}\, v_{i-1}^{(i)} \tag{1.4}$$

where

$$v_{i-1}^{(i)} = x_{i-1}^{(i)} - x_{i-1}^{(i-1)} \quad i = 1, 2, \ldots, n; \quad k = i, i + 1, \ldots, n \,,$$

a_i is the i-th row of the matrix A and b_i is the i-th component of the vector b. Let $x_0^{(0)} = (0, \ldots, 0)^T$ and

$$x_0^{(k)} = (0, \ldots, t_k, \ldots, 0)^T \tag{1.5}$$

where $t_k = 1$. Let A be a strictly regular matrix, then

$$(a_i, v_{i-1}^{(i)}) \neq 0 ,$$

the matrix of the vectors $v_{i-1}^{(i)}$ is upper triangular with unit elements on the diagonal and the point $x_n^{(n)}$ is the solution of (1.3). Let A be a strictly regular, symmetric matrix. Let $x_0^{(0)} = (0, \ldots, 0)^T$ and let $x_0^{(k)}$ be in the form (1.5), then

$$(Av_{i-1}^{(i)}, v_{j-1}^{(j)}) = 0 \quad i \neq j .$$

According to (1.5) the algorithm (1.4) requires $O(\frac{1}{3}n^3)$ number of multiplications and additions. The symmetry of the matrix has no influence on the number of arithmetical operations, but allows to generate conjugate directions. The total storage requirements are less than $\frac{1}{4}n^2 + n + 2$. Input of data is very convenient for the algorithm since a single row of the matrix is required by each iteration.

In [14] a generalization of the described algorithm is suggested. Let us consider the quadratic function

$$f(x) = (Ax, x) - 2(b, x) + c \tag{1.6}$$

where A is a positive definite, symmetric matrix, b is the right-hand side vector of (1.3) and c is a scalar.

General algorithm [14].

Let $x_0^{(0)}, x_0^{(1)}, \ldots, x_0^{(n)}$ be $n + 1$ linearly independent points of the space R_n. Then the algorithm for minimization of (1.6) is defined as follows:

$$x_i^{(k)} = x_{i-1}^{(k)} + \alpha_{i-1}^{(k)} v_{i-1}^{(i)} \tag{1.7}$$

where

$$v_{i-1}^{(i)} = x_{i-1}^{(i)} - x_{i-1}^{(i-1)}$$

and $\alpha_{i-1}^{(k)}$ are scalar coefficients,

$$f(x_{i-1}^{(k)} + \alpha v_{i-1}^{(i)}) = \min! \quad i = 1, 2, \ldots, n ; \quad k = i, i + 1, \ldots, n . \tag{1.8}$$

$$\alpha = \alpha_{i-1}^{(k)}$$

It has been shown [14] that $v_{i-1}^{(i)}$ are mutually conjugate vectors and at the point $x_n^{(n)}$ function (1.6) achieves its minimum, i.e., the algorithm has the quadratic termination property.

Special case: Let $v_0^{(k)} = x_0^{(k)} - x_0^{(0)}$ be in the form

$$v_0^{(k)} = (0, \ldots, t_k, \ldots, 0)^T$$

where $t_k = 1$. Then for $\alpha_{i-1}^{(k)}$ defined by (1.8) it holds

$$\alpha_{i-1}^{(k)} = \frac{b_i - (a_i, x_{i-1}^{(k)})}{(a_i, v_{i-1}^{(i)})} \quad i = 1, 2, \ldots, n\,; \quad k = i, i+1, \ldots, n\,.$$

In the next sections we shall view the algorithm (1.7)–(1.8) as an iterative algorithm for finding the minimizer of a strictly convex function. As is known, it is not unreasonable to expect that any algorithm, which is convergent when applied to the minimization of a strictly convex function and which solves the problem of the minimization of (1.6) in a small number of steps, will also converge rapidly towards the end in the minimization of a strictly convex function. As we shall see the algorithm can be also considered a generalized algorithm of Gauss-Seidel type. We will also show that the sparseness of the occurrence matrix influences the structure of the algorithm. The algorithm will be analyzed from the parallel computation point of view.

2. DESCRIPTION OF THE ALGORITHM

Let $f : R_n \to R_1$ be a continuously differentiable strictly convex function. Let $x_0^{(0)}, x_0^{(1)}, \ldots, x_0^{(n)}$ be $n+1$ linearly independent points of the space R_n. Let $x_0^{(0)}$ be a starting point and let $x_0^{(k)} = x_0^{(0)} + v_0^{(k)}$, $v_0^{(k)} = (0, \ldots, t_k, \ldots, 0)^T$ where $t_k = \lambda$ is a suitable positive real number. Then the algorithm for the minimization of $f(x)$ is defined as follows:

(2.1) Algorithm.

Step(i): For given $x_0^{(0)}$, $x_0^{(k)} = x_0^{(0)} + v_0^{(k)}$ do the calculation by the recurrent relation

$$x_i^{(k)} = x_{i-1}^{(k)} + \alpha_{i-1}^{(k)} w_{i-1}^{(i)}$$

where

$$w_{i-1}^{(i)} = v_{i-1}^{(i)} / \|v_{i-1}^{(i)}\|\,, \quad v_{i-1}^{(i)} = x_{i-1}^{(i)} - x_{i-1}^{(i-1)}$$

and $\alpha_{i-1}^{(k)}$ is defined by

$$(2.2) \qquad f(x_{i-1}^{(k)} + \alpha w_{i-1}^{(i)}) = \min! \quad i = 1, 2, \ldots, n\,; \quad k = i, i+1, \ldots, n\,.$$

$$\alpha = \alpha_{i-1}^{(k)}$$

Step(ii): Replace $x_0^{(0)}$ by $x_n^{(n)}$ and go to Step(i).

Comment : $k = i$ it is more advantageous from the computational point of view to consider

$$x_i^{(i)} = x_{i-1}^{(i-1)} + \alpha_{i-1}^{(i-1)} w_{i-1}^{(i)}$$

where $\alpha_{i-1}^{(i-1)}$ is defined by

$$f(x_{i-1}^{(i-1)} + \alpha w_{i-1}^{(i)}) = \min!\,.$$
$$\alpha = \alpha_{i-1}^{(i-1)}$$

According to the choice of λ, which defines the vectors $v_0^{(k)}$ in Step(i), we obtain the following algorithms:

Algorithm I: $\lambda = h$, where h is a constant; for example $h = 1$ (0.5). In this case the vectors $w_{i-1}^{(i)}$ need not be in general normalized.

Algorithm II: $\lambda = \min(h, \|x_n^{(n)} - x_0^{(0)}\|)$.

Algorithm III: $\lambda = \min(h, |f(x_n^{(n)}) - f(x_0^{(0)})|)$.

Algorithm IV: $\lambda = \min(h, \|f'(x_0^{(0)})\|)$, where $f'(x) = (\partial f/\partial x_1, \ldots, \partial f/\partial x_n)$.

Theorem 1. *Let $f : R_n \to R_1$ be a continuously differentiable strictly convex function which satisfies*

$$\lim_{\|x\| \to \infty} f(x) = +\infty\,. \tag{2.3}$$

Then for any $x_0^{(0)} \in R_n$ the algorithm (2.1) is well defined and $x_n^{(n)}$ converges to the unique minimizer of $f(x)$. In order to prove this theorem we shall use definitions and theorems presented and proved in [1].

Definition 1. *If $f : D \subset R_n \to R_1$ then any nonempty set of the form $L(\gamma) = \{x \in D \,|\, f(x) \leqq \gamma\}$, $\gamma \in R_1$ is a level set of f.*

Definition 2. *A sequence $\{p^{(k)}\} \in R_n$, $p^{(k)} \neq 0$ is uniformly linearly independent if there exist a constant $\gamma > 0$ and indices $m \geqq n$, $k_0 \geqq 0$ such that for each $k \geqq k_0$,*

$$\max\{|x^T p^{(j)}|/\|x\|\,\|p^{(j)}\| \,|\, j = k+1, \ldots, k+m\} \geqq \gamma\,, \quad \forall x \in R_n, x \neq 0\,. \tag{2.4}$$

Let $w_0^{(1)}, w_1^{(2)}, \ldots, w_{n-1}^{(n)}$ be vectors defined by the algorithm (2.1). In each cycle we obtain a new set of these vectors. Let us denote

$$p^{(s+1)} = q^{(s)}\,, \quad q^{(s)} = w_{s(\mathrm{mod}n)}^{(s)(\mathrm{mod}n)+1}\,, \quad s = 0, 1, \ldots. \tag{2.5}$$

Let $k_0 \geqq 0$. The sequence $\{p^{(k)}\}_{k=1}^{\infty} \in R_n$, $p^{(k)} \neq 0$ defined by (2.5) has the property that for $k \geqq k_0$, $p^{(j)}$, $j = k+1, \ldots, k+n$ are linearly independent vectors and after a rearrangement the matrix of these vectors is upper triangular with $\lambda/\|v_{i-1}^{(i)}\|$ on the diagonal. Since the vector $v_{i-1}^{(i)}$ is a linear combination of the vectors $v_0^{(k)}$, $k = 1, 2, \ldots, i$ [13] and $L_0 = L(f(x_0^{(0)}))$ is compact according to (2.3), it follows that $\|v_{i-1}^{(i)}\| \leqq K\lambda$ and $\lambda = \|v_0^{(k)}\|$, $K > 0$, so that $\lambda/\|v_{i-1}^{(i)}\| \geqq 1/K = c > 0$.

Theorem 2. *The sequence $\{p^{(k)}\}_{k=1}^{\infty}$, $p^{(k)} \neq 0$ defined by (2.5) is uniformly linearly independent.*

Proof. Let $m = n$ and $k_0 = 0$. Since the vectors $p^{(j)}$, $j = k + 1, \ldots, k + n$ for each $k \geqq 0$ are linearly independent vectors and after a rearrangement the matrix of these vectors is upper triangular, they span R_n. Let us denote

$$\|x\|_k = \max \{|x^T p^{(j)}| / \|p^{(j)}\| \mid j = k + 1, \ldots, k + n\}$$

for $k \geqq 0$. $\|x\|_k$ defines a norm on R_n and there exists a constant $\gamma_k > 0$ such that $\|x\|_k \geqq \gamma_k \|x\|$ for all $x \in R_n$. Let us denote $L_0 = L(f(x_0^{(0)}))$. According to the assumption (2.3) the level set L_0 is compact. All points $x_i^{(i)}$ defined by the algorithm (2.1) which define the vectors $p^{(k)}$, lie in L_0. Let us denote

$$\gamma = \inf \gamma_k , \quad k \geqq 0 .$$

Then we have

$$\|x\|_k \geqq \gamma_k \|x\| \geqq \gamma \|x\| , \quad \gamma > 0$$

and (2.4) holds.

Proof of Theorem 1. The condition (2.3) ensures that $x_n^{(n)}$ is well defined and $L_0 = L(f(x_0^{(0)}))$ is compact while (2.2) ensures that $x_n^{(n)}$ lies in L_0. By virtue of Theorem 14. 2. 10 [1] and Theorem 14. 1. 3 [1] and with regard to the fact that the sequence $\{p^{(k)}\}$ is uniformly linearly independent Theorem 14.6.4 [1] applies and Theorem 14.1.5 [1] shows the convergence of $\{x_n^{(n)}\}$.

3. SPARSE SYSTEMS

In most mathematical representations of real problems the relations of the system contain only a few common variables. In the case of a system of linear algebraic equations the sparseness of the matrix influences the structure of the algorithms. According to the algorithm (1.4), the following statement can be proved [13]:

Let A be a strictly regular, q-diagonal band matrix. Let $v_0^{(k)} = x_0^{(k)} - x_0^{(0)}$ be in the form

$$v_0^{(k)} = (0, \ldots, t_k, \ldots, 0)^T \tag{3.1}$$

where $t_k = 1$. Then

$$x_i^{(k)} = x_i^{(i)} + v_0^{(k)} \quad k > \tfrac{1}{2}(q - 1) + i ; \quad i = 1, 2, \ldots, n ; \quad k = i, i + 1, \ldots, n . \tag{3.2}$$

Likewise we may derive reduced algorithms for other structures of sparse matrices such as block diagonal matrices, bordered block diagonal matrices, bordered band matrices. The sparseness of the matrix also influences the structure of the algorithm for the minimization of the quadratic function (1.6):

Let A be a positive definite, symmetric, q-diagonal band matrix. Then for the algorithm (1.7)–(1.8) it holds [14]

$$x_i^{(k)} = x_i^{(i)} + v_0^{(k)} \quad k > \tfrac{1}{2}(q - 1) + i ; \quad i = 1, 2, \ldots, n ; \quad k = i, i + 1, \ldots, n ,$$

where $v_0^{(k)}$ is in the form (3.1). For illustration, the number of function values of (1.6) required for minimization by the above described method and the method of conjugate gradients with exact difference formulas is as follows [14]:

full matrix	$n^2 + 2n$	$2n^2 + 3n$
5-diagonal band matrix	$7n - 6$	$2n^2 + 3n$
3-diagonal band matrix	$5n - 2$	$2n^2 + 3n$.

The structure of the matrix has no influence on the total number of function values in the case of gradient methods using difference formulas.

In order to obtain a more general result we first introduce a definition (see also in [4]). Let $f : R_n \to R_1$ be a continuously differentiable strictly convex function. Let us consider the system

$$\partial f/\partial x_i = f_i(x_1, x_2, \ldots, x_n) = 0 \quad i = 1, 2, \ldots, n . \tag{3.3}$$

Let us denote $f'(x) = (\partial f/\partial x_1, \partial f/\partial x_2, \ldots, \partial f/\partial x_n)$. The system (3.3) will be assigned a Boolean matrix, called the occurrence matrix:

An element of the matrix, s_{ij}, is either a Boolean 1 or 0 according to the rule

$$s_{ij} = \begin{cases} 1 \text{ if the } j\text{-th variable appears in the } i\text{-th equation} \\ 0 \text{ otherwise.} \end{cases}$$

This matrix then indicates the occurrence of the dependent variables in each of the relations and, which is equivalent, the functional dependence between the corresponding variables. The occurrence matrix corresponding to a function $f(x)$ is known either a priori or can be found.

Theorem 3. *Let $f : R_n \to R_1$ be a continuously differentiable strictly convex function satisfying*

$$\lim_{\|x\| \to \infty} f(x) = +\infty . \tag{3.4}$$

Let the occurrence matrix corresponding to the function $f(x)$ be a q-diagonal band matrix. Let $x_0^{(0)} \in R_n$ be an arbitrary starting point and let $v_0^{(k)} = x_0^{(k)} - x_0^{(0)}$ be in the form

$$v_0^{(k)} = (0, \ldots, t_k, \ldots, 0)^T$$

where $t_k = \lambda$. Then for $x_i^{(k)}$ defined by the algorithm (2.1) it holds

$$x_i^{(k)} = x_i^{(i)} + v_0^{(k)} \quad k > \tfrac{1}{2}(q - 1) + i ; \quad i = 1, 2, \ldots, n ; \quad k = i, i + 1, \ldots, n .$$

Proof. For $i = 1$ according to the algorithm (2.1) we have

$$x_1^{(1)} = x_0^{(1)} + \alpha_0^{(1)} w_0^{(1)} = x_0^{(0)} + \alpha_0^{(0)} w_0^{(1)}$$

$$x_1^{(k)} = x_0^{(k)} + \alpha_0^{(k)} w_0^{(1)} \quad k = 2, 3, \ldots, n .$$

For $x_1^{(k)}$ we obtain

$$x_1^{(k)} = x_0^{(k)} + \alpha_0^{(k)} w_0^{(1)} = x_0^{(0)} + \alpha_0^{(k)} w_0^{(1)} + v_0^{(k)} \quad k = 2, 3, \ldots, n \,. \tag{3.5}$$

For $x_1^{(k)}$ it holds that

$$\left(f'(x_1^{(k)}), w_0^{(1)}\right) = 0 \quad k = 1, 2, \ldots, n \,,$$

i.e., in view of the form of $w_0^{(1)}$ and $f'(x_1^{(k)})$ we have

$$f_1(x_{11}^{(k)}, x_{12}^{(k)}, \ldots, x_{1p}^{(k)}) = 0 \quad k = 1, 2, \ldots, n \,. \tag{3.6}$$

whereby $p = \frac{1}{2}(q-1) + 1$. Let $\alpha_0^{(0)} = \alpha_0^{(k)}$ for $k > p = \frac{1}{2}(q-1) + 1$. Then according to (3.5) the equation (3.6) will be fulfilled for $k < \frac{1}{2}(q-1) + 1$. The assumption (3.4) ensures that there exists only one such $\alpha_0^{(k)}$ for which the equation (3.6) is fulfilled for given k, hence for $\alpha_0^{(k)}$ defined by the algorithm (2.1) we have

$$\alpha_0^{(0)} = \alpha_0^{(k)} \quad k > \tfrac{1}{2}(q-1) + 1$$

and therefore

$$x_1^{(k)} = x_0^{(k)} + \alpha_0^{(k)} w_0^{(1)} = x_0^{(0)} + \alpha_0^{(0)} w_0^{(1)} + v_0^{(k)} = x_1^{(1)} + v_0^{(k)} \quad k > \tfrac{1}{2}(q-1) + 1 \,. \tag{3.7}$$

For $i = 2$ according to the algorithm (2.1) we obtain

$$x_2^{(2)} = x_1^{(2)} + \alpha_1^{(2)} w_1^{(2)} = x_1^{(1)} + \alpha_1^{(1)} w_1^{(2)}$$

$$x_2^{(k)} = x_1^{(k)} + \alpha_1^{(2)} w_1^{(2)} \quad k = 3, 4, \ldots, n \,. \tag{3.8}$$

By means of (3.7) we have

$$x_2^{(k)} = x_1^{(k)} + \alpha_1^{(k)} w_1^{(2)} = x_1^{(1)} + \alpha_1^{(k)} w_1^{(2)} + v_0^{(k)} \quad k > \tfrac{1}{2}(q-1) + 2 \,. \tag{3.9}$$

For $x_2^{(k)}$ it holds according to (2.2) that

$$\left(f'(x_2^{(k)}), w_1^{(2)}\right) = 0 \quad k = 2, 3, \ldots, n$$

where

$$w_1^{(2)} = (w_{11}^{(2)}, w_{12}^{(2)}, 0, \ldots, 0)^T \,,$$

i.e., in view of the form of $w_1^{(2)}$ and $f'(x_2^{(k)})$,

$$f_1(x_{21}^{(k)}, x_{22}^{(k)}, \ldots, x_{2p-1}^{(k)})\, w_{11}^{(2)} + f_2(x_{21}^{(k)}, x_{22}^{(k)}, \ldots, x_{2p}^{(k)})\, w_{12}^{(2)} = 0 \tag{3.10}$$

$$k = 2, 3, \ldots, n \,,$$

where $p = \frac{1}{2}(q-1)+2$. Let $\alpha_1^{(1)} = \alpha_1^{(k)}$ for $k > \frac{1}{2}(q-1)+2$. Then according to (3.9) the equation (3.10) will be fulfilled for $k > \frac{1}{2}(q-1)+2$. Since the assumption (3.4) ensures that there exists only one such $\alpha_1^{(k)}$ for given k that the equation (3.10) is fulfilled, hence for $\alpha_1^{(k)}$ defined by the algorithm (2.1) we have

$$\alpha_1^{(1)} = \alpha_1^{(k)} \quad k > \tfrac{1}{2}(q-1)+2$$

and therefore

$$x_2^{(k)} = x_1^{(k)} + \alpha_1^{(k)} w_1^{(2)} = x_1^{(1)} + \alpha_1^{(1)} w_1^{(2)} + v_0^{(k)} = x_2^{(2)} + v_0^{(k)} \quad k > \tfrac{1}{2}(q-1)+2\,.$$

Let

$$x_j^{(k)} = x_j^{(j)} + v_0^{(k)} \quad k > \tfrac{1}{2}(q-1)+j\,. \tag{3.11}$$

For $i = j+1$ we have

$$x_{j+1}^{(j+1)} = x_j^{(j+1)} + \alpha_j^{(j+1)} w_j^{(j+1)} = x_j^{(j)} + \alpha_j^{(j)} w_j^{(j+1)}$$

$$x_{j+1}^{(k)} = x_j^{(k)} + \alpha_j^{(k)} w_j^{(j+1)} \quad k = j+2, \ldots, n\,.$$

According to (3.11) we have

$$x_{j+1}^{(k)} = x_j^{(k)} + \alpha_j^{(k)} w_j^{(j+1)} = x_j^{(j)} + \alpha_j^{(k)} w_j^{(j+1)} + v_0^{(k)} \quad k > \tfrac{1}{2}(q-1)+j+1\,. \tag{3.12}$$

For $x_{j+1}^{(k)}$ it holds that

$$\left(f'(x_{j+1}^{(k)}), w_j^{(j+1)}\right) = 0 \quad k = j+1, \ldots, n$$

whereby

$$w_j^{(j+1)} = \left(w_{j1}^{(j+1)}, \ldots, w_{jj+1}^{(j+1)}, 0, \ldots, 0\right)^T,$$

.e , in view of the form of $f'(x_{j+1}^{(k)})$ and $w_j^{(j+1)}$,

$$\sum_{i=1}^{j+1} f_i(x_{j+1}^{(k)})\, w_{ji}^{(j+1)} = 0 \quad k = j+1, \ldots, n \tag{3.13}$$

and $p = \frac{1}{2}(q-1)+j+1$ denotes the last component $x_{j+1\,p}^{(k)}$ of $x_{j+1}^{(k)}$ in $f_{j+1}(x_{j+1}^{(k)})$. Let $\alpha_j^{(j)} = \alpha_j^{(k)}$ for $k > (q-1)/2 + j + 1$. Then according to (3.12) the equation (3.13) will be fulfilled for $k > \frac{1}{2}(q-1)+j+1$. Therefore according to the assumption (3.4), for $\alpha_j^{(k)}$ defined by the algorithm (2.1) it holds

$$\alpha_j^{(k)} = \alpha_j^{(j)} \quad k > \tfrac{1}{2}(q-1)+j+1$$

and for $x_{j+1}^{(k)}$ we have

$$x_{j+1}^{(k)} = x_j^{(k)} + \alpha_j^{(k)} w_j^{(j+1)} = x_j^{(i)} + \alpha_j^{(j)} w_j^{(j+1)} + v_0^{(k)} = x_{j+1}^{(j+1)} + v_0^{(k)}$$

$$k > \tfrac{1}{2}(q-1)+j+1\,.$$

Since Theorem 3 holds for an arbitrary starting point $x_0^{(0)}$ it can be applied to each iteration defined by the algorithm (2.1). For $q = 1$ we obtain the nonlinear Gauss-Seidel iteration. It means that the nonlinear Gauss-Seidel method is efficient if the occurrence matrix is diagonal. Theorem 3 enables us to reduce the computational time required for the minimization of $f(x)$.

4. COMPUTATIONAL CONSIDERATIONS

In this part we will consider an alternative implementation of the algorithm (2.1).

(4.1) Algorithm.

Step(0): Define λ and set $x_0^{(k)} = x_0^{(0)} + v_0^{(k)}$ for $k = 1, 2, \ldots, n$; set $i = 1$.

Step (1): Compute

$$x_i^{(i)} = x_{i-1}^{(i-1)} + \alpha_{i-1}^{(i-1)} w_{i-1}^{(i)}$$

where

$$w_{i-1}^{(i)} = v_{i-1}^{(i)} / \|v_{i-1}^{(i)}\|, \quad v_{i-1}^{(i)} = x_{i-1}^{(i)} - x_{i-1}^{(i-1)}$$

and $\alpha_{i-1}^{(i-1)}$ is a scalar coefficient such that

$$f(x_{i-1}^{(i-1)} + \alpha w_{i-1}^{(i)}) = \min!$$

$$\alpha = \alpha_{i-1}^{(i-1)}$$

Step (2): Compute

$$x_{i-1}^{(k)} = x_{i-1}^{(k)} - (x_{i-1}^{(k)} - x_i^{(i)}, w_{i-1}^{(i)})\, w_{i-1}^{(i)} \quad k = i + 1, \ldots, n .$$

Step (3): Compute

$$x_i^{(k)} = x_{i-1}^{(k)} + \alpha_{i-1}^{(k)} w_{i-1}^{(i)}$$

where $\alpha_{i-1}^{(k)}$ is a scalar coefficient such that

$$f(x_{i-1}^{(k)} + \alpha w_{i-1}^{(i)}) = \min!$$

$$\alpha = \alpha_{i-1}^{(k)} \quad k = i + 1, \ldots, n .$$

Step (4): Set $i = i + 1$. If $i \leqq n$ then go to Step (1); else, go to Step (5).

Step (5): If $\|x_n^{(n)} - x_0^{(0)}\| = 0 \vee |f(x_n^{(n)}) - f(x_0^{(0)})| = 0 \vee \|f'(x_n^{(n)})\| = 0$ then stop; else, $x_0^{(0)} = x_n^{(n)}$ and go to Step (0).

In Step (2) we define the orthogonal projections of the point $x_i^{(i)}$, defined in Step (1), on to the cotresponding parallel directions using the fact that the vectors $w_{i-1}^{(i)}$ are normalized. In the proof of Theorem 3 we have seen that the vectors $v_i^{(k)} = x_i^{(k)} - x_i^{(i)}$ for $k > \frac{1}{2}(q-1) + i$ are orthogonal to $w_{i-1}^{(i)}$. These vectors are nearly orthogonal in the case of well-conditioned problems and in the neighbourhood of the global minimizer. By a suitable choice of λ the orthogonal projections, defined in Step (2),

are good initial approximations of the local minimizers on the corresponding directions. In Step (3) we consider the linear minimizations on the corresponding parallel directions using the result of Step (2). Step (2) may be considered a predictor and Step (3) a corrector of the local minimizers.

The above described algorithm terminates theoretically after a finite number of linear minimizations for functions for which the equiconstant level sets are concentric ellipsoids, e.g., $g = \exp[f(x)]$ where $f(x)$ is a quadratic function. In this case the algorithm requires in general $\frac{1}{2}n\,(n+1)$ linear minimizations and $2n-1$ linear minimizations if the occurrence matrix is 3-diagonal band matrix. The other methods of this type [7], [8], [3], [6] require n^2 linear minimizations per iteration and the structure of the occurrence matrix has no influence on the number of linear minimizations. In numerical examples we have calculated local minimizers with an accuracy of 10^{-6}. If λ is a constant it must not be chosen too small. The main advantage of the algorithm is that the sparseness of the occurrence matrix enables us to define effective reduced algorithms which reduce substantially the number of linear minimizations and require small storage facilities.

5. NONLINEAR SYSTEMS

Let $F : R_n \to R_n$ be continuously differentiable with $F'(x)$ symmetric and satisfying for some $c > 0$

$$h^T F'(x)\, h \geqq ch^T h \quad \forall x,\ h \in R_n\ ;$$

then the function $g : R_n \to R_1$,

$$g(x) = \int_0^1 x^T F(tx)\, \mathrm{d}t$$

according to 4.1.6 [1], 4.3.6 [1] and 3.4.6 [1] is uniformly convex, $g'(x) = (Fx)^T$ and $\lim g(x) = +\infty$ for $\|x\| \to \infty$, so that the unique minimizer of $g(x)$ is the unique solution of $Fx = 0$.

6. PARALLEL COMPUTATION AND STORAGE REQUIREMENTS

The above described algorithms are suitable for implementation on a parallel computer. The minimizations on parallel directions are independent from the computational point of view and can be calculated simultaneously. If we assume that the minimizations on parallel directions require the same computational time and that they are calculated simultaneously then one step of the above described algorithms on a parallel computer is equivalent to one step of the generalized Gauss-Seidel method and the rate of convergence is substantially better. The sparseness of the occurrence matrix influences the number of processors. Each processors of a parallel computer has to store one vector. The total storage requirements for a sequential

computer are less than $\frac{1}{4}n^2 + 2n + 2$. If the occurrence matrix is a q-diagonal band matrix then it is necessary to store $\frac{1}{2}(q-1) + 1$ vectors. In [5], [9], [10], [11] other parallel iterative algorithms are described.

References

[1] *J. M. Ortega, W. C. Rheinboldt:* Iterative solution of nonlinear equations in several variables. AP, New York, 1970.

[2] *M. J. D. Powell:* A survey of numerical methods for unconstrained optimization. SIAM Review, 1 (1970), 79—97.

[3] *D. Chazan, W. Miranker:* A nongradient and parallel algorithm for unconstrained minimization. SIAM J. Control, 2 (1970), 207—217.

[4] *D. M. Himmelblau:* Decomposition of large-scale problems. North-Holl. publ. comp., New York, 1973.

[5] *R. M. Karp, W. L. Miranker*: Parallel minimax search for a maximum. J. of Combinatioral Theory, 1 (1968), 19—35.

[6] *R. P. Brent:* Algorithms for minimization without derivatives. Prentice-Hall, Englewood Cliffs, New Jersey, 1973.

[7] *W. I. Zangwill:* Minimizing a function without calculating derivatives. Comp. J., 7 (1967), 293—296.

[8] *M. J. D. Powell:* An efficient method for finding minimum of a function of several variables without calculating derivatives. Compt. J., 7 (1964), 155—162.

[9] *H. T. Kung, J. F. Traub:* On the efficiency of parallel iterative algorithms for non-linear equations. Symposium on complexity of sequential and parallel numerical algorithms, Cornegie-Mellon University, 1973.

[10] *W. Miranker:* Parallel methods for approximating the root of a function. IBM J. of Research and Development, vol. 13, 1967, 297—301.

[11] *S. Winograd:* Parallel iteration methods, Complexity of computer computations. R. E. Miller and J. W. Thatcher, Plenum Press, New York, 1972, 53—60.

[12] *N. Anderson, A. Brörck:* A new high order method of regula falsi type for computing a root of an equation. BIT, 13 (1973), 253—264.

[13] F. Sloboda: A parallel projection method for linear algebraic systems, to appear.

[14] F. Sloboda: Parallel method of conjugate directions for minimization. Apl. mat. 6 (1975), 436—446.

Súhrn

NELINEÁRNE ITERAČNÉ METÓDY A PARALELNÝ VÝPOČET

FRIDRICH SLOBODA

V článku sú skúmané nelineárne iteračné metódy pre minimalizáciu spojite diferencovateľných ostro konvexných funkcií, je popísané zobecnenie jednej priamej projekčnej metódy pre riešenie systému lineárnych algebraických rovníc. Metóda je vhodná pre paralelný výpočet a pre riedke matice výskytu.

Author's address: Dr. *Fridrich Sloboda,* Ústav technickej kybernetiky SAV, Dúbravská cesta 1. 809 31 Bratislava.