

비선형 최적화의 수렴속도 개선에 관한 연구

이영진, 이권순, 이준탁
동아대학교 전기공학과

A Study on Development of Convergence Time in Nonlinear Optimization Problem

Young J. Lee, Kwon S. Lee, Jun T. Lee
Dept. of Electrical Eng.
Dong-A University, Pusan, Korea

Abstract

The simulated annealing(SA) algorithm is a stochastic strategy for search of the ground state and a powerful tool for optimization, based on the annealing process used for the crystallization in physical systems. It's main disadvantage is the long convergence time. Therefore, this paper shows that the new algorithm using SA can be applied to reduce the computation time. This idea has been used to solve the estimation problem of the nonlinear parameter .

1. 서론

현재, 많은 최적화 알고리즘들이 순회 판매원 문제(Traveling Salesman Problem, TSP), LSI의 설계에서 부품의 최적화 배치등과 같은 비선형 파라미터 추정 (nonlinear parameter estimation, NPE)에서 시스템의 해를 쉽게 구할 수 없는 문제들에 적용되어 왔다[2]. 그러나, 일반적으로 최적화 알고리즘에 의하여 이들 문제가 해결될 때에는 초기값의 선정에 따라 최적한 해를 구하지 못하고 극소점(local minimum)에 빠져 버리거나 최적해를 구하는데 많은 시간을 소비하는 등 여러가지 문제가 뒤따르고 있다. 때문에 보다 나은 개선책으로 시뮬레이티드 에닐링(simulated annealing, SA)법이 개발되었고, 최근 들어 이러한 NP(nondeterministic polynomial)-complete 문제의 최적 상태인 최소점 에너지(global minimum energy)를 찾기 위하여 SA법이 많이 적용되고 있다[3].

하지만, 계산 시간이 너무 막대하게 걸려서 실용적이지 못하다는 단점을 갖고 있기 때문에 실행 속도를 빠르게 하기 위한 적절한 초기온도(starting temperature)를 설정하거나 초기 온도로 부터 보다 빠른 속도로 수렴해 가도록 온도를 낮추어 나가는 방법등을 선택함으로써 SA법을 개선해 보기 위하여 시도해 보고 있지만, 그다지 좋은 결과를 얻지 못하고 있다. 따라서, 본 논문에서는 SA 알고리즘이 가지는 많은 문제점들 중 계산 시간을 줄이기 위한 방법으로서 초기값 선정에 강건(robustness)하며 보다 높은 에너지로의 이동(uphill movement)이 바탕이 되어 최적해에 도달할 확률을 높일 수 있는 알고리즘을 SA법을 토대로 개선하여 얻고자 하며 동시에 SA법이 갖는 계산시간을 줄일 수 있는 방법을 제안하고자 한다.

2. 시뮬레이티드 에닐링법

SA법의 개념의 도입은 야금학에서 담금질(metallurgical annealing)의 유사성으로부터 시작되었다. 야금학 기술에 의

하면 담금질법은 금속을 거의 녹는점까지 가열한 후 서서히 실온 상태까지 냉각시키는 것을 말한다. 즉, 담금질의 개념은 전위(轉位)나 다른 크리스탈격자 분열(crystal lattice disruption)이 높은 온도에서 온도 격동(temperature agitation)으로 제거되어질 수 있고, 금속을 천천히 식힘으로써 새로운 전위(轉位) 형성을 막을 수 있지만, 온도 강하(dropping)에 따라 발생하는 어떤 전위(轉位)를 회복시키는 데는 시간이 필요하다 는 것이다. 다시 말해서, 이 과정의 본질은 금속의 전(全)에너지 함수(global energy function)가 결과적으로 어떤 절대적 최소값(absolute minimum value)에 도달하는 것이다. 만약, 그 온도가 급격히 떨어지면 격자 전위는 금속 내부에서 동결되기 때문에 금속 격자 에너지는 전(全)에너지 함수의 최소값보다 더 높아지게 된다.

SA법의 기본 개념은 Nicholas Metropolis, Arianna Rosenbluth 등에 의해서 1953년에 제안된 이후 그들은 각 분자들의 상호작용으로 이루어진 물질의 상태방정식과 같은 특성을 조사할 수 있는 Monte Carlo법에 관심을 가졌다. 이 방법은 상반되는 목적들로 조합된 목적함수(objective function)와 대단히 많은 자유도(degrees of freedom)를 갖는 문제를 해결하기 위한 것이며, 그림 1과 같이 매우 많은 극소점들 중 절대적

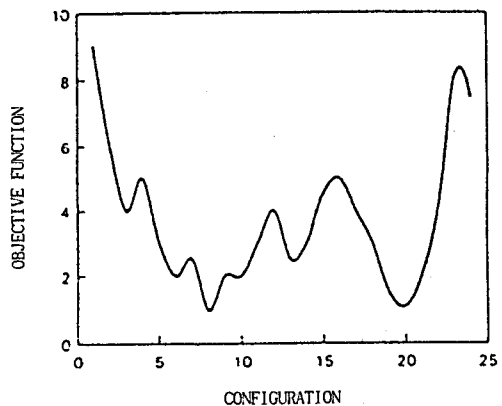


그림 1 상반된 목적을 갖는 다중 시스템에 대한 가상의 목적함수.

최소값을 찾는 문제로 축약될 수 있다[4]. 따라서, 이 방법은 쉽게 최적화 문제로 확장될 수 있으며, 그 개념은 알고리즘의 전과정에서 임의로 선택된 시스템의 일부분을 한번에 일시적으로 변화시켜, 이 변화가 전(全)시스템 에너지 상태를 낮추어 간다면 그 변화가 받아들여지고, 이 변화가 전(全)시스템에

너지 상태를 높게 되면 이 변화는 볼츠만 분포로 주어진 확률 P로 받아 들어진다. 온도를 서서히 내려가면서 확률적인 다이내믹스(dynamics)를 실행하여 최소점을 구해 나가는 SA법은 이진 벡터(binary vector) X의 실수값 함수(real-valued function) f(x)를 최소화 하기 위한 조합문제(combinatorial problem) 최적화의 한 방법으로 제안되었다. 그 알고리즘은 다음 7가지 순서로 축약될 수 있다:

- 1) 난수(random numbers)적으로 벡터 X와 초기온도 T중에서 큰값을 선택한다(모든 에너지 변화 Δ에 대하여 $\exp(-\Delta/T) \geq 0.999$ 인 초기온도를 선택한다).
- 2) 균등하게 발생시킨 난수(uniform random numbers)에서 X의 성분 X_i 를 임의로 선택한다.
- 3) 0에서 1사이 혹은 1에서 0사이중 적절하게 변화 되어진 성분 X_i 값을 제외하고는 $X = X^*$ 로 두면, $\Delta = f(X^*) - f(X)$ 이다.
- 4) 만약 $\Delta < 0$ 이면, $X = X^*$ 라 놓고 6)으로 간다.
- 5) 만약 $\Delta \geq 0$ 이면, 확률 $P = \exp(-\Delta/T)$ 를 가지며 $X = X^*$ 라 놓는다. 즉, 균일한 확률 분포를 통하여 0과 1사이의 난수 ζ를 선택하여 $\zeta < P = \exp(-\Delta/T)$ 이면, $X = X^*$ 로 놓고, 그렇지 않으면 X를 그대로 둔다.
- 6) 만약 현재의 온도값에서 X의 성공적인 변화값이나 최대 허용 변화량이 일어나면, T를 α의 크기로 감소시킨다. 여기서, α는 일반적으로 0.8 - 0.9999 사이의 온도로서 쿨링 스케줄(cooling schedule) 상수이다.
- 7) 만약 f의 최소값이 주어진 ε보다 더 감소하지 않는다면 멈추고, 그렇지 않으면 2)로 되돌아 간다.

예를들어, TSP에 대하여 살펴보면, N개의 도시가 주어지고 우리는 한번에 모든 도시를 최단거리로 여행하여 다시 시작지점으로 돌아오기를 원한다. 이때 각 도시간의 거리를 비용함수(cost function)로 두고 쿨링 스케줄에 따라 비용함수를 최소화 시켜나가는 것이며, 생각할 수 있는 모든 가능한 여행경로의 수가 $N!/2N$ 로 되는 이 문제는 N이 증가함에 따라 계산량은 거의 지수적으로 증가하므로 좋은 해법이 요구되어지는데, SA법에 의하여 근사적으로 해결되어질 수 있었고 이 근사해는 거의 최적의 범위 내에 있다고 할 수 있다[5]. 하지만, 앞서술한 바와 같이 많은 시간을 필요로 한다는 단점을 해결하기 위하여 다음의 몇가지 방법이 고려된다.

3. 수렴속도 개선책

3-1. 초기 온도 결정

먼저 다음과 같이 볼츠만 분포함수(Boltzmann distribution function)를 SA법에 적용하여 보았을 때, 어떤 변화가 시스템 전체 에너지의 증가로 귀착한다면 이 변화는 다음의 식

$$P = \exp(-\Delta E/T) \quad (1)$$

에 의한 확률 P에 의하여 이루어지며, 여기서 $\Delta E = E_{\text{new}} - E_{\text{old}}$ 이고, T는 상태 천이 규칙에서 이용한 온도 파라미터이다.

식 (1)로 부터 에너지가 최소가 될때에 상태가 최대 확률로 나타난다는 것을 알 수 있으며 온도 T가 0의 극한에서 최소 에너지 상태를 가질 확률이 1로 된다. 앞서 지적한 바도 있지만, 식 (1)로 부터도 온도 파라미터 T가 SA법에 있어서 대단히 중요하다는 것을 알 수 있다. 어쨌든, 온도가 낮을수록 최소값을 가질 확률이 매우 크게 될 것이다. 만약, 온도 T를 처음부터 0의 상태에서 출발시킨다면 상태는 에너지 함수 극소점으로 수렴하기 때문에, 반드시 최소점을 취할 수가 없으므로 초기온도의 선택과 이 온도를 줄여나가는 방법은 계산 시간적 측면에서 보면 대단히 중요한 요소 중의 하나이다. 이외에도 다음의 식

$$P = T/[T^2 + \Delta E^2] \quad (2)$$

과 같이 코쉬 분포 함수(Cauchy distribution function)라고 불리는 분포 함수를 선택함으로써 볼츠만 분포함수보다 큰 스텝으로 증가하는 분포를 갖게 하여 속도를 고속화할 수 있는 것도 보고되고 있다. 여기서 $\Delta E = E_{\text{new}} - E_{\text{old}}$ 이고, T는 온도 파라미터이다. 실제로 그림 2에서 코쉬 분포가 볼츠만 분포 보다 큰 스텝으로 증가한다는 것을 알 수 있다. 식 (1)과

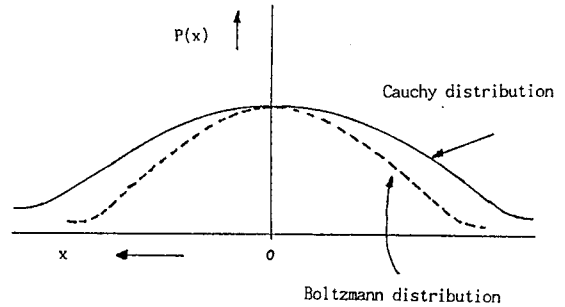


그림 2 볼츠만 분포에 대한 코쉬 분포의 비교.

(2)에서 보는 바와 같이 실행 확률은 초기 온도가 높으면 에너지의 작은 차이에서는 둔감하게 되며, 초기 온도가 매우 낮으면 민감하게 되므로 시스템에 대한 가장 적절한 초기온도 즉, 실행 시간을 최소화 하면서 해의 최소값을 가지도록 하기위한 온도를 설정하는 문제는 최적화의 의미에서 대단히 중요하다고 생각된다.

3-2. 쿨링 스케줄

또한, 온도를 줄여 나가는 방법(cooling schedule)을 생각해 볼 수 있는데 이는 지금까지 많은 사람들에 의하여 제안되어 왔다. 특히, 그 중에서 응용 수학자 Geman형제가 화상 해석에 볼츠만 기계를 응용한 논문에서 화상을 랜덤 마르코프 과정으로 해석하여 볼츠만 기계(Boltzmann machine)의 최적 상태로 수렴하도록 다음의 조건을 만족하게 낮추어 가면 반드시 최소점에 수렴할 수 있다고 증명하였다[6].

$$T(t) = T_0 / \log(1 + t) \quad (3)$$

여기서 T_0 는 초기온도이며, t는 시간을 나타낸다. 물론 SA법이나 볼츠만 기계에서는 온도를 낮추어 가면서 에너지 함수의 어떤 극소에서 다른 극소로 천이하는데 시간이 많이 걸린다. 때문에 현재로서는 시뮬레이션 시간이 너무 길어져 실용적이지 못한 단점도 있지만, 아직까지 이 방법은 극소에 빠지지 않고 최소를 구하면서 해의 수렴이 보장되고 있는 유일한 방법이라 할 수 있다. 실제로 식 (3)의 온도 변화는 대수적으로 매우 완만하므로 시뮬레이션 시간이 너무 막대하게 될 문제점이 있다. 때문에 SA법 적용시에 다소간의 발견법적(heuristic)인 고속화를 행하고 있으며, 그 한 방법으로써 Szu와 Hartley에 의하여 1987년에 개발된 다음의 식

$$T = T_0 / (1 + t) \quad (4)$$

과 같은 코쉬 트레이닝(Cauchy training)이라고 불리는 방법을 코쉬 분포 함수와 함께 이용하여 수렴 속도를 고속화할 수 있는 것도 보고되고 있다[7]. 여기서 T_0 는 초기온도이며, t는 시간을 나타낸다. 식 (4)에서 보는 바와 같이 대수적인 식 (3)대신 역수적인 식으로 속도를 고속화한다는 것을 알 수 있다.

3-3. 제안된 알고리즘

지금까지의 관점은 달리 기존의 최적화 알고리즘이 갖는 장 단점을 이용하여 새로운 알고리즘을 제안하는 것을 생각해 보면, 프로그램 상에서 주어지는 파라미터의 초기값(initial guess)에 대한 강인성과 극소점에 수렴하는 기존의 단점을 보완하기 위하여 그림 3과 같은 알고리즘을 생각해 보자. 제안

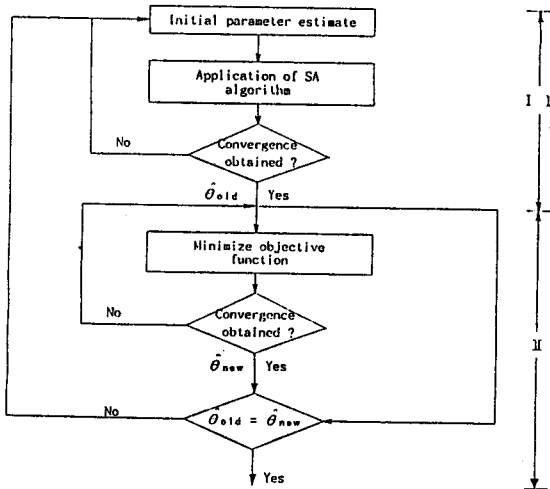


그림 3 새로운 알고리즘에 관한 흐름도 차트.

된 알고리즘은 크게 세부분으로 구성되어 있으며 그중 첫번째 단계인 I은 거의 어느 초기값에 대해서도 발산하지 않는 SA 알고리즘을 적용하여 임의의 초기값으로부터 어느 정도 최소점 부근의 파라미터를 찾는다. 그러나, 이 방법을 계속 적용하게 되면 최소점 부근에서 방대한 양의 계산을 요구하므로 막대한 비용이 요구되어진다.

그러므로, 단계 II에서는 다변수 비선형 함수(multivariable nonlinear function)의 최소점을 찾기 위하여 Fletcher와 Reeves 알고리즘과 같이 steepest descent 방향과 conjugate 방향에 의하여 최소점의 탐색이 이루어진다. 그러나 여기서 얻은 값이 최소점에 도달하기에는 무리가 있으므로 극소점에 머무를 수가 있다. 그러므로 단계 III에서 극소점에서 보다 높은 에너지로의 이동을 통하여 최소점을 찾아간다.

본 알고리즘에서는 최소점에 도달한 후 계산 종료 여부를 결정하는 기준(termination criteria)으로서 다음 식을 사용한다.

$$|F(x_{i+1}) - F(x_i)| \leq \epsilon, \quad i = 1, 2, 3, \dots, N \quad (5)$$

그러나, 임의의 초기값으로부터 극소점 부근까지의 이동을 위한 단계 I과 최소점을 찾아가는 단계 III의 ϵ 값은 달리 설정되어져야 계산시간을 줄여야 하는 경제적 측면에서 합리적인 것이라 생각된다.

4. 시뮬레이션 결과 및 토의

본 논문에서는 TSP문제에 적용하여 수행시간을 비교하여 보았다. TSP문제는 그 해를 쉽게 구할 수 없다고 앞서 지적한바 있지만 그 수행시간이 대단히 많이 소요된다는 단점을 가지고 있기 때문에 이를 줄일 수 있는 방법으로서 초기온도 및 쿨링 스케줄을 변화시켜서 비교한 결과를 보였다.

이는 모두 10개의 도시에 대하여 수행하였으며 이 문제의 $10!/20$ 의 경우의 수를 가진다. 그림 4(a)와 4(b)는 10개의 도시에 대하여 최초로 선정된 경로와 SA법에 의하여 최적화된 경로를 도시하여 보았다.

Traveling Salesman Problem

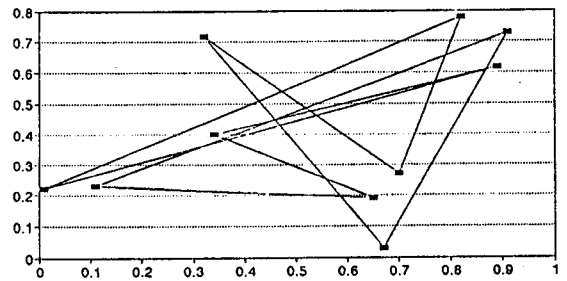
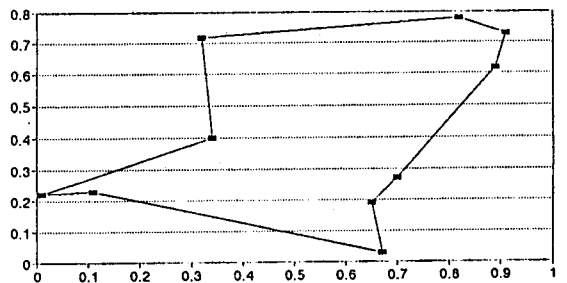


그림 4(a) 10개의 도시에 대한 초기 경로.

TSP solved by simulated annealing



-- city

그림 4(b) 10개의 도시에 대한 SA법에 의하여 최적화된 경로.

표 1은 초기 온도에 대한 결과 값들의 비교이며, 표 2와 3은 쿨링 스케줄에 대한 수행 결과 값들을 비교한 것이다.

Starting temperature	Path length	Counter
0.05	2.69	12,000
0.5	2.69	22,608
10.5	2.69	28,393

표 1 쿨링 스케줄의 변화에 따른 수행 시간 비교.

Cooling schedule	Path length	Counter
0.9	2.69	47,457
$T/\log(1+\text{count})$	2.69	5,314
$T/(1+\text{count})$	2.69	2,563

표 2 초기온도의 변화에 따른 수행시간 비교.

Cooling schedule	Path length	Counter
0.9	2.69	22,608
$T/\log(1+\text{count})$	2.69	2,970
$T/(1+\text{count})$	2.69	2,153

표 3 볼츠만 분포에 의한 쿨링 스케줄 변화에 따른 수행 시간 비교.

그림 5는 볼츠만 분포에 따른 초기온도의 비교를 나타내었으며 그림 6과 7는 각각 볼츠만과 코쉬 분포에 따른 쿨링 스케줄을 비교 수행하여 보았다.

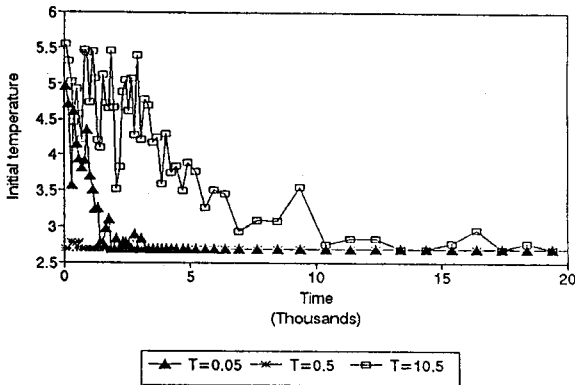


그림 5 볼츠만 분포에 대한 초기온도의 수행 시간 비교.

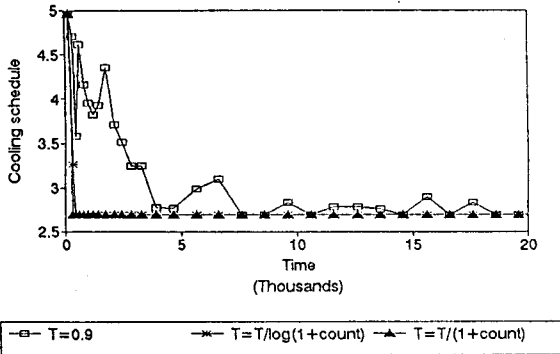


그림 6 볼츠만 분포에 대한 쿨링 스케줄의 수행 시간 비교

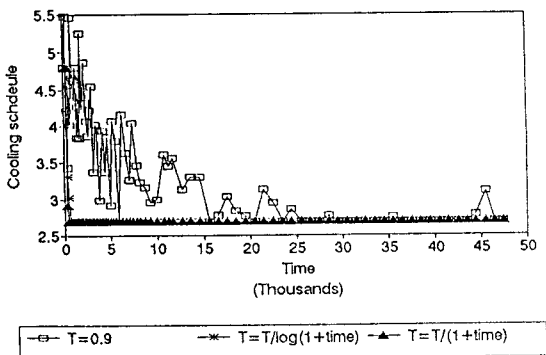


그림 7 코쉬 분포에 대한 쿨링 스케줄의 수행 시간 비교.

참고문헌

[1] Harold SZU & Ralph HARTLEY, "Fast Simulated Annealing," Physics Letters A, Vol. 122, No. 3,4, pp. 157-162, 1987.
 [2] H. William, et al., Numerical Recipes: The Art of Scientific Computing, pp. 326-334, Cambridge Univ. Press, New York, 1986.

[3] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, "Optimization by Simulated Annealing," Science, Vol. 220, pp. 671-680, 1983.
 [4] Robert Hecht-Nielsen, Neurocomputing, pp. 192-197, Addison-Wesley Publishing Company, New York, 1990.
 [5] Mario P. Vecchi, et al., "Global Wiring by Simulated Annealing," IEEE Trans. Computer-Aided Design, Vol. CAD-2, No. 4, pp. 215-22, 1983.
 [6] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions and Bayesian Restoration of Images," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 6, pp. 721-741, 1984.
 [7] Philip D. Wasserman, Neural Computing: Theory and Practice, pp. 81-92, Van Norstrand Reinhold, New York, 1989.