PUSAN KYŎNGNAM MATHEMATICAL JOURNAL 69
Vol. 2, 69—74, 1986

ON THE LINE SEARCH METHOD USING MUTUALLY ORTHONORMAL DIRECTIONS

CHANG WOOK KIM

1. Introduction

Unconstrained optimization deals with the problem of minimizing a function in the absence of any constraints.

The well-known techniques solving unconstrained minimization problems are the simplex method ([2]), the pattern search method ([4]), the gradient method ([8]), the Newton's method ([1]) and the conjugate gradient method ([3]). Several unconstrained optimization techniques can be extended in a natural way to provide and motivate solution procedures for constrained problems ([6], [7]).

In this paper, we are concerned with the line search method using mutually orthonormal directions and the cyclic coordinate method. We give algorithms about two techniques and establish the covergence theorem for the more general algorithms which cover two algorithms.

2. The line search method using mutually orthonormal directions

This method uses the mutually orthonormal vectors in R^n as the search directions. At k-th iteration, suppose that we have a point x^k and a set of mutually orthonormal directions $d_1^k, \cdots, d_n^k$. We obtain x^{k+1} by minimizing f sequentially along the directions $d_1^k, \cdots, d_n^k$. Suppose that $x^{k+1} = x^k + \sum_{i=1}^{n}$

$\lambda_j^k d_j^k$. We generate the new collection of mutually orthonormal directions $d_1^{k+1}, \cdots, d_n^{k+1}$ in R^n by the Gram-Schmidt orthogonalization process ([5]) as follows;

$$a_j^{k+1}=\begin{cases} d_j^k & , \text{ if } \lambda_j^k=0 \\ \sum_{j=i}^{n} \lambda_j^k d^k, & \text{if } \lambda_j^k=0 \end{cases}$$

$$b_j^{k+1}=\begin{cases} a_j^k & , j=1 \\ a_j^k-\sum_{i=1}^{j-1} \{(a_{jk})^t d^{k+1}\} d_i^{k+1}, & j\geq 2 \end{cases} \tag{1-1}$$

$$d_j^{k+1}=\frac{b_j^{k+1}}{||b_j^{k+1}||}.$$

We summarize below the line search method using orthonormal directions for minimizing a function $f: R^n \to R$ as follows.

Initialization step: Let $\varepsilon>0$ be the termination scalar, Choose $d_1^1, \cdots, d_n^1$ as the coordinate directions, where d_j^1 is a vector of zeros except for a one at the j-th position. Choose a starting point x^1, let $y_1^1=x^1$, $k=j=1$ and go to the main step.

Main step: 1. Let λ_j^k be an optimal solution to the problem to minimize $f(y_j^k+\lambda d_j^k)$ subject to $\lambda \in R$, and let $y^k{}_{j+1}=y_j^k+\lambda_j^k d_j^k$. If $j<n$, replace j by $j+1$, and repeat step 1. Otherwise, go to step 2.

2. Let $x^{k+1}=y^k{}_{n+1}$ If $||x^{k+1}-x^k||<\varepsilon$, then step. Otherwise, let $y_1^{k+1}=X^{k+1}$, replace k by $k+1$ let $j=1$ and go to step 3.

3. Form a new set of mutually orthonormal search directions by (1-1). Repeat step 1.

3. The cyclic coordinate method

This method uses the coordinate axes as the search directions. More specially, the method searches cyclically along the directions $d_1, \cdots, d_n$ in R^n, where d_j is a vector of zeros except for a one at the j-th position. Thus, along the search direction d_j, the j-th variable x_j is changed while all other variables are kept fixed.

We summarize below the cyclic coordinate method for minimizing a function $f: R^n \to R$ as follows.

Initialization step: Choose a scalar $\varepsilon > 0$ to be used for terminating the algarithm and let $d_1, \cdots, d_n$ be the coordinate directions. Choose an initial point x^1, let $y_1{}^1 = x^1$, let $k = j = 1$, and go to the main step.

Main step: 1. Let $\lambda_j{}^k$ be an optimal solution to the problem to minimize $f(y_j{}^k + \lambda d_j{}^k)$ subject to $\lambda \in R$, and let $Y^k{}_{j+1} = y_j{}^k + \lambda_j{}^k d_j{}^k$ If $j < n$, replace j by $j+1$, and repeat step 1. Otherwise, if $j = n$, go to step 2.

2. Let $x^{k+1} = y^k{}_{n+1}$. If $||x^{k+1} - x^k|| < \varepsilon$, then stop. Otherwise, let $y_1{}^{k+1} = x^{k+1}$, let $j = 1$, replace k by $k+1$, and repeat step 1.

4. The convergence theorem

We produce a more general algorithm which covers the line search method using mutually orthonormal directions and the cyclic coodinate method as follows: At *k-th* iteration, suppose that we have a point x^k and a set of directions $d_1{}^k$, $\cdots$, $d_n{}^k$ in R^n, where $||d_j{}^k|| = 1$ for each j and that $|\det D^k| \geq \delta$ for some $\delta > 0$, where $D = (d_1{}^k, \cdots, d_{nk})$

We determine a new point x^{k+1} and a new set of directions as follows minimizing a function $f: R^n \to R$ strating from x^k sequentially in the directions $d_1{}^k, \cdots, d_n{}^k$ produces $x^{k+1} = x^k + D^k \lambda^k$, where $\lambda^k = (\lambda_1{}^k, \cdots, \lambda_n{}^k)$ is a vector indicating the distance moved in each direction, that is,

$$y_1{}^k = x^k. \tag{3-1}$$

$$y^k{}_{i+1} = y_i{}^k + \lambda_i{}^k d_i{}^k, \quad i = 1, \cdots, n, \tag{3-2}$$

$$x^{k+1} = y^k_{n+1} \tag{3-3}$$

$$f(y_i^k + \lambda_i^k d_i^k) \leq f(y_i^k + \lambda d_i^k) \text{ for all } \lambda \in R, \quad \text{for } i = 1, \cdots n. \tag{3-4}$$

Now D^{k+1} obtained as a function of the triple $(\lambda^k, D, {}^k x^k)$ by some specified rule such that $||d_j^{k+1}|| = 1$ for each j and $|\det D^{k+1}| \geq \delta$.

The following Lemma is needed to prove the convergence theorem for the above scheme which covers the line search method using mutually orthonormal directions and the cyclic coordinate method.

LEMMA. *Let $f: R^n \to R$ be continuous and $x^1 \in R_n$. Let $\{d^k\}$ be a sequence of directions such that $||d^k|| = 1$ for each k. Starting from x^1, let $\{x^k\}$ be a sequence obtained by minimizing f from x^{k-1} in the direction d^k to obtain x^k. If $\{x^k\}$ is contained in a compact set X and if f has a unique minimum on any line, then $||x^{k+1} - x^k|| \to 0$ as $k \to \infty$.*

PROOF. For each $x \in X$ and d with $||d|| = 1$, let $s(x, d)$ be a number such that $f[x + s(x, d)d] \leq f(x + \lambda d)$ for all λ.

Then for $\varepsilon > 0$, let

$h(\varepsilon) = \inf\{f(x) - f[x + s(x, d)d] | s(x, d) \geq \varepsilon, \quad x \in X, \quad ||d|| = 1\}$

By contradiction, suppose that $h(\varepsilon) = 0$ for some $\varepsilon > 0$. By the definition of infimum and the compactness of X, we can find an infinite sequence K of positive integers so that for $k \in K$, $x^k \to \bar{x}$, $d^k \to \bar{d}$, $s^k = s(x^k, d^k) \to s \geq \varepsilon$ and such that $f(x^k) - f(x^k + s^k d^k) \to 0$. Since $f(x^k + s^k d^k) \leq f(x^k + \lambda d^k)$ for each λ, then $f(\bar{x}) = f(\bar{x} + sd) \leq f(\bar{x} + \lambda d)$ for each λ. This contradicts the assumption that f has a unique minimum on any line, Thus $h(\varepsilon) > 0$ for each $\varepsilon > 0$.

Since $\{x^k\}$ is contained in a compact set X and f is continuous, then the sequence $\{f(x^k)\}$is bounded below. Since the sequence $\{f(x^k)\}$ is decreasing, the sequence $\{f(x^k)\}$ has a limit. Hence, for each $\varepsilon>0$, only finitely many step sizes $\|x^{k+1}-x^k\|$ are greater than ε. Thus, $\|x^{k+1}-x^k\|\to 0$ as $k\to\infty$.

THEOREM. *Let $f:R^n\to R$ be continuously differentiable and suppose that it has a unique minimum along any line. Starting with x^1, suppose that the above algorithm produces the sequence $\{x^k\}$, Furthermore, suppose that all points generated by the algorithm are contained in a compact set. Then each accumulation point $\bar{x}$ of the sequence $\{x^k\}$ must satisfy $\nabla f(\bar{x})=0$.*

PROOF. Let $\{y_i^k\}$ and $\{\lambda_i^k\}$ be the sequences generated according to (3-1) through (3-4). By the lemma, $\lambda_i^k\to 0$ for each i, as $k\to\infty$.

Now let $\bar{x}$ be an accumulation point of $\{x^k\}$, so that $x^k\to\bar{x}$ as $k\to\infty$, in some unbounded set K. By (3-1) and (3-2), $y_i^k=x^k+\sum_{j=1}^{i}\lambda_j^k d_j^k$, and hence $y_j^k\to\bar{x}$ as $k\to\infty$ in K. Since ∇f is continuous, for each $\varepsilon>0$ there is a N so that $\|\nabla f(y_i^k)-\nabla f(x)\|<\varepsilon$ for $k>N$, $k\in K$.

From (3-4), $f(y^k)^t d^k=0$

Thus, for each $k>N$, $k\in K$.

$$
\begin{aligned}
|\nabla f(\bar{x})^t d^k| &\le \|[\nabla f(\bar{x})-\nabla f(y_i^k)]^t d_i^k\|+\|\nabla f(y_i^k)^t d^k\| \\
&\le \|\nabla f(\bar{x})-f(y_i^k)\|\cdot\|d_i^k\| \\
&= \|\nabla f(\bar{x})-\nabla f(y_i^k)\| \\
&< \varepsilon
\end{aligned}
$$

Therefore, $f(\bar{x})^t D^k\to 0$ as $k\to\infty$ in K. Since each column of D^k has norm 1, there exists an n X n matrix D such that $D^k\to D$ as $k\to\infty$ in K', where K' is an unbounded subset of K. Since $|\det D^k|\ge\delta$, then $|\det D|\ge\delta$ Hence, columns of D

are linearly independent. Thus $\nabla f(\bar{x})=0$.

References

1. E.W. Cheney and A.A Goldstein, Newton's method for convex programming and Tchebycheff Approximation, Nume rische Mathematik 1, 253-268, 1959.
2. J.W. Daniel, The conjugate gradient method for linear and nonlinear operator equations, SIAM J. Numer. Anal., 4, 10-26, 1967.
3. J. W. Daniel, Convergence of the conjugate gradient method with computationally convenient modifications, Numerische Mathematik, 10, 125-131, 1967.
4. R. Hooke and T.A. Jeeves, Direct search solution of numerical and statistical problems, J. Assoc. Comp. Mach., 8, 212-221, 1961.
5. S. Lang, Linear Algebra, Addison-Wesley Pub. Co., 1973.
6. D. G. Luenberger, Linear and Nonlinear Programming, Addison-Wesley Pub, Co., 1984.
7. D.M. Simmons, Nonlinear Programming for operations research, Prentice-Hall Inc., 1975.
8. P. Wolfe, Convergence conditions for ascent methods II: Some corrections, Siam Review, Vol 13, No2, April, 1971.

Korea Maritime University
Pusan 606
Korea