Computers Math. Applic. Vol. 19, No. 12, pp. 65–69, 1990
Printed in Great Britain.

0097-4943/90 $3.00 + 0.00

A METHOD OF INEXACT STEEPEST DESCENT FOR SYSTEMS OF LINEAR EQUATIONS

T. ALTMAN
Department of Computer Science, University of Kentucky, Lexington, KY 40506, U.S.A.

(*Received* 30 *May* 1989)

Abstract—Our approach combines the method of inexact steepest descent with the method of *contractor directions* to obtain an algorithm for solving systems of linear equations. In order to enhance the scope of applicability, we consider an iterative method with variable step-size iterations. We prove the convergence and given an error estimate for our method.

The algorithm is well-suited for parallel computation. In fact, for systems with m equations and n unknowns, each iteration may be computed in parallel time $O(\log m + \log n)$, on an EREW PRAM with $O(mn)$ processors.

1. INTRODUCTION

The well-known method of steepest descent can also be applied to solve systems of linear algebraic equations. However, in practice, it is usually impossible to carry out the steepest descent computations exactly, e.g. due to the floating point arithmetic error. For this reason, we introduce a method of inexact steepest descent, or a gradient type method, with a variable step-size.

Our method incorporates Altman's method of *contractor directions* [1, 2] together with some of the ideas from Ref. [3]. It is an iterative method with backtracking whose individual iteration steps can be computed with a high degree of parallelism. Moreover, the method of contractor directions allows for an additional level of parallel computation in that the computations for backtracking may be performed simultaneously. The method may be applied to systems of equations where m, the number of equations, is not equal to n, the number of unknowns.

2. DESCRIPTION OF THE METHOD

Let A be an $(m \times n)$-matrix, $\mathbf{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$, $\mathbf{b} = (b_1, b_2, \ldots, b_m)$. Our system of linear equations is

$$A\mathbf{x} = \mathbf{b}. \tag{1}$$

Instead of equation (1), we solve the equation

$$F(\mathbf{x}) = \| A\mathbf{x} - \mathbf{b} \|^2 = 0,$$

where $F(\mathbf{x})$ is a nonlinear functional, and denote by $F'(\mathbf{x})$ its gradient at $\mathbf{x}$. Thus, we have

$$\langle F'(\mathbf{x}), h \rangle = 2\langle A^{\mathrm{T}}(A\mathbf{x} - \mathbf{b}), h \rangle. \tag{2}$$

The following is the iterative method under consideration: let $0 < \bar{q} < q < 1$ be given constants. We put

$$\mathbf{x}_{i+1} := \mathbf{x}_i - \epsilon_i h_i. \tag{3}$$

We choose h to satisfy the following relation:

$$|\langle F'(\mathbf{x}), h \rangle - F(\mathbf{x})| \leqslant \bar{q} F(\mathbf{x}) \tag{4}$$

or

$$|2\langle A^{\mathrm{T}}(A\mathbf{x} - \mathbf{b}), h \rangle - \| A\mathbf{x} - \mathbf{b} \|^2| \leqslant \bar{q} \| A\mathbf{x} - \mathbf{b} \|^2 \tag{5}$$

and

$$\|h\| \leqslant c\sqrt{F(\mathbf{x})} = c\,\|A\mathbf{x}-\mathbf{b}\|, \tag{6}$$

for some constant $c > 0$.

If

$$h = \frac{\|A\mathbf{x}-\mathbf{b}\|^2}{2|A^{\mathrm{T}}(A\mathbf{x}-\mathbf{b})\|^2} A^{\mathrm{T}}(A\mathbf{x}-\mathbf{b})$$

and $\epsilon_i = 1$, we obtain

$$2\langle A^{\mathrm{T}}(A\mathbf{x}-\mathbf{b}), h\rangle = \|A\mathbf{x}-\mathbf{b}\|^2$$

and method (3) is then

$$\mathbf{x}_{i+1} := \mathbf{x}_i - \frac{\|A\mathbf{x}_i-\mathbf{b}\|^2}{2\|A^{\mathrm{T}}(A\mathbf{x}_i-\mathbf{b})\|^2} A^{\mathrm{T}}(A\mathbf{x}_i-\mathbf{b}), \tag{7}$$

that is, we obtain the method of steepest descent, which is a special case of method (3), and condition (6) is satisfied if, e.g. A is a nonsingular square matrix. Then we get

$$\|h\| \leqslant \frac{1}{2}\frac{\|A\mathbf{x}-\mathbf{b}\|^2}{\|A^{\mathrm{T}}(A\mathbf{x}-\mathbf{b})\|} \leqslant \frac{1}{2c_1}\frac{\|A\mathbf{x}-\mathbf{b}\|^2}{\|A\mathbf{x}-\mathbf{b}\|} = c\,\|A\mathbf{x}-\mathbf{b}\|,$$

where $c = (2c_1)^{-1}$ in expression (6) provided $\|A^{\mathrm{T}}(A\mathbf{x}-\mathbf{b})\| \geqslant c_1\|A\mathbf{x}-\mathbf{b}\|$.

In order to determine the step-size ϵ_i, let us put

$$\Phi(\epsilon, h, \mathbf{x}) = \frac{|F(\mathbf{x}-\epsilon h) - (1-\epsilon)F(\mathbf{x})|}{\epsilon}. \tag{8}$$

If

$$\Phi(1, h_i, x_i) \leqslant qF(x_i),$$

then put

$$\epsilon_i := i. \tag{9}$$

If $\epsilon_i = 1$ does not satisfy condition (9), put

$$\epsilon_i := \frac{\epsilon_i}{2},$$

and keep dividing it by 2 until

$$\beta q F(\mathbf{x}_i) \leqslant \Phi(\epsilon_i, h_i, \mathbf{x}_i) \leqslant qF(\mathbf{x}_i), \tag{10}$$

where $0 < \bar{q}/q < \beta < 1$ is fixed independently of i.

In this way, the following *fundamental inequality for contractor directions*, (see Ref. [2]), is satisfied for each iteration:

$$|F(\mathbf{x}_i - \epsilon_i h_i) - (1-\epsilon_i)F(\mathbf{x}_i)| \leqslant \epsilon_i q F(\mathbf{x}_i). \tag{11}$$

Let us also define the following sequence:

$$t_0 = 0, \quad t_{i+1} = t_i + \epsilon_i, \quad 0 < \epsilon_i \leqslant 1.$$

It follows inductively from condition (11) that

$$F(\mathbf{x}_i) \leqslant F(\mathbf{x}_0)\exp(-(1-q)t_i). \tag{12}$$

In fact, by induction from (11) we get

$$F(\mathbf{x}_{i+1}) \leqslant (1-(1-q)\epsilon_i)F(\mathbf{x}_i) \leqslant \exp(-(1-q)\epsilon_i)F(\mathbf{x}_i) \leqslant F(\mathbf{x}_0)\exp(-(1-q)(t_i+\epsilon_i)).$$

Lemma 2.1

The following estimate holds

$$\sum_{n=0}^{\infty} \epsilon_i \| h_i \| \leqslant c[\tfrac{1}{2}(1-q)]^{-1}\sqrt{F(\mathbf{x}_0)}\exp(\tfrac{1}{2}(1-q)). \tag{13}$$

Proof. From conditions (12) and (6) we have

$$\| h_i \| \leqslant c\sqrt{F(\mathbf{x}_0)}\exp(-\tfrac{1}{2}(1-q)t_i). \tag{14}$$

Furthermore,

$$\begin{aligned}\sum_{i=0}^{\infty} \epsilon_i \exp(-\tfrac{1}{2}(1-q)t_i) &= \sum_{i=0}^{\infty} (t_{i+1}-t_i)\exp(-\tfrac{1}{2}(1-q)t_i)\\ &= \sum_{i=0}^{\infty} (t_{i+1}-t_i)\exp(-\tfrac{1}{2}(1-q)t_{i+1})\exp(\tfrac{1}{2}(1-q)\epsilon_i)\\ &\leqslant \exp(\tfrac{1}{2}(1-q)) \sum_{i=0}^{\infty} \int_{t_i}^{t_{i+1}} \exp(-\tfrac{1}{2}(1-q)\theta)\,\mathrm{d}\theta\\ &= \exp(\tfrac{1}{2}(1-q)) \int_0^{\infty} \exp(-\tfrac{1}{2}(1-q)\theta)\,\mathrm{d}\theta\\ &= [\tfrac{1}{2}(1-q)]^{-1}\exp(\tfrac{1}{2}(1-q)).\end{aligned}$$

Hence, relation (13) results from relation (14). □

It results from Lemma 2.1 that $\mathbf{x} = \mathbf{x}_0 + \epsilon_0 h_0 + \epsilon_1 h_1 + \cdots$ is convergent.

Lemma 2.2

Suppose that $\{\mathbf{x}_i\}$ is convergent to some $\mathbf{x}$ and $\{h_i\}$ is bounded. Then

$$\epsilon_i^{-1}|F(\mathbf{x}_i - \epsilon_i h_i) - F(\mathbf{x}_i) + \epsilon_i\langle F'(\mathbf{x}_i), h_i\rangle| \to 0 \quad \text{as} \quad \epsilon_i \to 0. \tag{15}$$

Proof. We have, by equation (2)

$$\begin{aligned}\epsilon_i^{-1}[F(\mathbf{x}_i - \epsilon_i h_i) - F(\mathbf{x}_i) + \langle F'(\mathbf{x}_i), \epsilon_i h_i\rangle) &= \epsilon_i^{-1}\int_0^1 \langle F'(\mathbf{x}_i - \theta\epsilon_i h_i) - F'(\mathbf{x}_i), -\epsilon_i h_i\rangle\,\mathrm{d}\theta\\ &\leqslant \| h_i \| \int_0^1 \| F'(\mathbf{x}_i - \theta\epsilon_i h_i) - F'(\mathbf{x}_i)\|\,\mathrm{d}\theta.\end{aligned}$$

But,

$$\| F'(\mathbf{x}_i - \theta\epsilon_i h_i) - F'(\mathbf{x}_i)\| = \| 2A^{\mathrm{T}}[A(\mathbf{x}_i - \epsilon_i\theta h_i) - \mathbf{b}] - 2A^{\mathrm{T}}(A\mathbf{x}_i - \mathbf{b})\| = 2\epsilon_i\theta\| A^{\mathrm{T}}Ah_i\| \to 0,$$

as $\epsilon_i \to 0$, uniformly in $0 \leqslant \theta \leqslant 1$. Hence, relation (15) follows. □

Lemma 2.3

Under the assumption of Lemma 2.2, we get for $i \to \infty$,

$$(\beta q - \bar{q})F(\mathbf{x}_i) \leqslant \Phi(\epsilon_i, h_i, \mathbf{x}_i) \to 0, \quad \text{if } \epsilon_i \to 0. \tag{16}$$

Proof. We have

$$\Phi(\epsilon_i, h_i, \mathbf{x}_i) \leqslant \epsilon_i^{-1}|[F(\mathbf{x}_i - \epsilon_i h_i) - F(\mathbf{x}_i) + \langle F'(\mathbf{x}_i), \epsilon_i h_i\rangle] - \epsilon_i[\langle F'(\mathbf{x}_i), h_i\rangle - F(\mathbf{x}_i)]|.$$

Hence, by relation (4) we get

$$\Phi(\epsilon_i, h_i, \mathbf{x}_i) \leqslant \epsilon_i^{-1}|F(\mathbf{x}_i - \epsilon_i h_i) - F(\mathbf{x}_i) + \langle F'(\mathbf{x}_i), \epsilon_i h_i\rangle| + \bar{q}F(\mathbf{x}_i),$$

and relation (16) follows, by relation (10). It follows that if relation (15) is true for relation (1), then relation (3) converges to a solution. □

Theorem 2.1

Suppose that conditions (4) and (6) are satisfied. Then the equation $A\mathbf{x} = \mathbf{b}$ has a solution $\mathbf{x}$ such that

$$\|\mathbf{x}_i - \mathbf{x}\| \to 0, \quad \text{as} \quad i \to \infty, \tag{17}$$

where $\{\mathbf{x}_i\}$ is determined by relation (3). The error estimate is given by the formula

$$\begin{aligned}\|\mathbf{x}_i - \mathbf{x}\| &\leqslant c\,\|A\mathbf{x}_0 - \mathbf{b}\|\exp(-\tfrac{1}{2}(1-q))\int_{t_i}^{\infty}\exp(-\tfrac{1}{2}(1-q)\theta)\,\mathrm{d}\theta \\ &= c\,\|A\mathbf{x}_0 - \mathbf{b}\|\exp(-\tfrac{1}{2}(1-q))2(1-q)^{-1}\exp(-\tfrac{1}{2}(1-q))t_i). \end{aligned} \tag{18}$$

Proof. Relations (6) and (13) imply that the sequence $\{\mathbf{x}_i\}$ converges to some $\mathbf{x}$, that is, condition (17) holds. Since $t_0 = 0$ and $t_i = \Sigma_{j=0}^{i-1}\epsilon_j$, we consider two cases:

(a) $t_i \to \infty$ as $i \to \infty$. Then we get from relation (12) that $\|A\mathbf{x}_i - \mathbf{b}\| \to 0$ as $i \to \infty$. Hence, $A\mathbf{x} = \mathbf{b}$.

(b) Suppose that $\lim_{i\to\infty} t_i < \infty$. Then $\epsilon_i \to 0$ as $i \to \infty$. But it results from relation (16) that $F(\mathbf{x}_i) \to 0$ as $i \to \infty$, i.e. $\|A\mathbf{x}_i - \mathbf{b}\| \to 0$ as $i \to \infty$. Hence, $A\mathbf{x} = \mathbf{b}$. This completes the proof. □

3. PARALLEL COMPUTATION

Below, we show that an iteration step in our algorithm can be performed in parallel time $O(\log m + \log n)$ on an EREW PRAM with $O(mn)$ processors. Moreover, the determination of the step-size ϵ_i, can itself be determined by evaluating the Φ function (with different ϵ step-sizes) simultaneously, thereby eliminating the sequential nature of condition (10). In Ref. [3], we have shown that for a given iteration, the number of different ϵ values is bounded, and could be determined *a priori*, although it would not be cost effective to do so.

Let us examine an iteration step of our method:

$$\mathbf{x}_{i+1} := \mathbf{x}_i - \epsilon_i \frac{\|A\mathbf{x}_i - \mathbf{b}\|^2}{2\|A^{\mathrm{T}}(A\mathbf{x}_i - \mathbf{b})\|^2} A^{\mathrm{T}}(A\mathbf{x}_i - \mathbf{b}).$$

The product $A\mathbf{x}_i$ will take $O(\log n)$ steps, whereas the computation of the norms

$$\|A\mathbf{x}_i - \mathbf{b}\|^2 \quad \text{and} \quad \|A^{\mathrm{T}}(A\mathbf{x}_i - \mathbf{b})\|$$

will take additional $O(\log m + \log n)$ time.

It follows that evaluation of function Φ in the determination of ϵ_i, also takes $O(\log m + \log n)$ steps. Hence, on a machine with $O(mn)$ processors, the computation time for each iteration is

$$O(\log m + \log n).$$

4. CONCLUDING REMARKS

We have presented a method for solving systems of linear algebraic equations. The method may be applied to systems of equations where m, the number of equations, is not equal to n, the number of unknowns. Our approach combines the method of inexact steepest descent with the method of *contractor directions* to obtain an iterative backtracking algorithm that is also well-suited for parallel computation, as was demonstrated in Refs [3, 4]. Each iteration step may be computed in parallel time $O(\log m + \log n)$. Moreover, the line search determination allows for an additional level of parallelism, since the computation for the backtracking steps may be performed simultaneously.

REFERENCES

1. M. Altman, *Contractors and Contractor Directions Theory and Applications: A New Approach to Solving Equations, Lecture Notes in Pure and Applied Mathematics*. Dekker, New York (1977).
2. M. Altman, Iterative methods of contractor directions. *Nonlinear Analysis* **4**, 761–772 (1980).
3. T. Altman, Application of contractor directions for solving systems of nonlinear equations. *Computers Math. Applic.* **19**(6), 9–14 (1990).
4. S. Rajendran, Experiences with a parallel algorithm for solving systems of nonlinear equations. M.S. Thesis, Dept. of Computer Science, University of Kentucky, Lexington, Ky (1989).