相互直交化된 方向벡터를 利用한 線追跡法에 對한 硏究

金 章 郁*

On the Line Search Method Using Mutually Orthonormal Directions

Chang-Wook Kim

〈目 次〉

1. 서 론
2. 상호직교화된 방향벡터를 이용한 선추적법
3. 반복 좌표방향벡터를 이용하는 방법
4. 수렴정리
참고문헌

1. 서 론

비제한적 최적화문제는 제한되어 있지 않는 함수를 최소화시키는 문제를 다루는 것이다. 비제한적 최소화문제의 해를 구하는 방법중 널리 알려진 방법들은 심플렉스방법([2]), 패턴추적법([4]), 구배법([8]), 뉴우톤방법([1]), 그리고 공액구배법([3]) 등이 있다. 이러한 방법들은 제한적 최소화문제를 푸는데 자연스럽게 이용된다([6],[7]).

본 논문에서, 상호 직교화된 방향벡터를 이용하는 선추적법과 반복 좌표방향벡터를 이용하는 방법에 대해 다룬다. 이 두 방법에 대한 알고리즘과 이 두 알고리즘을 포함할 수 있는 일반화된 알고리즘을 만들고 이에 대한 수렴정리를 확립한다.

2. 상호직교화된 방향벡터를 이용한 선추적법

이 방법은 추적방향으로 R^n 내에 있는 상호직교화된 벡터를 이용한다.

k 번째 단계에서, 우리는 한 점 x^k 와 상호직교화된 방향벡터 $d_1^k, \cdots, d_n^k$ 를 가진다. 방향벡터 $d_1^k, \cdots, d_n^k$ 에 따라 계속적으로 함수 f 를 최소화시켜서 x^{k+1} 를 얻는다. $x^{k+1}=x^k+\sum_{j=1}^{k}\lambda_j^k d_j^k$ 라 하자. 이때 우리는 다음과 같은 Gram—Schmidt 직교화 과정([5])에 의해 새로운 직교화된 방향벡터들의 집합을 얻는다.

* 正會員, 한국해양대학

$$a_j^{k+1}=\begin{cases} d_j^k & , \ \lambda_j^k=0 \\ \sum_{i=j}^{n}\lambda_i^k d_i^k, & \lambda_j^k \neq 0 \end{cases}$$

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

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

직교화된 방향벡터를 이용한 선추적법을 다음과 같이 알고리즘화 한다.

이것은 함수 $f(:R^n \to R)$을 최소화하기 위한 것이다.

초기단계 : $\varepsilon>0$ 을 결정상수라 하고 좌표방향벡터인 $d_1', \cdots, d_n'$ 을 선택한다. 그리고 출발점 x^1 을 선택하여 $y_1^1=x^1$, $k=j=1$ 라 두고 주 단계로 간다.

주 단 계 : 1. $\lambda \in R$ 에 제한하여 $f(y_j^k+\lambda d_j^k)$를 최소화하는 문제의 해를 λ_j^k 라 하고, $y^k{}_{j+1}=y_j^k+\lambda_j^k d_j^k$ 라 하자.

만약 $j<n$ 이면 j 대신에 $j+1$ 에 대해 1단계를 반복한다.

만약 $j=n$ 이면 step 2 로 간다.

2. $x^{k+1}=y^k{}_{n+1}$ 라 두자.

만약 $\|x^{k+1}-x^k\|<\varepsilon$ 이면 x^{k+1} 를 구하는 해로 한다.

만약 $\|x^{k+1}-x^k\|\geq\varepsilon$ 이면 $y_1^{k+1}=x^{k+1}$ 라 두고 k 대신에 $k+1$ 로 대치한다.

그리고 $j=1$ 라 두고 3단계로 간다.

3. (2—1)에 의해 상호직교화된 추적방향벡터의 집합을 만들고 1단계를 반복한다.

3. 반복 좌표방향벡터를 이용하는 방법

이 방법은 좌표축들을 추적방향으로 하여 반복적으로 시행하는 것이다. 이 방법은 다음과 같이 알고리즘화 할 수 있다.

초기단계 : $\varepsilon>0$ 을 결정상수로 취하고 $d_1, \cdots, d_n$ 을 각 좌표방향이라 하자.

x^1 을 출발점으로 하고 $y_1^1=x^1$, $k=j=1$ 라 두어 주단계로 간다.

주 단 계 : 1. $\lambda \in R$ 에 제한하여 $f(y_j^k+\lambda d_j^k)$를 최소화하는 문제의 해를 λ_j^k 라 하고, $y^k{}_{j+1}=y_j^k+\lambda_j^k d_j^k$ 라 하자.

만약 $j<n$ 이면 j 대신에 $j+1$ 에 대해 1단계를 반복한다.

만약 $j=n$ 이면 step 2 로 간다.

2. $x^{k+1}=y^k{}_{n+1}$ 라 두자.

만약 $\|x^{k+1}-x^k\|<\varepsilon$ 이면 x^{k+1} 를 구하는 해로 한다.

만약 $\|x^{k+1}-x^k\|\geq\varepsilon$ 이면 $y_1^{k+1}=x^{k+1}$ 라 두고 k 대신에 $k+1$ 로 대치한다.

그리고 $j=1$ 라 두고 1 단계로 한다.

4. 수 렴 정 리

상호직교화된 방향벡터를 이용한 선추적법과 반복좌표 방향벡터를 이용한 방법을 모두 포함시키는 일반화된 알고리즘을 다음과 같이 만든다.

k 번째 단계에서 한 점 x^k 와 방향벡터 $d_1^k, \cdots, d_n^k$ 를 가졌다고 하자.

방향벡터를 $d_1^k, \cdots, d_n^k$ 에 대한 조건은 각 j 에 대해 $\|d_j^k\|=1$ 이고, $|\det D^k|\geq\delta$ 인 양수 j 가 존재하는 것이다. 여기서 $D^k=(d_1^k, \cdots, d_n^k)$ 이다.

x^k 를 출발점으로 하여 방향벡터 $d_1^k, \cdots, d_n^k$ 을 따라 반복적으로 최소값 $\lambda_1^k, \cdots, \lambda_n^k$ 를 얻어 한 점 $x^{k+1}=x^k+D^k\lambda^k$ 를 얻는다. 여기서 $\lambda^k=(\lambda_1^k, \cdots, \lambda_n^k)$ 는 각 방향벡터 d_n^k 를 따라 움직인 거리를 말한다.

즉,

$$\begin{cases} y_1^k=x^k & (4-1) \\ y^k_{i+1}=y_i^k+\lambda_i^k d_i^k, \quad i=1, \cdots, n & (4-2) \\ x^{k+1}=y^k_{n+1} & (4-3) \\ f(y_i^k+\lambda_i^k d_i^k)\leq f(y_i^k+\lambda d_i^k), \quad \lambda\in \boldsymbol{R}, \quad i=1, \cdots, n & (4-4) \end{cases}$$

새로운 D^{k+1} 는 각 j 에 대해 $\|d_j^{k+1}\|=1$ 를 만족하고, $|\det D^{k+1}|\geq\delta$ 를 만족하는 한 특수한 법칙에 의해 함수값이 생성되는 (λ^k, D^k, x^k) 에 대한 함수에 의해 결정된다.

이 알고리즘은 상호직교화된 방향벡터를 이용한 선추적법과 반복좌표 방향벡터를 이용한 방법을 일반화시킨 것이다.

다음의 보조정리는 위에서 논의된 알고리즘에 대한 수렴정리를 증명하기 위해 필요하다.

【보조정리】 함수 $f(:\boldsymbol{R}^n\to\boldsymbol{R})$ 가 연속이고 $x^1\in\boldsymbol{R}^n$ 이라 하자. 또 $\{d^k\}$ 를 $\|d^k\|=1$ 인 방향벡터의 수열이라 하자. 점 x^1 으로부터 출발하여 함수 f 를 x^{k-1} 로부터 방향 d^k 에 따라 최소화하여 x^k 를 얻어 수열 $\{x^k\}$ 를 만들었다고 하자. 만약 수열 $\{x^k\}$ 가 어떤 콤팩트집합 X 에 포함되고 함수 f 가 어떠한 선상에서도 유일한 최소값을 가진다고 하면 $\lim_{k\to\infty}\|x^{k+1}-x^k\|=0$ 이다.

【증명】 각 $x\in X$ 와 $\|d\|=1$ 인 $d\in\boldsymbol{R}^n$ 에 대해 $S(x, d)$ 를 모든 λ 에 대해 $f[x+s(x, d)d]\leq f(x+\lambda d)$ 가 되는 한 수라 하자. 그리고 $\varepsilon>0$ 에 대해 $h(\varepsilon)=\inf\{f(x)-f[x+s(x, d)d] : S(x, d)\geq\varepsilon, x\in X, \|d\|=1\}$ 라 두자. 모순에 의해 각 $\varepsilon>0$ 에 대해 $h(\varepsilon)>0$ 임을 밝힌다.

어떤 $\varepsilon>0$ 에 대해 $h(\varepsilon)=0$ 라 가정하자. 하한의 정의와 X 의 콤팩트성질에 의해 $x^k\to\bar{x}$, $d^k\to\bar{d}$, $s^k=s(x^k, d^k)\to S\geq\varepsilon$, $f(x^k)-f(x^k+s^kd^k)\to 0$ 인 자연수의 무한집합 $\{k\}$ 를 얻을 수 있다.

각 λ에 대해 $f(x^k+s^kd^k)\leq f(x^k+\lambda d^k)$이기 때문에 각 λ에 대해 $f(\bar{x})=f(\bar{x}+s\bar{d})\leq f(\bar{x}+\lambda\bar{d})$이다.

이것은 함수 f가 임의의 선에 대해 유일한 최소값을 가진다는 것에 모순이 되므로 각 $\varepsilon>0$에 대해 $h(\varepsilon)>0$.

수열 $\{f(x^k)\}$가 아래로 유계이고 감소하므로 수열 $\{f(x^k)\}$는 극한값을 가진다. 따라서 각 $\varepsilon>0$에 대해 유한개의 $\|x^{k+1}-x^k\|$만 ε보다 크다. 따라서 $\lim_{k\to\infty}\|x^{k+1}-x^k\|=0$.

【정리】 함수 $f(:R^n\to R)$가 연속적으로 미분가능이라 하고, 이 함수가 임의의 선상에서 단 하나의 최소값을 가진다고 가정하자. 또한 x^1을 출발점으로 하여 위에서 논의된 일반화된 알고리즘에 의해 수열 $\{x^k\}$를 얻었다고 하자. 나아가서, 위에서 논의된 일반화된 알고리즘에 의해 얻어지는 모든 점들이 어떤 콤팩트집합에 포함된다고 가정하자. 이때, 수열 $\{x^k\}$의 각 집적점 $\bar{x}$는 반드시 $\nabla f(\bar{x})=0$을 만족한다.

【증명】 $\{y_i^k\}$와 $\{\lambda^k{}_i\}$를 (4—1)에서 (4—4)까지에 따라 생성된 수열이라 하자. 이때, 보조정리에 의해 각 i에 대해 $\lambda_i^k\to 0$이다.

$\bar{x}$를 수열 $\{x^k\}$의 한 집적점이라 하자.

이때 비유계집합 K가 존재하여 $k\to\infty$, $k\in K$일 때 $x^k\to\bar{x}$이다.

(4—1)과 (4—2)에 의해 $y_i^k=x^k+\sum_{j=1}^{i}\lambda_j^kd_j^k$

따라서, $k\to\infty$, $k\in K$일 때 $y_i^k\leftrightarrow\bar{x}$이다.

∇f가 연속이므로 각 $\varepsilon>0$에 대해 $k\in K$이고 $k>N$이면 $\|\nabla f(y_i^k)-\nabla f(\bar{x})\|<\varepsilon$되는 자연수 N가 존재한다.

(4—4)에 의해 $f(y_i^k)^td_i^k=0$.

따라서, 각 $k\in K$, $k>N$에 대해

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

따라서 $k\to\infty$, $k\in K$일 때 $f(\bar{x})^tD^k\to 0$이다.

D^k의 각 행벡터들의 노름이 1이기 때문에 $k\to\infty$, $k\in K'$일 때 $D^k\to D$인 n차 행열 D와 K의 비유계부분집합 K'가 존재한다.

$|\det D^k|\geq\delta$이기 때문에 $|\det D|\geq\delta$이다.

따라서, D의 행벡터들은 일차독립이다.

그러므로, $\nabla f(\bar{x})=0$이다.

참 고 문 헌

1) E. W. Cheney and A. A. Goldstein, Newton's method for convex Programming and Tchebycheff Approximation ' Numerische 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, No. 2, April, 1971.